Dissertation

Submitted to the

Combined Faculties of Natural Sciences and Mathematics

of the

Ruperto-Carola-University of Heidelberg, Germany

for the degree of

Doctor of Natural Sciences

Put forward by

Christof Sauer

born in Fürth, Germany

Oral examination: October 24, 2023

Measurement of the Triple-Differential Cross-Section for the Production of Multijet Events using $139\,\mathrm{fb}^{-1}$ of Proton-Proton Collision Data at $\sqrt{s} = 13\,\mathrm{TeV}$ with the ATLAS Detector to Disentangle Quarks and Gluons at the Large Hadron Collider

Referees:

Prof. Dr. André Schöning

Prof. Dr. Ulrich Uwer

Nekrolog

An dem wolkenlosen Morgen des sechzehnten Aprils des krisengeschüttelten Jahres 2021 um präzise 10:37 Uhr Ortszeit ereilte mich in Heidelberg die niederschmetternde Kunde der erst kürzlich diagnostizierten, schwerwiegenden Krebserkrankung meines Vaters, welche den Auftakt einer über Monate währenden Chronologie sich gegenseitig abwechselnder Hoffnungen, des ständigen Bangens und herber Niederschläge markierte. Diese dornenreiche Zeit hat meinen Blickwinkel auf die Welt nachhaltig verschoben und mir vergegenwärtigt, dass lediglich zwei Aspekte des menschlichen Lebens wirklich von Relevanz zu sein scheinen: die Liebe und der Tod – beide als Brennpunkte der Ellipse des menschlichen hienieden Daseins.

Wenige Momente im Leben vermögen die Welt auf vergleichbare Weise aus ihren Angeln zu heben, die tönernen Füße wankender Existenz zu zertrümmern, wie die unbeschreibliche, zermarternde Furcht, einen über alle Maßen geliebten Menschen zu verlieren. Die permanente Sorge und Angst erfüllt den gesamten Gedankenraum und tätigt psychosomatische Symptome; drohend steht das Menetekel an der Wand; unaufhörlich wetzt sich die Klinge funkensprühend am Walzstein; beständig schwebt des Damokles' Schwert bedrohlich über dem bangenden Haupt. Als Familie haben wir – mein Vater, meine Mutter sowie meine Person – diese schwere Zeit der Bedrängnis gemeinsam bewältigt und gegenseitig das Joch des anderen zu Tragen versucht, wobei insbesondere mein Vater sein schweres Los in ritterlicher, bewundernswürdiger Tapferkeit ertrug. Seinem ganzen Wesen entsprechend galten seine Gedanken (auch in größter persönlicher Drangsal!) seiner Familie, die stets das Zentrum seines Lebens bildete und der er aufopferungsvoll seine ganze Hingabe – fürderhin: seine gesamte Existenz – schenkte.

Des Weiteren ist die fast übermenschliche Leistung und Selbstaufopferung meiner Mutter hervorzuheben, die in unbedingter Hingabe und grenzenloser Liebe zu ihrem schwer erkrankten Gatten (unter Vernachlässigung der eigenen Person) nichts unversucht ließ, seine Genesung zu erzielen, sein Leid zu mindern und ihm ohne Unterlass die Hand zum Troste reichte sowie in inniger Liebe das tränenbenetzte Haupt zum Kusse senkte. Hier wirkten Kräfte, mächtiger als der Tod.

Nichts in meiner bisherigen Existenz prägte mich wie diese Zeit; nichts hat das Koordinatensystem meines Lebens auf vergleichbare Art und Weise verschoben und mich daran erinnert, dass das, was wirklich zählt, die Liebe ist, welche die wenigen, schattenverwirrten Jahrzehnten unseres hienieden Daseins mit Sinn und Glückseligkeit erfüllt. Was sonst sollte Sinn stiften sein in einem Leben, in dem jeder aus kalter Materie geformte Nagel, den wir kraftvoll in das Holz treiben, das Potential in sich trägt, langer zu weilen, als unser fleichlicher, sterblicher Leib es jemals vermöchte? Im furchtbarsten Elend offenbart sich das Prinzip der schönsten und schöpferischsten Kraft im gesamten Kosmos. Was sind die wenigen Kostbarkeiten, diese unbeschreiblichen Glücksmomente, welche wir am Ende unseres irdischen Seins zutage fördern, wenn wir den Sand vergangener Reminiszenzen durch das grobmaschige Sieb der Erinnerung schütteln? Was bleibt am Ende, wenn wir die löchrigen Gewänder der Zivilisation und ihre wertlosen Abzeichen und Epauletten abstreifen? Es ist das Wissen, bedingungslos geliebt zu werden durch eine Kraft, die möchte, dass wir sind und bleiben. Hierin nun liegt das Unerhörte des Todes: dass er zu trennen trachtet, was die Liebe zur unscheidbaren Einheit geschmiedet. Doch selbst wenn die Moiren den Lebensfaden endgültig durchtrennen, so lässt sich das Band der Liebe nicht zerscheiden; es besteht weiter und verbindet uns durch Raum und Zeit – durch Dies- und Jenseits. Wer im Brennpunkt der Liebe steht (oder stand), stirbt wahrlich nie.

Eines Tages kehren wir alle zurück in den Schoß der Mutter. Unsere sterblichen Überreste vergehen und in wenigen Generationen verflüchtigen sich ebenso die Erinnerungen an uns, wenn niemand mehr ist, der unserer gedenken könnte. Die ägyptischen Pharaonen fürchteten dies so sehr, dass sie ihre Monumente mit Mriaden Kartuschen übersäten, welche in kunstvollen Rahmen und Hieroglyphen ihren Namen trugen und als Kristallisationspunkte der Erinnerung dienlich sein sollten. Doch dies ist ohne Belang, denn wir sind und damit werden wir auch immer gewesen sein, so wie jedes Präsenz zwangsläufig zum Futurum exaktum reift, selbst dann, wenn die letzte Sonne verglüht ist und sich der Kosmos in ewige Dunkelheit hüllt. Die Tatsache, dass wir sind und waren, dass wir gemeinsam lachten, uns zankten und liebten, ist

auf alle Ewigkeit geschrieben in das kollektive Gedächtnis der Alleinheit unabhängig davon, ob jemand sich dessen entsinnt.

Mein lieber Vater, der Gevatter Tod hat dich viel zu früh von uns genommen; alle Tage unsres Lebens tragen wir Dich im Herzen geborgen; kein Moment vergeht, an dem du nicht bei uns bist. Der Wind trägt deine Stimme zu uns, über den Äther versprüht sich deine Herzensgüte, als seist Du noch mitten unter uns. Unsere Liebe spricht zu dir, wo auch immer du dich verborgen hältst. Schließe ich meine schweren Lieder abends zur Ruhe, so sehe ich dein freudestrahlend' Antlitz frei von Schmerz vor mir erscheinen. In der Dunkelheit der Nacht zeichnet das Mondlich zärtlich deine Silhouette. Wo ich auch bin, trage ich den Gedanken an Dich und all meine Dankbarkeit in mir. Dein Beispiel lehrte mich ein mannhaft Herz, das stark bleibt in Freuden und Leiden, im Triumphe nicht prahlt und im Sturze nicht verzagt. Wenn die Sonne sich senkt hinterm Horizont und mein Haupt gen Himmel ich erhebe, zu schauen, wie Stern um Stern sich entzündet zum Ballett der Nacht, dann schwingt mein Geist fort zu dir in die Unendlichkeit. Und wenn die Trauer mich übermannt in düsteren Winternächten, dann denke ich, wie leise und unbemerkt unter dem kalten Schnee, der draußen glitzert, schon der kommende Frühling wieder keimt. Dann tröste ich mich und weiß: *Sämtlicher Tod ist nur Schein.*

Ich, *Christof Sauer*, widme die vorliegende Dissertationsschrift zur Erlangung des akademischen Grades „Doktor der Naturwissenschaften“ mitsamt *aller* darin gewonnenen Erkenntnisse und dargelegten Einsichten meinen beiden geliebten Eltern: meinem Vater, *Hubertus Udo Sauer* – dem es unseligerweise nicht mehr vergönnt war, diesen Moment leibhaftig mitzuerleben –, sowie meiner Mutter, *Ursula Elaine Sauer*. Dieses bescheidene Denkmal (nicht aus Stein geschlagen, nicht aus Gold gegossen, sondern aus Wissen geformt und in Liebe gebunden) sei ewig für euch errichtet.

Wir wissen nichts von diesem Hingehn, das\ nicht mit uns teilt. Wir haben keinen Grund,\ Bewunderung und Liebe oder Haß\ dem Tod zu zeigen, den ein Maskenmund

tragischer Klage wunderlich entstellt.\ Noch ist die Welt voll Rollen, die wir spielen.\ Solang wir sorgen, ob wir auch gefielen,\ spielt auch der Tod, obwohl er nicht gefällt.

Doch als du gingst, da brach in diese Bühne\ ein Streifen Wirklichkeit durch jenen Spalt\ durch den du hingingst: Grün wirklicher Grüne,\ wirklicher Sonnenschein, wirklicher Wald.

Wir spielen weiter. Bang und schwer Erlerntes\ hersagend und Gebärden dann und wann\ aufhebend; aber dein von uns entferntes,\ aus unserm Stück entrücktes Dasein kann

uns manchmal überkommen, wie ein Wissen\ von jener Wirklichkeit sich niedersenkend,\ so daß wir eine Weile hingerissen\ das Leben spielen, nicht an Beifall denkend.

– Rainer Maria Rilke (1875 – 1926), aus *“Neue Gedichte”* (1907)

Obituary

On the cloudless morning of the sixteenth of April in the crisis-ridden year of 2021 at precisely 10:37 a.m., I received devastating news about the recently diagnosed serious cancer of my father, which was the prelude to a period of alternating hopes, constant trepidation and bitter defeats that lasted for months. This thorny time has permanently shifted my perspective on the world as a whole and made me aware that there are only two aspects of human life that are really of relevance: love and death – both as foci of the ellipsis of human life.

Just a few moments in someone's life unhinge the world the same way as the indescribable, gruelling fear of losing a loved one does. The state of constant worry consumes all energy and takes up all thoughts. The omen stands threateningly written on the wall; the blade is constantly sharpened, spraying sparks on the whetstones; Damocles' sword permanently hovers menacingly above the trembling head. As a family, the three of us – my father, my mother and myself – have overcome this difficult time of affliction together. We tried to bear each other's yoke, whereby my father in particular endured his difficult fate with chivalrous, admirable bravery. In accordance with his whole nature, his thoughts were always surrounding his family (even in the greatest personal tribulations!), which always formed the centre of his life and to which he sacrificially dedicated all his devotion as well as his entire existence.

Furthermore, I'd like to emphasise and honour the almost Herculean performance and self-sacrifice of my mother who, in unconditional devotion and boundless love for her seriously ill husband, left no stone unturned to bring about his recovery, to alleviate his suffering, and at any moment held out her hand to comfort him and in heartfelt love always lowered her tear-stained head to a kiss. Here were forces at work far more powerful than death.

Nothing in my existence to date has shaped me like this time; nothing has shifted the coordinate system of my life comparably and reminded me that what actually counts is the love that gives meaning to the few shadowy decades of our existence here on earth with meaning and bliss. What else should give meaning to our existence in a universe in which each nail driven into wood has the potential to last longer than our ephemera body. However, in the most terrible misery, the principle of the most beautiful and creative force in the entire cosmos reveals itself to us mortal beings. What are the precious instances, these indescribable moments of happiness, which we bring to light at the end of our earthly existence when we scrape the sand of past reminiscences through the coarse-meshed sieve of memories? What remains at the end when we strip off the holey robes of civilization and its worthless badges? It is the knowledge that we are loved unconditionally by a power that wants us to be and to remain. This now the outrageousness of death: it seeks to separate what love has forged into an indistinguishable unity. But even if the Mœræ finally cut the thread of life, the bond of love cannot be severed; it continues to exist and connects us through space and time – through this world and the hereafter. Whoever stands (or stood) in the focal point of love never dies.

One day we all return to the womb of Mother Earth. Our remains will pass away and just in a few generations, the memories of us will fade away when there is no one left to remember. The Egyptian pharaohs feared this condition to such an extent that they decorated their monuments with myriads of cartouches, which bore their names in elaborate frames and hieroglyphics to serve as a focal point of memory for future generations to come. But this is of no importance because we *are* and thus we will always have *been* just as every simple presence inevitably matures to the past perfect simple even when the last sun has burned up and the cosmos has been wrapped in eternal darkness. The fact that we are (and have been), that we laughed together, quarreled occasionally and loved together, is written for all eternity in the collective memory of all-encompassing unity – regardless of whether anyone remembers it or not.

My dear father, the Grim Reaper has taken you from us far too soon; all the days of our lives we carry you in our hearts; not a moment passes when you are not with us. The wind carries your voice to us; your kindness of heart spreads over the ether as if you were still among us. Our love speaks to you wherever you are hidden. When I close my heavy eyelids to rest in the evening, I see your joy-beaming face freed from pain appearing before me. In the darkness

of night, the moonlight tenderly draws your silhouette. Wherever I am, I carry the thought of you and all my gratitude. Your example taught me a manly heart that remains strong in joys and sufferings, does not boast in triumph and does not despair in troublesome times. When the sun sinks behind the horizon and I raise my head to the sky to see how star after star is ignited the ballet of the night, then my spirit swings away to you in the infinity. And when sadness overcomes me on gloomy winter nights, I think how quietly and unnoticed under the cold snow, that glitters outside, the coming spring is already sprouting again. Then I comfort myself and know: *all death is only an illusion.*

I, *Christof Sauer*, dedicate the thesis at hand for the degree "Doctor of Natural Sciences" together with *all* the knowledge gained during my studies including *any* insights presented therein to my two beloved parents: my father, *Hubertus Udo Sauer* – who, unfortunately, was not able to experience this moment in flesh – and my mother, *Ursula Elaine Sauer*. This modest memorial (not carved from stone, not cast in gold, but formed from knowledge and bound in love) be eternally erected for you.

We know nothing of this going away, that\ shares nothing with us. We have no reason,\ whether astonishment and love or hate,\ to display Death, whom a fantastic mask

of tragic lament astonishingly disfigures.\ Now the world is still full of roles which we play\ as long as we make sure, that, like it or not,\ Death plays, too, although he does not please us.

But when you left, a strip of reality broke\ upon the stage through the very opening\ through which you vanished: Green, true green,\ true sunshine, true forest.

We continue our play. Picking up gestures\ now and then, and anxiously reciting\ that which was difficult to learn; but your far away,\ removed out of our performance existence,

sometimes overcomes us, as an awareness\ descending upon us of this very reality,\ so that for a while we play Life\ rapturously, not thinking of any applause[1].

– Rainer Maria Rilke (1875 – 1926), from *"New Poems"* (1907)

[1] This wonderfull translation by Cliff Crego of Rainer Maria Rilke's poem *"Death Experience"*, which somehow manages to keep the peculiar feeling of this text, has been taken from Ref. [1].

“Descend Into the Sun’s Barque ...”

Oh father! Already the last grains pearls through the glass of life; Oh father! Not a single spark is caught in the candel’s wick anymore; the light of life is extinguished in a plume of smoke; the last embers die out. Behold, The last grain has slid through the hourglass. You are no more ... Dear Father, as your tormented head rested in my hands, as you blew out the breath of life, I suddenly remembered a verse from ancient Egypt – several millennia old –, which gave me comfort at that moment.

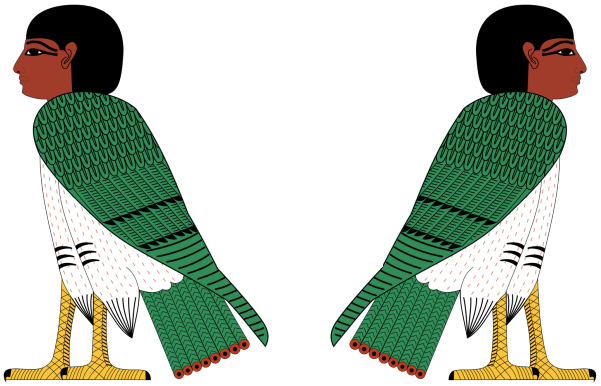
Death is in my sight today, a cure | of a sick man, like going outside

after | confinement. Death is in my sight today,

like the smell of | myrrh, like sitting under a sail

on a windy day. Death is in my sight today.

– *“Dispute of a man with his ‘ba’”* [2] (anon.), XIIth Dynasty (1991 BC–1802 BC)



Dedicated to my beloved parents, Hubertus Udo[†] *and Ursula Elaine Sauer.*

‘Ποῦ σου, θάνατε, τό κέντρον; Ποῦ σου, ἅδη, τό νῖκος;’

(»Tod, wo ist dein Stachel? Totenreich, wo ist dein Sieg?«)

– 1. Korinther 15:55-58, Schlachter-Bibel, (1905)

[†]05.04.1956–09.11.2021

Acknowledgments

The dissertation marks an important step in the life of a young scientist as it provides the opportunity for a career in research not only to follow one's passion, but also as a way of earning one's living – to combine a profession with pleasure. In such moments, it's only natural (and *to some extend* of course justified) to feel a sense of pride and satisfaction for what has been archived; however, it should also be an occasion to pause for a while and to remember that this accomplishment in life was only possible due to preconditions that are far, far beyond my own influence, but without which the work at hand would have been *absolutely* unthinkable. How often do we grid ourselves with pride and a feeling of supercity not keeping in mind that almost all of what we are is due to unmerited grace and pure coincidence. If born just one brick wall next door, our life might have turned out to be completely different. Just one hundred centimeters that decide on whether one is born in a lovely surrounding, a fertile soil on which one's potential can flourish into full bloom, or be thrown into desolate circumstances that do not allow one to pursue own interests and preferences. Consequently, the feeling of pride should be accompanied by an even greater humility and gratitude for all the good fortune that has come one's way in life. Success is a gift that comes to us undeservedly, and certainly not an outstanding merit of any human being in itself. By human standards, some may fall through the coarse sieve of social norms; however, no human being can fall lower than into the loving hand of the one power that wants us to be. Accordingly, no title should be worn as a medal or epaulette at the worthless and arbitrary uniform of society's standards, but – in silent and secret gratitude – should sharpen our eyes for our fellow men to whom the fortune of chance has been less kind. The more successful we are in life, the more humble we should be and the more respect and appreciation we should pay to those who, despite all adverse circumstances, master their lives.

This PhD thesis owes its existence to the support, help, and inspiration of numerous people who should be acknowledged at this occasion. Given the enormous amount of souls encountered and interacted with during my way of life – all of which had some kind of causal effect up to the present day –, this acknowledgment will *inevitably* commit an injustice to all of those that remain unnamed, unhonoured and without an explicit expression of gratitude. I deeply apologize to all those people who remain unappreciated due to my ignorance and fragmentary reminiscence, but who deserve to be mentioned here. Be assured of my whole-hearted, deep-felt gratitude!

Please note: the arrangement or ordering of names does not reflect a hierarchy of quality ranking.

First and foremost, I want to express my sound sense of acknowledgment and esteem to my supervisor, mentor (and friend), *Prof. Dr. André Schöning*, for the great opportunity to spend a substantial portion of my previous academic education under his patronage in the ATLAS group at the Physikalisches Institut at the Ruprecht-Karls-Universität, Heidelberg.

I must not forget to give thanks to *Prof. Dr. Ulrich Uwer* for his willingness to be the second referee of this report.

Besides the two main referees of this thesis, I want express my gratitude to *Prof. Dr. Ulrich Anton Glasmacher* from the Institute of Earth Sciences University of Heidelberg as well as *Prof. Dr. Björn Malte Schäfer* from the Institut of Theoretical Astrophysics for their willingness of being part of my examination committee.

Apart from my advisors, I want to articulate my deepest sense of gratitude and appreciation to *Dr. Matt LeBlanc* for his marvelous encouragement, guidance, and patience throughout a substantial part of this PhD project. Our fruitful, professional, and personal conversations were of inestimable value and will remain cherished memories.

A special thanks goes to the "International Max Planck Research School for Precision Tests

of Fundamental Symmetries (IMPRS-PTFS) in Particle Physics, Nuclear Physics, Atomic Physics and Astroparticle Physics at the Heidelberg University" for supporting my scientific work financially and logistically over the entire period of my research.

Furthermore, I would like to express my gratitude to our secretary at the IMPRS-PTFS, *Britta Schwarz*, for all her work and dedication. I want to thank you for your help and support during my fathers illness.

I want to thank the entire ALTAS collaboration for all their work, from the very first idea of this ambiguous project in someone's head till today. Projects like this one are a great example of human willpower and a beacon in the night, a glimmer of hope in a gruesome world in which uncountable *human beings* are constantly turned through the mill of blood and power.

Today's science always is a joint effort. In my daily work in our two research groups under the direction of my supervisor, ATLAS and Mu3e, I have been blessed with friendly and cheerful colleagues. My thanks go to (in alphabetical order): *Dr. Annie Meneses González*, *Antara Paul*, *Benjamin Weinläder*, *David Maximilian Immig*, *Dr. Heiko Augustin*, *Joachim Zinßer*, *Kadir Tastepe*, *Dr. Luigi Vigani*, *Lukas Mandok*, *Dr. Marta Czurylo*, *Mohd Talha*, *Poppy Hicks*, *Sachin Gupta*, *Dr. Sebastian Dittmeier*, *Dr. Tamasi Kar*, *Dr. Thomas Rudzki*.

Among my former colleagues (who left academia or moved to different universities), I thank: *Dr. Adrian Herkert*, *Anjali Krishnan*, *Dr. Arthur Bolz*, *Dr. Christoph Falk Anders*, *Dohun Kim*, *Dr. Louis Helary*, *Dr. Mathis Kolb*, *Dr. Mykhailo Lisovy*. I must not forget to mention my former colleague *Dr. Danilo Enoque Ferreira de Lima*, who taught me a lot during my master's thesis.

In the group of my colleagues, I'd like to give a particularly deep gratitude to *David Maximilian Immig*, *Dohun Kim* and *Dr. Tamasi Kar* for providing invaluable emotional support during the difficult phase of my father's agony. In addition to what has been said, I want to thank David-Maximilian Immig once again for his friendship and support.

I'd also like to extend my gratitude to our secretary *Claudia Wallenwein*, without whose support and involvement the academic work at the institute would be absolutely impossible to maintain. Furthermore, I would like to thank Claudia for numerous hours of her emotional support – while herself already being involved in the support of a friend in contemplation of her inevitable, soonish death – and for many tears that have been shed together from the very beginning of my father's disease up to his death.

In my private surrounding, I want to thank *Marius-Hergen* and *Björn-Arved Rauch* for their long-standing loyal friendship and mental support for more than twenty years. In addition, my gratitude goes to my friends and fellow former students *Alexander Uth* and my alter ego *Markus Kardorff*, whose friendship I value beyond measure.

I would like to thank the former school principal, *Ursula Engelberger*, as well as my former maths teacher, *Alfons Frink*, for their support in a very difficult phase of life when I was lost in a Daedalus-like labyrinth of inner and outer disorientation, with the horrible Minotaur threatening to tear me appart. I also want to express special thanks to my erstwhile physics teacher at the Gymnasium am Römerkastell, *Gudrun Hattemer*, for her encouragement and confirmation to study physics – a decision I have never regretted to the present day, even though my view of this natural science (natural science in general) has changed considerably over time.

I would like to express my special heartfelt thanks and gratitude to *Christine Heuchert* for her long-term friendship that persists to the present, although the opportunity rarely arose to meet each other in person. I thank you for numerous conversations and many merry evenings full of laughter, thoughtfulness – and wine ... plenty of wine. Thank you for forgiving me my stupidity on a regularly basis. I hope that we will stay in touch for many more years to come.

Furthermore, I voice my gratitude to *Luisa Melina Melenk* who was exactly the right person I met at precisely the right moment in time thanks to the grace of chance. Thank you for many beautiful, memorable moments.

Speaking about my fellow students of my former school, I must not forget to mention *Jonathan Trauth* and *Patrik Freudenberger*, with whom I have had no contact for a long period of time, but which are very often present in my mind.

I also wish to thank (and she is probably the least person who expects this) *N**** for her support, friendship during one phase in my life as well as for the realization of the importance of love (as the basic structure of all being) in people's lives. Always remember (even in the darkest moments): *"Nur die Liebe vermag alle Knoten zu lösen"* – Lew Nikolajewitsch Graf Tolstoi (1828-1910), diaries 1889.

An immeasurable, vast sense of gratitude, thankfulness and admiration – that can hardly be put into profane words of every day language – is addressed to *Melissa Flassig*, who provided inestimable support of all sorts during the saddest and most troublesome times of my life up to this point. Times have been very tough every now and then, but I thank *you*, Melissa, from the bottom of my heart for your patients, your loving and caring hugs, your warm kisses ... that kept me grounded and gave me something to hold on to whenever I ran the risk of slowly losing touch with reality. I am overflowing with gratitude given the grace and gift of your acquaintance! Thank you for numerous pleasant conversations, plenty moments of heartiness, intimacy and tenderness, hours of beautiful, touching guitar and violin music, as well as the blessing of your presence and your existence. Without you, the work at hand would have been absolutely unthinkable. You are truly a gem among all women, a treasure of humanity and a jewel on the diadem of stars. In light of your mesmerizing (inner and outer) beauty, Emanuel Schikaneder's words in the lovely libretto of Wolfgang Amadeus Mozart's opera "The Magic Flute" (K. 620) become reality:

> *(Pamina) Love sweetens every sorrow,* | *All creatures pay it homage.* | *(Papageno) Love adds spice to our days on earth,,* | *Love is at work throughout all nature.* | *(Both) Its exalted goal is manifest:* | *Nothing is more noble than man and wife.* | *Man and wife, and wife and man,* | *Attain divinity.*
>
> *(Pamina) Die Lieb' versüßet jede Plage,* | *ihr opfert jede Kreatur.* | *(Papageno) Sie würzet uns're Lebenstage,* | *sie wirkt im Kreise der Natur.* | *(Beide) Ihr hoher Zweck zeigt deutlich an:* | *nichts Edler's sei, als Weib und Mann.* | *Mann und Weib, und Weib und Mann,* | *reichen an die Gottheit an.*

Like the thirsty man in the scorching heat of the desert sand doth longs for a thirst-quenching spring; so doth my heart desire you. All that has been said only scratches the surface, but – as the Austrian philosopher Ludwig Josef Johann Wittgenstein (1889–1951) in his landmark essay *"Tractatus Logico-Philosophicus"* (1921) once said – *"[t]hat which cannot be said must not be said[;] [t]hat which cannot be said, one must be silent thereof[.]" (Tractatus 7).*

Last, but certainly not least, I take the opportunity to express my profound, boundless gratitude and love to my parents, *Ursula Elaine Sauer* and *Hubertus Udo Sauer*, who supported me throughout my entire life, spiritually and materially, and encouraged me in all of my life decisions up to this very point. Thank you for never telling me what I'm *supposed* to be, but always asking what I *want* to be instead.

My gratefulness to all of you is inscribed in eternity – not even time is capable to grind it down. This thanksgiving is completed by Galateia's final melody in Händel's beautiful opera "Acis and Galatea" (HWV 49):

> *Heart, the seat of soft delight,* | *Be thou now a fountain bright!* | *Purple be no more thy blood,* | *Glide thou like a crystal flood.* | *Rock, thy hollow womb disclose!* | *The bubbling fountain, lo! it flows;* | *Through the plains he joys to rove,* | *Murm'ring still his gentle love.*
>
> *Herz, der Liebe süsser Born,* | *Sei fortan ein Silberquell!* | *Purpur sei nicht mehr dein Blut,* | *Gleite gleich krystall'ner Flut.* | *Fels, entschliess' den dunkeln Schoss!* | *Der Sprudelquell, sieh'! er strömt;* | *Und er rauscht das Tal entlang,* | *Murmelnd seinen Liebessang.*

Abstract

At hadron-hadron colliders, it is almost impossible to obtain pure samples in either quark- or gluon-initialized hadronic showers as one always deals with a mixture of particle jets. The analysis presented in this dissertation aims to break the aforementioned degeneracy by extracting the underlying fractions of (light) quarks and gluons through a measurement of the relative production rates of *multijet* events.

A measurement of the triple-differential multijet cross section at a centre-of-mass energy of 13 TeV using an integrated luminosity of $139\,\mathrm{fb}^{-1}$ of data collected with the ATLAS detector in proton-proton collisions at the Large Hadron Collider (LHC) is presented. The cross section is measured as a function of the transverse momentum p_{T}, two categories of pseudorapidity η_{rel} defined by the relative orientation between the jets, as well as a Jet Sub-Structure (JSS) observable $\mathcal{O}_{\mathrm{JSS}}$, sensitive to the quark- or gluon-like nature of the hadronic shower of the two leading-p_{T} jets with $250\,\mathrm{GeV} < p_{\mathrm{T}} < 4.5\,\mathrm{TeV}$ and $|\eta| < 2.1$ in the event.

The JSS variables, which have been studied within the context of this thesis, can broadly be divided into two categories: one set of JSS observables is constructed by iteratively declustering and counting the jet's charged constituents; the second set is based on the output predicted by Deep Neural Networks (DNNs) derived from the "deep sets" paradigm to implement permutation invariant functions over sets, which are trained to discriminate between quark- and gluon-initialized showers in a supervised fashion.

All JSS observables are measured based on Inner Detector tracks with $p_{\mathrm{T}} > 500\,\mathrm{MeV}$ and $|\eta| < 2.5$ to maintain strong correlations between detector- and particle-level objects. The reconstructed spectra are fully corrected for acceptance and detector effects, and the unfolded cross section is compared to various state-of-the-art parton shower Monte Carlo models. Several sources of systematic and statistical uncertainties are taken into account that are fully propagated through the entire unfolding procedure onto the final cross section. The total uncertainty on the cross section varies between 5 % and 20 % depending on the region of phase space.

The unfolded multi-differential cross sections are used to extract the underlying fractions and probability distributions of quark- and gluon-initialized jets in a solely data-driven, model-independent manner using a statistical demixing procedure ("jet topics"), which has originally been developed as a tool for extracting emergent themes in an extensive corpus of text-based documents. The obtained fractions are *model-independent* and are based on an *operational definition* of quark and gluon jets that does not seek to assign a binary label on a jet-to-jet basis, but rather identifies quark- and gluon-related features on the level of individual distributions, avoiding common theoretical and conceptional pitfalls regarding the definition of quark and gluon jets.

The total fraction of gluon-initialized jets in the multijet sample is (IRC-safely) measured to be $60.5 \pm 0.4(\mathrm{Stat}) \oplus 2.4(\mathrm{Syst})$ % and $52.3 \pm 0.4(\mathrm{Stat}) \oplus 2.6(\mathrm{Syst})$ % in central and forward region, respectively. Furthermore, the gluon fractions are extracted in several exclusive regions of transverse momentum.

Kurzzusammenfassung

An Hadronenbeschleunigern ist es schwerlich möglich, Datensätze zu generieren, welche sich jeweils ausschließlich aus quark- sowie gluonen-basierten Teilchenjets zusammensetzen; entsprechend schwierig gestaltet sich das Studium spezifischer Eigenschaften dieser fundamentalen Elementarteilchen an besagten Apparaturen. Die im Rahmen dieser Dissertation dargebotene Analyse ist ein Beitrag, die vorgenannte Problematik bezüglich der unbekannten Komposition hinsichtlich Quarks und Gluonen zu adressieren, um deren prozentuale Anteile in einem Ensemble von hadronischen Jets zu bestimmen.

Die vorliegenden Arbeit präsentiert eine Messung des dreifach differentiellen Wirkungsquerschnitts der relativen Rate von Kollisions-Ereignissen mit *mindestens* zweier Teilchenjets (Multijets), basierend auf insgesamt $139\,\mathrm{fb}^{-1}$ Proton-Proton Kollisionen, welche mit dem ATLAS-Detektor am Großen Hadronen-Speicherring am CERN im Zeitraum von 2015 bis einschließlich 2018 mit eine Schwerpunktsenergie von $13\,\mathrm{TeV}$ aufgezeichnet wurden. Der Produktions-Wirkungsquerschnitt wird in Abhängigkeit des transversalen Impulses p_T, zwei Kategorien der relativen Jet-(Pseudo-)Rapitität η_rel sowie einer topologischen Jet-SubStruktur-Variablen (JSS) $\mathcal{O}_\mathrm{JSS}$, welche sensitiv bezüglich des hadronischen Endzustandes in Gluon- und Quark-Schauern ist, für die beiden hochenergetischsten Teilchenjets der Kollision mit $250\,\mathrm{GeV} < p_\mathrm{T} < 4.5\,\mathrm{TeV}$ und $|\eta| < 2.1$ gemessen.

Die JSS-Variablen, die im Rahmen dieser Arbeit gemessen werden, lassen sich grob in zwei Kategorien unterteilen: Ein Satz von Struktur-Observablen ergibt sich aus dem iterativen „Zählen“ der Emissionen eines mittels des Soft-Drop-Algorithmus' rekonstruierten beziehungsweise rekombinierten Teilchenjets; die zweite Gruppe bildet die Ausgabe zweier Neuronaler Netzwerke, welche allesamt auf der „Deep-Set-Architecktur“ basieren und derart trainiert wird, um zwischen quark- und gluonen-initialisierten Jets zu diskriminieren.

Alle JSS-Observablen werden auf Grundlage innerer Detektorspuren mit einem Transversalimpuls von $p_\mathrm{T} > 500\,\mathrm{MeV}$ sowie $|\eta| < 2.5$ gemessen, um eine ausrechende Korrelation zwischen Detektor- und Teilchenebene für den nachfolgenden Entfaltungsprozess zu gewährleisten. Die gemessenen Spektren werden vollständig entfaltet und für Detektor- sowie Akzeptanzeffekte korrigiert und die finalen Wirkungsquerschnitte mit den Vorhersagen einiger moderner, zeitgenössischer Monte-Carlo-Generatoren verglichen. Zusätzlich werden eine Vielzahl an zugehörigen systematischen und statistischen Messunsicherheiten bezüglich des relativen Produktions-Wirkungsquerschnitts bestimmt. Resümierend lässt sich sagen, dass die Unsicherheit des Wirkungsquerschnittes zwischen 5 % und 20 % je nach Phasenraumelement variiert.

Der gemessene (differentielle) Wirkungsquerschnitt dient als Grundlage, die zugrunde liegenden Anteile an quark- und gluonen-initialisierten Jets sowie deren Wahrscheinlichkeitsverteilungen auf eine rein datengestützte, modellunabhängige Art und Weise zu ermitteln. Dabei wird auf eine adaptiertes Verfahren namens „Jet-Topics“ zurückgegriffen, welches ursprünglich im Rahmen der Computerwissenschaften zur Extrahierung thematischer Inhalte in einem umfangreichen Korpus von textbasierten Dokumenten entwickelt wurde. Die gewonnenen Anteile an Quarks und Gluonen sowie die dazugehörigen Wahrscheinlichkeitsverteilungen sind vollständig *modellunabhängig* und basieren auf einer *operationalen* Definition von Quark- und Gluonen-Jets, die quark- und gluonen-bezogene Charakteristika auf Ebene des rekonstruierten Wirkungsquerschnitts anstatt auf der Stufe individueller Teilchenjets definiert.

Der prozentuale Anteil an mit Gluonen kompatiblen Jets liegt bei $60.5 \pm 0.4(\mathrm{Stat}) \oplus 2.4(\mathrm{Syst})$ % sowie $52.3 \pm 0.4(\mathrm{Stat}) \oplus 2.6(\mathrm{Syst})$ % in den jeweiligen (Pseudo-)Rapiditäts-Kategorien. Des Weiteren werden die Anteile an Gluonen in mehreren disjunkten Intervallen des Transversal-Impulses bestimmt.

Preface

In the context of this dissertation project – which lasted over a period of three-odd years –, the author of this thesis (Christof Sauer) was involved in several activities beyond the central theme (multijet cross section measurement and quark/gluon fraction extraction with the ATLAS experiment at the LHC) like, for instance, heavy flavor jet (top) tagging based on ATLAS' (large-R) Unified Flow Objects (UFOs); anomaly searches/tagging ("one-class classification") based on the latent space learned by (variational) autoencoders; as well as possible applications of "jet topics" in the context of an α_{S}-extraction based on a measurement of a JSS observables that are sensitive to quark- and gluon-initialized jets. The above-mentioned involvements within the context of this PhD resulted in the two publications listed below.

[1] B. M. Dillon, T. Plehn, C. Sauer, and P. Sorrenson. Better Latent Spaces for Better Autoencoders. *SciPost Phys.*, 11:061, 2021.

Abstract

Autoencoders as tools behind anomaly searches at the LHC have the structural problem that they only work in one direction, extracting jets with higher complexity but not the other way around. To address this, we derive classifiers from the latent space of (variational) autoencoders, specifically in Gaussian mixture and Dirichlet latent spaces. In particular, the Dirichlet setup solves the problem and improves both the performance and the interpretability of the networks.

[2] M. LeBlanc, B. Nachman, and C. Sauer. Going off topics to demix quark and gluon jets in α_S extractions. *JHEP*, 02:150, 2023.

Abstract

Quantum chromodynamics is the theory of the strong interaction between quarks and gluons; the coupling strength of the interaction, α_{S}, is the least precisely-known of all interactions in nature. An extraction of the strong coupling from the radiation pattern within jets would provide a complementary approach to conventional extractions from jet production rates and hadronic event shapes, and would be a key achievement of jet substructure at the Large Hadron Collider (LHC). Presently, the relative fraction of quark and gluon jets in a sample is the limiting factor in such extractions, as this fraction is degenerate with the value of α_{S} for the most well-understood observables. To overcome this limitation, we apply recently proposed techniques to statistically demix multiple mixtures of jets and obtain purified quark and gluon distributions based on an operational definition. We illustrate that studying quark and gluon jet substructure separately can significantly improve the sensitivity of such extractions of the strong coupling. We also discuss how using machine learning techniques or infrared- and collinear-unsafe information can improve the demixing performance without the loss of theoretical control. While theoretical research is required to connect the extract topics with the quark and gluon objects in cross section calculations, our study illustrates the potential of demixing to reduce the dominant uncertainty for the α_{S} extraction from jet substructure at the LHC.

»Versuche es, geneigter Leser!, in dem feenhaften Reiche voll herrlicher Wunder, die höchste Wonne wie das tiefste Entsetzen in gewaltigen Schlägen hervorrufen [...] – ja! in diesem Reiche, das uns der Geist so oft, wenigstens im Traume aufschließt, versuche es, geneigter Leser!, die bekannten Gestalten, wie sie täglich, wie man zu sagen pflegt im gemeinen Leben, um dich herwandeln, wieder zu erkennen.«

– E. T. A. Hoffmann (1776 – 1822), *»Der goldne Topf«* (1814)

Contents

List of Figures

List of Plots

List of Tables

Introduction

The subject of (high-energy) elementary particle physics aims to fathom the most *fundamental* entities of material being that make up the (in a figurative sense) "tangible" proportion of the surrounding (physical) reality. To what extent this reductionist approach may lead to *true* insights in a holistic world is another matter[3] ... The author of this thesis fully agrees to Arthur Schoppenhauer's opinion according to which *"[p]hysics is unable to stand on its own feet, but needs a metaphysics on which to support itself, whatever fine airs it may assume towards the latter"* – Arthur Schopenhauer, "*The World as Will and Representation*" (orig. title: "Die Welt als Wille und Vorstellung"), trans. E. F. J. Byrne (1958), Vol. 2, 172[4]. Regardless of whether this claim is foolish or not[5], the research area of elementary particle physics has undoubtedly made a significant and precious contribution to the general understanding of surrounding *material* reality that surrounds us and led to valuable, exciting insights into the underlying laws that conduct nature.

Through the years, a huge variety of methods have been developed to gain knowledge and insight of the elementary particle content of the universe as well as the fundamental forces of nature that govern their interactions. One method, which has proven itself very successful throughout modern history, is the analysis of the scattering products produced in particle collisions at high energies. Strictly speaking, this is the logical continuation of microscopes after they reached their physical limitations. This well-tried method is employed, for instance, at the Large Hadron Collider (LHC) in gigantic dimensions at the European Organization for Nuclear Research, in general better known as CERN (French: *Conseil Européen pour la Recherche Nucléaire*). At the LHC, protons are accelerated to unprecedented energies and their collision products are studied in dedicated measuring apparatuses (like, e.g., the ATLAS detector) to gain insights into the fundamental mechanism of nature.

One of the four *known* fundamental forces of nature is the strong interactions that is described by the (gauge) field theory of Quantum Chromo(Greek: χρῶμα, "color")Dynamics (QCD). The strong (nuclear) force is responsible for the confinement of the fundamental quanta of the elementary fields in QCD, i.e., quarks and gluons, to color-single hadron states; furthermore, the strong interaction is the underlying force that binds neutrons and protons into atomic nuclei. The strength of the interaction in QCD is described by the strong coupling "constant" α_{S}, which is the least precisely-known interaction in nature in terms of its uncertainties. The coupling constant actually depends on the energy scale of the interaction.

[3]After some years in physics, I more and more arrive at the (personal) conclusion that the reductionist approach to gain understanding of "the world" may lead to a continuously increased precision (in a quantitative sense) and is indeed beneficial in terms of the unbundling of "cause-and-effect relations" (principle/axiom of causality) that can be translated into technical artifacts, *but* – inevitably! – results in a drain of ontological meaning of the gained "insights". While getting more precise in the calculation of physical quantities (physicists are, for instance, very proud of measuring an agreement to more than ten significant figures with respect to calculations of the magnetic moment of the electron), the ontological correspondence is gradually lost – "what does this actually mean?" It appears to me that precision and ontological meaning are contrasting each other and move into opposite directions. After all, the term "precision" (from the Latin word *praecisionem*) means "cutting off", "abstraction" – literally: "cutting of [reality]". Consequently, a reductionist's approach – I'm convinced – will *never* result in a "true" insight of the world. Like the great Russian novelist Nikolai Vasilyevich Gogol (1809–1852) wrote in his famous, unfinished novel "Dead Souls" (spoken by the philosopher Kiefa Makejitsch): "[...] 'Je mehr du dich in die Natur versenkst, desto weniger begreifst du sie' [...]" (Gogol, "Die toten Seelen" oder "Tschitschikows Abenteuer", 1842 (1. Teil) beziehungsweise 1855 (2. Teil, Fragment); Herausgeber: Diogenes (1. Januar 1977), Übersetzung von Philipp Löbenstein 1846, S. 352). To cite a beautiful verse from Novalis' *"Von der geheimen Welt"*: *"[...] Unendliche Ferne der Blumenwelt [...]"*. Indeed, no matter how far humans will disassemble the world (of flowers), the thing-in-itself (Kant) will remain unreachable – at least in mind-based relationship between a *subject* and an *object* under investigation.

[4]This is by no means a degradation of this wonderful research area – on the contrary: only taking its epistemological limitations into account, (natural) science can be fully appreciated.

[5]Surrounded by a reality in which the inhuman philosophy of so-called "transhumanism" spreads like the plague, it is important to remember that (at least according to the author's opinion) the reducionist approach of today's natural science will not provide an answer to *non* of the questions that are actually important to human existence (first formulated by Immanuel Kant in the "Critique of Pure Reason" (KrV, A 805/B 833)): *(1) What can I know? (2) What ought I to do? (3) What may I hope for? [(4) What is man?]*

This so-called "running of the coupling" encodes the underlying dynamics of hadron physics, which results in (color) confinement at large distances and asymptotic freedom at high energies, i.e., short distances.

Protons are composite objects that are composed of quarks and gluons, which are commonly referred to as *partons*. Those partons are the particles that actually undergo interactions in proton-proton collisions. The internal structure of the proton (as well as other hadrons) is described by the Parton Distribution Functions (PDFs) that describe the probability of finding a parton of some flavor[6] at some energy scale Q with a certain fraction of the proton's momentum x. The evolution of the PDFs as a function of the energy scale Q is predicted by means of perturbative QCD – in contrast, the dependence on the momentum faction x can *not* be calculated from first principles in QCD directly and therefore must be extracted from experimental data. The PDFs of all flavors are well constrained in the low- and medium-x regime, mainly done in deep inelastic scattering experiments. In the case of high momentum fractions, however, the uncertainty on the gluon's (as well as for the quark's) PDF is still "comparatively" large.

The PDF of the various quark flavors as well as the gluon are related to the fractions of the respective parton in the final state of the event. Those fractions play an important role in the extraction of the strong coupling from the radiation pattern within jets by means of a fit to a jet substructure observable with a known cross section. The main challenge in this extraction of the strong coupling constant from the measurement of a substructure observable at hadron colliders like the LHC is the *degeneracy* of quark- and gluon-initialized jet, that is, the fraction of those particles is generally unknown as one always deals with a *mixture* of jets instead of "pure" regions in phase space given a certain type of jet. This situation makes it very challenging to study quark/gluon-related properties at hadron colliders, like the LHC, in general.

In principle, the relative fraction of quark and gluon jets can be calculated using the aforementioned PDFs and propagating the probability of the respective parton flavor throughout the calculation of the underlying hard interaction. This approach, however, is sensitive to the non-negligible uncertainties on the PDFs as well as theoretical uncertainties due to limited precision of the calculation of the underlying event and variations of the renormalization and factorization scale, which results in large uncertainties on the calculated quark/gluon fractions. Alternatively, the quark/gluon fractions can be extracted from data by means of a cross section measurement. The latter approach suffers from the already mentioned degeneracy of quarks and gluons at hadron colliders because neither the fractions nor the distributions are known. Usually, this problem is addressed by extracting the underlying fractions from simulations by means of some state-of-the-art Monte Carlo event generators or by generating templates of the cross section for quark and gluon jets[7]. This approach works well and breaks the degeneracy; however, it poses the problem of resulting in model-dependent fractions of unknown uncertainty (model-dependent because using a different generator will yield different results for the extracted quark/gluon templates/fractions), which is particular important for the aforementioned (fit-based) extraction of α_{S}. Furthermore, the fractions derived from simulation rely on a jet-by-jet assignment of the jet's flavor. Despite being commonly treated as well-defined objects, there is *no* precise definition of quark and gluon jets to the present day. This complication arises because higher-order contributions (at Born-level, the jet flavor can unambiguously be defined) to parton-level quarks and gluons may result in identical, i.e., indistinguishable final state configurations, which – according to quantum mechanics – results

[6] Flavor is a *quantum number* that corresponds to the "type" of an *elementary* particle. In QCD, there are six (known) flavors of quarks.

[7] This important aspect will be dealt with later in adequate detail, but for the sake of understanding, consider the following decomposition:

$$P_{M_\alpha}(x) = f_q P^q_{M_\alpha}(x) + f_g P^g_{M_\alpha}(x),$$

whereby $P_{M_\alpha}(x)$ denotes the *measured* distribution with respect to some jet substructure feature space 'x' over two Mixtures M_α of jets, e.g., $\alpha \in \{\text{central, forward}\}$ or $\alpha \in \{\text{multijet, Z+jets}\}$. The assumption is that the measured spectra are a linear composition of contributions from quarks and gluons, whereby the weights $f_{q/g}|_{M_\alpha}$ are given by the fractions over the jet mixtures M_α. This system of equations is usually degenerate.

in a superposition of states. The reconstructed jets therefore can *not* uniquely be assigned to a quark or gluon anymore on a *jet-on-jet basis*.

The analysis presented in this report pursues an alternative approach to the extraction of quark- and gluon-related jet features based on a cross section measurement that does not suffer from the above-mentioned issue of imposing model-dependence on the final fractions or distributions, but tries to solve the degeneracy of jet samples at hadron colliders in a solely data-driven, model-independent manner. This is done by using an operational definition of quark- and gluon-*like* jets (that treats the jet-flavor rather as an *adjective* than a noun) as well as a statistical demixing procedure called "jet topics" [29], which is based on probabilistic topic modeling [48] that has originally been developed as a tool for extracting emergent themes in an extensive corpus of text-based documents. Generally, the method assumes the measured cross section of, e.g., multijet or Z+jets events, to be a mixture of contributions from quarks and gluons, independent of their precise definition. The cross section is measured as a function of some jet substructure observables that provides sensitivity to the quark- or gluon-like nature of the hadronic shower, as well as in several regions of transverse momentum with the objective to extract the fraction of quark- and gluon-like jets in several p_{T} bins. Furthermore, the unmixing procedure requires two orthogonal samples that are enriched by either of quarks and gluons, respectively, to extract the underlying features associated with the two classes. For this purpose, the cross section is measured in two disjoint categories of different regions, *central* and *forward*, that are defined by the relative orientation of the two leading-p_{T} jets in the event. Generally, the jet being reconstructed with the smaller *absolute* pseudorapidity $|\eta|$ is assigned to the central rapidity category, while the other jets populate all distributions in the forward region[8]. In total, the data-driven extraction of quark and gluon fractions is based on the measurement of the (inclusive) triple-differential multijet (dijet) cross section as a function of the transverse momentum p_{T}, the relative pseudorapidity η_{rel}, as well as a jet substructure observable $\mathcal{O}_{\mathrm{JSS}}$ that acts as a quark-gluon discriminator according to

$$\frac{1}{\sigma_{\mathrm{fid}}}\frac{\mathrm{d}^3\sigma}{\mathrm{d}p_{\mathrm{T}}\,\mathrm{d}\eta_{\mathrm{rel}}\,\mathrm{d}\mathcal{O}_{\mathrm{JSS}}} = \frac{1}{N_{\mathrm{tot}}}\frac{N^{\mathrm{multijet}}}{\Delta p_{\mathrm{T}}\,\Delta\eta_{\mathrm{rel}}\,\Delta\mathcal{O}_{\mathrm{JSS}}}, \tag{1}$$

whereby N^{multijet} denotes the number of (*inclusive* dijet) multijet events in a given phase space element $\Delta p_{\mathrm{T}}\,\Delta\eta_{\mathrm{rel}}\,\Delta\mathcal{O}_{\mathrm{JSS}}$. The measurement is inclusive because all events with *at least* two jets that pass the event selection contribute to the cross section measurement, but only the two leading-p_{T} jets being used to reconstruct the distributions. The relative event rate (normalized cross section) is measured because systematic uncertainties between nominator and denominator cancel to the degree of their correlation. For instance, the uncertainty related to the Luminosity drops out completely in this measurement. The cross section is measured from jets with $250\,\mathrm{GeV} < p_{\mathrm{T}} < 4.5\,\mathrm{TeV}$ and $|\eta| < 2.1$ from an integrated luminosity of $139\,\mathrm{fb}^{-1}$ of proton-proton collisions at $\sqrt{s} = 13\,\mathrm{TeV}$ that was reconstructed with the ATLAS detector during Run 2 (2015–2018).

One challenge of this measurement is the unfolding of the measured spectra to correct for detector and acceptance effects due to the jets substructure observable. For the unfolding of JSS observable to work, sufficient correlations between detector- and particle-level objects must be maintained in order to obtain sufficiently diagonal response matrices. For this reason, the substructure variable is measured based on Inner Detector ($|\eta| < 2.5$) tracks with $p_{\mathrm{T}} > 500\,\mathrm{MeV}$ that have been matched to the reconstructed small-$R(= 0.4)$ anti-k_t Particle Flow (PFlow) jets – the topo-cluster or PFlow-based jets can *not* be used to calculate the JSS variable as the resolution is far too low to maintain the necessary correlations; single-particle resolution is needed. This can easily be understood since the (comparatively) coarse granularity of the calorimeter (compared with the tracking-part of the detector at least) often results in several particles being reconstructed/absorbed in the very same cell of the calorimeter, which reduces resolution.

The unmixing procedure described above (unmix the contributions from quarks and gluons

[8]The pseudorapidity has some sensitivity to differences between quark and gluon jets, because the forward particles are correlated with the longitudinal momentum fraction, which for larger momenta is more likely to correspond to a quark due to the structure of the underlying PDF of the proton.

to the measured cross section in central and forward region) is still underconstrained; in order for the extraction to result in unique underlying quark/gluon probability distributions and fractions, the utilized jet substructure observable must fulfill an additional (mathematical) property, i.e., it must results in regions in phase space that are (almost) pure in quark and gluon jets. In more technical terms, the underlying (probability) distributions of the JSS observables must be *mutually irreducible* (loosely speaking: not a mixture by itself). The jet shapes used in this analysis have been selected in light of this property as well as their sensitivity to differences between quark and gluons. Broadly speaking, two classes of observables have been studied that can be subsumed as being *counting*- or *NN*-based.

In the first category, the cross section was measured with respect to the number of tracks n_{trk} (which is a proxy for the number of charged constituents) in the jet, as well as the Soft Drop (SD) multiplicity n_{SD}, with the latter one being constructed by iteratively declustering and counting charged jet constituents that pass the SD condition, which constraints the emissions' phase space. Those observables are "simple" regarding their interpretation, yet sufficiently powerful to not be limited by Casimir scaling and hence being mutual irreducible.

The second class of observables is based on the output provided by a Deep Neural Network (DNN) that has been trained as a quark-gluon classifier. The models are based on the "Deep Sets" architecture that allows to constrain the set of possible functions learned by the DNN to permutation invariant maps over sets. This parameterization is tempting because it natively respects symmetries associated with jets. The resulting quark/gluon discriminators are very powerful, however, the decision-making process is hard to keep track of. Both categories therefore provide a balance between performance (in terms of classification) as well as interpretability of the result. Besides that, extracting the quark/gluon fractions for several measurements with respect to various observables provides a valuable cross check. Of course, the individual measurements of the differential multijet cross section are not independent in the sense of not being correlated – on the contrary: the measurements presented in this report are (partially highly) correlated as they are measured over the exact same set of jets.

To summarize, this thesis reports the measurement of the inclusive triple-differential dijet cross section as a function of several jet substructure observables. The reconstructed spectra are fully corrected for acceptance as well as detector effects and are unfolded to hadron level and their degree of agreement compared to several state-of-the-art Monte Carlo event generators. The uncertainty of the measurement is determined by taking several sources of systematic and statistical uncertainties into account and fully propagating them through the unfolding procedure onto the final cross section. After the (relative) cross section measurement has been concluded, the distributions are unmixed to obtain the quark- as well as gluon-like features and the corresponding quark/gluon fractions in each bin in transverse momentum using the "jet topics" procedure outlined above. The uncertainties on the relative event rate are propagated onto the extracted fractions and underlying quark/gluon probability distributions.

This thesis can broadly be divided into three parts. The first part ("*fundamentals*") introduces all the theoretical background needed to fully appreciate the physical motivation of the studies presented in this report. The second part of this thesis ("*methodology*") focuses on the more technical aspects as well as the different methods that have been used throughout this analysis. Finally, the third part ("*Analysis and Results*") is dedicated to all steps relevant for the actual analysis as well as the final results and conclusions.

In a bit more detail, this thesis is structured as follows: the very first Chapter 1 introduces the necessary theoretical framework including an introduction into the Standard Model of elementary particle physics, a quick review of (perturbative) QCD as well as some phenomenological aspects of QCD by means of jet physics. The next Chapter 2 builds on the previous one, but turns the attention to the simulation and event generation using Monte Carlo methods. The individual steps of an event generation chain in high-energy physics like, for instance, the calculation of the hard sub-process by means of the matrix element, the simulation of additional radiation using the parton shower approach, as well as the formation of color-singlet states in the hadronization step and the subsequent decay to final state particles.

The next Chapter 3 opens the methodology part and introduces the experimental setup that was used to produce the data that serves as the basis for the analysis at hand. The LHC at CERN as well as the ATLAS detector (including its components) are explained in appropriate

detail. Chapter 4 introduces the reconstruction of the physical objects from signals recorded with the ATLAS detector with a main focus on the reconstruction and calibration of hadronic jets. Once the physical objects are defined, the concept of detector unfolding is explained in Chapter 5. Chapter 6 jumps to a different topic (careful: the word "topic" is quite overloaded in this report as one might already guess). As previously mentioned, the triple-differential multijet cross section is measured with respect to the score predicted by a neural network. Accordingly, Chapter 6 provides a brief introduction into the broad area of machine learning with a focus on DNNs as well as a basic explanation of the "Deep Set" architecture, which is used for all models in this analysis. Chapter 7 explains all necessary background to understand the aforementioned data-driven feature extraction using topic modeling. This includes a quick review of some common pitfalls when talking about quark and gluon jets in general as well as an introduction into the operation definition of those objects.

The next Chapter 8 opens the third and last part of the thesis and introduces all samples in real pp-collision data and simulation that have been used in this analysis. In the subsequent Chapter 9, the reconstruction of the physical objects (mainly jets and tracks) that enter the measurement as well as the selection of events is discussed; furthermore, the trigger chain for the selection of analysis-relevant events is defined. After that, the measurement of the triple-differential cross section can – in principle – begin; however, one rather technical Chapter 10 is inserted that introduces all the sources of statistical and systematic uncertainties along with their calculation. After that, Chapter 11 finally turns the attention to the measurement of the triple-differential cross section with respect to the counting-based observables. This includes the reconstructed spectra at detector level, the unfolding of the cross section as well as the final, unfolded spectra including uncertainties. Chapter 12 basically mimics the previous one but for the NN-based observables. In Chapter 13, the unfolded cross section is used to extract the underlying quark/gluon probability distributions as well as their corresponding fractions using topic modeling. This Thesis ends with a conclusion in Chapter 14 that includes a discussion as well as some further research suggestions that could follow and build upon this analysis.

Part I

Fundamentals

The first part of this thesis establishes the theoretical foundations necessary to provide a profound motivation of the analysis presented in the document at hand. This minor excursion into the fundamentals of the broad field of elementary particle physics is mostly limited to aspects that are considered relevant for the studies carried out in the scope of this dissertation. Due to the sheer size of this matured branch of physics, this introduction barely scratches the surface; hence, the reader is assumed to be equip with a basic understanding of this discipline.

The first Chapter 1 of this introductory part provides a brief recap of the Standard Model of elementary particle physics including its historical development as well as its most important concepts. The main focus of this discussion is on the theory of the strong interaction and its phenomenology, which is the field of research this study is part of.

The second Chapter 2 builds on the lessons from the first one by providing an introduction into the topic of event generation (and simulation) in high-energy particle physics. It therefore builds a bridge between predictions in theory and observations in experiments.

1. Theoretical Framework[1]

The Standard Model (SM) is the fundamental theoretical (QFT-based) model of *elementary* particle physics based on an underlying, spontaneously broken $\mathrm{SU}(3)_c \times \mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ symmetry group that encapsulates the gathered knowledge of the fundamental (according to present knowledge), i.e., irreducible constituents of matter in a fertile interplay between theory and experiment. It provides an *astonishingly* precise description of the elementary particles known to mankind (see, e.g., Ref. [50]) and their mutual interactions that are governed by three out of four known fundamental forces occurring in nature: the *electromagnetic*, the *weak*, and the *strong force* – with the gravitational interaction withstanding a consistent quantum-field-theoretical description and, therefore, not being incorporated into the theory so far. The tumultuous history of the Standard Model (a name that has been given in the early '70s) is rich and goes back to the origin of modern particle physics at the beginning of the previous century. As the hour of birth, one could take the unification of the electromagnetic and the weak force by Sheldon Glashow, Abdus Salam, and Steven Weinberg in the '60s [51, 52, 53]. Ever since, the Standard Model has been tested in numerous experiments around the world, providing an accurate description of the microscopic world of elementary particles; making it the most accurate scientific theory known to humankind to the present day. This culminated in the successful discovery of the Higgs boson at the Large Hadron Collider (LHC) near the French-Swiss border in 2012 [54]; finally, making the Standard Model a self-consistent (albeit inherently incomplete [55]) theory of nature.

The purpose of this very first chapter is to serve as a general introduction into the topic area of the Standard Model of particle physics. A comprehensive introduction into the subject is, of course, beyond the scope of this work; hence, the focus lies on the fundamental aspects of the theory as well as some of the milestones in its history.

The first section (1.1) gives an outline of the Standard Model's history (1.1.1) as well as its elementary particle content. This is followed by a very brief introduction into the fundamental concepts behind Quantum Field Theory (QFT). Equipped with the necessary foundations, the Lagrangian of the Standard Model along with its individual terms is introduced with a focus on the strong sector of the theory (1.1.3). The subsequent section is dedicated to the domain of perturbative Quantum Chromodynamics (pQCD) and key concepts like the running of the strong coupling constant (1.2.1) and factorization theorems in QCD (1.2.2). This chapter closes with a careful look at some phenomenological aspects of QCD by means of the important concept of a particle jet (1.3.1, 1.3.2), which will be further developed later in this report.

1.1 The Standard Model of Particle Physics

This section roughly sketches the historical development of the Standard Model with the objective to embed the theoretical concepts, which are introduced in the following section, into its relative historical context and background that are relevant for the results presented in this analysis. Furthermore, it introduces the known elementary particles and their associated properties that are described by the Standard Model.

[1] Essential parts of the succeeding text passages have been adopted (almost) literally from my Master's thesis "*Towards a Data-Driven Simulation of QCD Radiation with Generative Models utilizing Machine Learning Methods*", 2019, Universität Heidelberg, Christof Sauer [49].

1.1.1 History of the Standard Model

He (the author speaks based on his very own experience) who encounters the state-of-the-art Standard Model of particle physics for the very first time might be dazzled and overwhelmed by its complexity and the diverse (partially counterintuitive) physical phenomena it is able to accurately describe. However, to understand and appreciate this highly advanced theory of nature, it is vital to reflect its historical development over the last decades, starting from the '30s of the previous century until today. Therefore, it is even more surprising that the history of the Standard Model is virtually never part of its introduction.

The Standard Model's history is an adventurous story full of misconceptions and embroilments, all of which represent steps on the path to higher knowledge[2]. It is a beautiful story; but unfortunately, too long to tell. Within the context of this thesis, only an incomplete overview of the milestones in the history of the Standard Model – according to the author's personal view! – is given.

It is quite "simple" to set a starting point for the historical development of the Standard Model. At the beginning of the '20th century, the world was more or less *classical* and described by classical fields as in Maxwell's laws of electromagnetism [56] and Einstein's theory of gravitation [57] that replaced Netwon's theory of gravity[3] [58]. Both theories at that time accurately described the two fundamental forces that are daily experienced by human beings in their meso- and macroscopic world – and one of them still does to this very day. This was the situation roughly up to the year 1930, when new discoveries and insights had changed the fundamental understanding of nature, with an enhanced understanding of Quantum Mechanics (QM) leading the charge. With the discovery of the neutron [59] – which was initially mistakenly assumed to be γ-radiation –, the postulation of weak interaction by Fermi to solve the puzzle of the continuous energy spectrum of the electron emitted in the beta-decay in his revolutionary essay *"[t]entativo di una teoria dellaemissione di raggi β"* [60] ("tentative theory of beta-decay") as well as the invention of Quantum Electrodynamics (QED) [61] by quantizing Maxwell's equations, the foundations of the Standard Model was laid. With the procedure of renormalization – invented in 1947 –, QED agreed with astonishing accuracy with the experiments (see, e.g., measurement of the anomalous magnetic dipole moment) and quickly became the most well-tested theory in physics. After World War II, when many scientists ended their rendezvous with the military and returned to their actual studies, the number of known particles significantly increased, creating a zoo of hundreds of, apparently, *elementary* particles; it was necessary to make sense out of this mess by trying to distill out the underlying structures. This ordering was partially achieved by the observation of apparent similarities between the different particles (e.g. mass, spin, electric charge etc.) and the discoveries of (approximate) symmetries such as the famous "eight-fold way" [62] to name only one out of many. It was a major "crisis", when it became clear that most symmetries only represent an approximation. The existence of numerous approximate symmetries confronted physicist with a daunting problem of interpretation.

A brilliant idea, which was independently developed by Murray Gell-Mann and Georg Zweig in 1961 [63, 64], gradually lifted the curtain of confusion: the so-called *parton* model pictures hadrons as a collection of point-like, i.e., elementary particles, which today are associated with the elementary quarks and gluons in quantum chromodynamics. It revealed an underlying structure of known (composite) particles and reduced those to a small number of elementary constituents of *material being* – the actual (at least according to our current knowledge) *atomos*. The parton model played a similar role in the development of modern particles physics like Pauli's exclusion principle did for chemistry in explaining the ordering of the elements in the periodic table.

Alongside the aforementioned parton model, the systematic development of gauge symmetry (also known as local gauge symmetry) was yet another important step in the evolution of

[2] Was ist (Natur-)Wissenschaft anderes als eine systematische Form des Sich-empor-Irrens? – Gerhard Vollmer (1943–today)

[3] Strictly speaking, this is not correct as the fundamental principles of quantum mechanics had already been discovered at that time. However, compared to the theory of relativity, it took somewhat more time to get used to the partially counter-intuitive insights.

the Standard Model. Although classical Electrodynamics may also be regarded as a gauge theory that is based on the $U(1)_Q$ symmetry group. The first methodical contribution in this matter came from Chen Ning Yang (楊振寧/杨振宁) and Robert Laurence Mills [65] who constructed a theory that was based on the non-abelian (Lie) group SU(2) with non-commuting "charges". The objective was to build a theory of the strong interaction; however, it was later applied to weak interactions that were already known to be a mixture of vector and axial-vector interactions.

The '50s, as well as the '60s, were eventful times full of revolutionary ideas and discoveries. Another notable one was the concept of intermediate or mediator particles, respectively, vector bosons to be more precise. This idea was independently developed by several scientists like Julian Seymour Schwinger [66], Sheldon Lee Glashow [51] and in joint work between Abdus Salam (عبد السلام) and Ward John Clive [67]. It had a tremendous impact on our understanding and interpretation of the fundamental nature of forces in nature as being transmitted by vector (and possibly tensor) bosons in contrast to the rather "vague" concept of (classical) potentials.

The Standard Model at that time was already able to describe a large range of phenomena; especially QED had proven as the most accurate theory in physics of all time, with the theoretical predicted electromagnetic fine-structure constant agreeing with the measured one within *ten-parts-per-billion* (1ppb= 10^{-9}) [68]. But despite these resounding successes and the improved understanding of the subatomic world, the Standard Model (at that time) had an apparent flaw: the requirement of local gauge invariance culminated in a theory with solely massless particles – which is in clear contradiction to the obvious (partially very) "massive" particles measured in the experiments ... This situation was a highly unsatisfactory since it caused inconsistencies in the Standard Model between predictions and observations – the merciless guillotine in physics that without hesitation beheads even the most elegant theories if they disagree with observations. Full of despair, physicist had to put in masses by hand into the theory; but, in doing so they accepted reduced predictability and predictive power of the theory and, additionally, made the theory non-renormalizable! A "novel" idea finally brought the long-awaited breakthrough with the desired way out of the dilemma: the concept of Spontaneous Symmetry Breaking (SSB), which is the final part of this summary.

The path to SSB was paved with obstacles. The problem was that it was proven by Goldstone, Salam and Weinberg that for each *exact* symmetry that is spontaneously broken there must be a mass- and spinless (scalar- resp. pseudoscalar) particle. At this time, a large number of approximate symmetries have been known, e.g., isospin symmetry or the aforementioned eight-fold way. But, besides the photon, no massless particle had been discovered (the gluon was yet unknown). In the end, one problem was replaced by another. At that time, Peter Ware Higgs [69] arrived on stage. He (and others) tried to find a way out of the Goldstone theorem and its ominous mass- and spin-less particles. He realized that the Goldstone theorem does not apply for local gauge symmetries, which are spontaneously broken. In this case, the Goldstone bosons do not manifest as real particles but remain in the theory and turn into the helicity-zero component of the gauge bosons (see Higgs-Kibble-Dinner). This not only gets rid of the non-observed Goldstone particles but simultaneously results in a mass for the gauge bosons and hence creates a theory of massive mediators (Higgs' idea was contemporaneously also discovered by the collaboration between Englert and Brout [70] and Guralnik, Hagen and Kibble [71] – therefore also the somewhat cumbersome name Englert-Brout-Higgs-Guralnik-Hagen-Kibble mechanism, usually just called Higgs mechanism). With the spontaneous breaking of the symmetry $SU(2)_L \times U(1)_Y$, the Higgs mechanism could also explain the masses of the fundamental fermions in the Standard Model – excluding the neutronos, which are known to be massive –; finally making it a consistent and (even more) predictive theory. Besides explaining the masses of massive vector bosons and fermions, the Higgs mechanism also predicted the presence of a new particle that is the quantum excitation of the respective Higgs field. This particle was predicted to be massive; hence, it could have escaped detection due to limited energies in particle colliding experiments at that time (e.g., at LEP at CERN). This was indeed the case. It took more than 50 years to finally discover the Higgs boson in 2012. A great example of human willpower.

Besides the discovery of the Higgs, the Standard Model has proven its validity and predictive power in a variety of other experiments. For example, the Standard Model predicted the existence of the $W^{\pm}$ and Z bosons in electroweak interactions, with the first observations of neutral currents in 1973 at the Gargamelle bubble chamber at CERN [72, 73] and the discovery of the actual particles back in 1983 that have been observed in 1986 for the very first time at the Super Proton Synchrotron also at CERN [74]. Or, the discovery of the gluon in 1979 with the PLUTO and TASSO experiment at DESY [75, 76, 77]. And, not to forget, the several quark flavors predicted by the Standard Model all that have been observed later in the experiment, like the *charm* and *top* quark [78, 79].

The Higgs mechanism finishes the journey through the history of the Standard Model. Of course, there would be far more to tell. However, it would be a mistake to think that these concepts have matured in an "ivory tower". On the contrary, the evolution of the Standard Model is a perfect example of how the interaction and mutual pollination of theoretical and experimental physics leads to a path of higher knowledge and a better understanding of the world around us [80].

1.1.2 Particle Content

As a quantum field theory, all the fundamental particles in the Standard Model are associated with a certain quantum field, more precisely, the quantum excitation that manifest as fundamental particles of the respective field. Roughly speaking, the elementary particles in the Standard Model can be subdivided into fermions (half-integer spin particles that obey the Fermi-Dirac statistic), vector (tensor) bosons (integer spin particles that mediate the fundamental forces), and one pseudoscalar particle, the Higgs, whose field plays an important role in the essential Higgs mechanism and the Yukawa interaction. The known fundamental fermions in the Standard Model are summarized in Table 1.1.

	Flavour	Mass	Q_{el}	T_3^L	Discovery
Leptons	ν_e	$< 2.2\,\text{eV}$	0	$+\frac{1}{2}$	1956
	e	$0.511 \pm 10^{-8}\,\text{MeV}$	-1	$-\frac{1}{2}$	1897
	ν_μ	$< 0.19\,\text{MeV}$	0	$+\frac{1}{2}$	1962
	μ	$105.7 \pm 4 \times 10^{-6}\,\text{MeV}$	-1	$-\frac{1}{2}$	1936
	ν_τ	$< 18.2\,\text{MeV}$	0	$+\frac{1}{2}$	2000
	τ	$1776.86 \pm 0.12\,\text{MeV}$	-1	$-\frac{1}{2}$	1975
Quarks	u	$2.3 \pm 0.7\,\text{MeV}$	$+\frac{2}{3}$	$+\frac{1}{2}$	1968
	d	$4.8 \pm 0.5\,\text{MeV}$	$-\frac{1}{3}$	$-\frac{1}{2}$	1968
	c	$1.275 \pm 0.025\,\text{GeV}$	$+\frac{2}{3}$	$+\frac{1}{2}$	1974
	s	$95 \pm 5\,\text{MeV}$	$-\frac{1}{3}$	$-\frac{1}{2}$	1968
	t	$173.2 \pm 0.9\,\text{GeV}$	$+\frac{2}{3}$	$+\frac{1}{2}$	1995
	b	$4.18 \pm 0.03\,\text{GeV}$	$-\frac{1}{3}$	$-\frac{1}{2}$	1977

Table 1.1: Fermions (subdivided into leptons and quarks) in the Standard Model (masses have been rounded) [47].

According to Table 1.1, the fermions in the Standard Model may be further subdivided into *leptons* and *quarks*. A lepton is an elementary spin-$^1/_2$ particle that does not experience the strong force. The known elementary leptons are the electron, the muon, the tauon, and the corresponding electrically neutral neutrino flavors with their respective antiparticles. The second category of fermions are the quarks that differ from leptons insofar as they carry a fractional electric charge as well as *color charge*, which is the subject of quantum chromodynamics. Usually, fermions are associated with matter, with quarks being the key building blocks of composite particles (hadrons) such as protons and neutrons. With leptons and quarks, the Standard Model knows twelve different fermion flavors and 24 fermions in total – including

the corresponding antiparticles. Furthermore, a distinction is made between fermions and bosons. Bosons (the name refers to the Indian physicist Satyendra Nath Bose (সত্যেন্দ্রনাথ বসু)) are described by the Bose-Einstein statistics. In the Standard Model, the elementary *vector bosons* ($s = 1$) are responsible for the mediation of the fundamental forces of nature. These elementary vector bosons are the photon γ, the gluon(s) g, as well as the neutral Z and the charged $W^{\pm}$ bosons.

	Boson	Interaction	Mass	Q_{el}
Scalar	higgs H_0	none	$125.18 \pm 0.16\,\text{GeV}$	0
Vector	photon γ	QED	$< 10^{16}\,\text{eV}$	0
	gluons g	QCD	0	0
	Z	EW	$91.1876 \pm 0.0021\,\text{GeV}$	0
	$W^{\pm}$		$80.379 \pm 0.012\,\text{GeV}$	± 1
Tensor	graviton[4] G	gravity	$< 6 \cdot 10^{-32}\,\text{eV}$	0

Table 1.2: Bosons in the Standard Model [47].

It should be mentioned that in a hypothetical quantum field theory of gravity there would be an elementary tensor boson with spin-2, the graviton, that is the "force-carrier" of the gravitational force. However, to be a consistent theory, the Standard Model requires the existence of (at least) one scalar particle ($s = 0$) that is associated with the mechanism that generates the mass of the fermions and bosons. This is the aforementioned Higgs boson. The bosons in the Standard Model are summarized in the Table 1.2.

The appearance of gauge bosons in the theory is a consequence of the required local gauge symmetry of the Standard Model Lagrangian. This concept is important in order to understand how elementary particles interact with each other through the exchange of force-carrying particles, which are quantum fields by themselves.

1.1.3 The QCD Lagrangian

The Standard Model of particle physics is a gauge-invariant quantum field theory that is based on the (spontaneously broken) unitary, non-abelian symmetry group[5]

$$G_{\text{SM}} := \text{SU}(3)_{\text{c}} \times \text{SU}(2)_{\text{L}} \times \text{U}(1)_{\text{Y}}, \tag{1.1}$$

whereby each of the compact (Lie) subgroups $\text{SU}(3)_{\text{c}}$, $\text{SU}(2)_{\text{L}}$ and $\text{U}(1)_{\text{Y}}$ introduces associated *gauge fields* that defines the nature of the interactions allowed in the theory. The gauge fields in the Standard Model ensures the underlying Lagrangian to be invariant under local gauge transformations concerning the group G_{SM}. The requirement of a local gauge symmetry is highly non-trivial and comes with a long history. First, it was introduced in analogy to QED and has then later been adopted to other theories. However, today there exists a deeper understanding of gauge transformations as the connection between equivalent coordinate bases for the same mathematical object (i.e., the field).

With the underlying group structure of the Standard Model introduced, it is about time to present the actual Lagrangian of the theory that encodes the dynamic of the fundamental quantum fields. The Lagrangian (density) $\mathcal{L}_{\text{SM}}$ of the Standard Model (reduced and in simplified

[4] Hypothetical elementary particle that mediates the force of gravity.

[5] The actual gauge group of the Standard Model is the reduced (quotient) $\text{SU}(3)_{\text{c}} \times \text{SU}(2)_{\text{L}} \times \text{U}(1)_{\text{Y}} \,/\, \mathbb{Z}_6$ symmetry group [81].

notation that fits on a coffee cup) in a legible form is given by:

$$\mathcal{L}_{\mathrm{SM}} = -\frac{1}{4}\mathrm{tr}\left[B_{\mu\nu}B^{\mu\nu}\right] - \frac{1}{4}\mathrm{tr}\left[W_{\mu\nu}W^{\mu\nu}\right] \tag{1.2}$$

$$+ \bar{\Psi}_q \gamma^\mu D_\mu \Psi_q - -\frac{1}{4}\mathrm{tr}\left[G_{\mu\nu}G^{\mu\nu}\right] \tag{1.3}$$

$$+ (D_\mu \Phi)^\dagger (D^\mu \Phi) + \mu^2 \Phi^\dagger \Phi - \frac{1}{2}\lambda(\Phi^\dagger \Phi)^2 \tag{1.4}$$

$$+ \bar{\Psi}_L \gamma^\mu D_\mu \Psi_L + \frac{1}{2}\Psi_L^T C \Phi H \Psi_L + h.c, \tag{1.5}$$

whereby the term "h.c." refers to the "hermitian conjugate", which is necessary in order to obtain observables that are free of non-physical "ghost" fields artifacts. Phenomena surrounding the strong force are described by Equation 1.3 in the expression above.

The strong force was (first phenomenologically) introduced in the 1930s to explain the binding of the constituents that form the atomic nuclei. Today, quantum chromodynamics (from Greek χρῶμα, "color") is the theory of the strong interaction with the underlying *non-abelian* symmetry group $\mathrm{SU(3)_c}$, whereby the subscript 'c' refers to the so-called *color charge*, which remains unbroken in the Standard Model. In the Lagrangian of the Standard Model according to Equation 1.5, QCD manifests itself by the term

$$\mathcal{L}_{\mathrm{QCD}} = \sum_q \bar{\Psi}_{q,a}(\gamma^\mu \partial_\mu \delta_{ab} - i g_3 \gamma^\mu G^\alpha_\mu t_\alpha)\Psi_{q,b} - \sum_q \bar{\Psi}_{q,a}\Psi_{q,b} m_q \delta_{ab} - \frac{1}{4}\mathrm{tr}\left[G_{\mu\nu}G^{\mu\nu}\right], \tag{1.6}$$

with t_α being related to the Hermitian and traceless Gell-Mann matrices (which are generators of the $\mathrm{SU(3)_c}$ group) and $\alpha \in \{1, ..., 8\}$ that correspond to eight gluons (fields) with different linear independent color charge combinations. The field strength tensor $G^a_{\mu\nu}$ is given by $G^a_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu - g_3 f_{abc} G^b_\mu G^c_\nu$, which describes the dynamic of the gluon fields.

The quarks acquire their mass via SSB of the subgroup $\mathrm{SU(2)_L} \times \mathrm{U(1)_Y} \rightarrow \mathrm{U(1)}_Q$ that also acts on the quark doublets Ψ_q. The gluons, which are considered to be in the adjoint representation of the gauge group and an octet under $\mathrm{SU(3)_c}$, remain massless though (which agrees with experimental observations) since the symmetry group $\mathrm{SU(3)_c}$ is unbroken in the Standard Model.

Quantum chromodynamics is a non-abelian gauge theory, i.e., the generators t_α of the underlying symmetry group $\mathrm{SU(3)_c}$ do *not* commute with each other $[t^a, t^b] = i f_{abc} t^c$. This property causes all kind of beautiful effects and makes QCD a rich theory. A well-known phenomenon in quantum chromodynamics, as it is also known for the weak interaction, is the so-called *self coupling* of the gauge bosons (see gluon self-interaction in Figure 1.1).

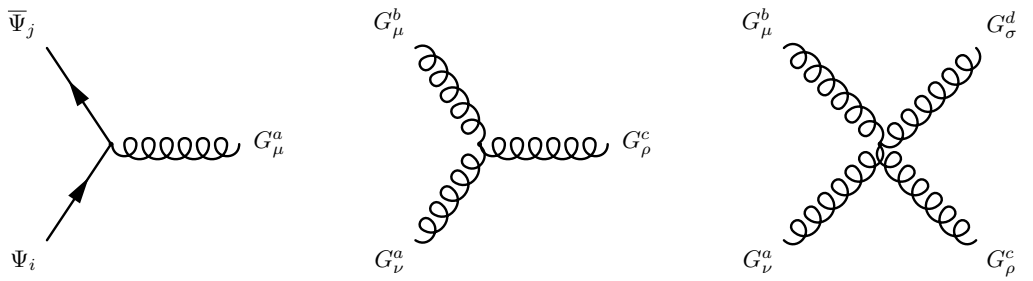


Fig. 1.1: The basic building blocks of QCD Feynman diagrams: (*left*) quark-gluon vertex (*middle*) three-gluon vertex (*right*) four-gluon vertex.

Contrary to QED, the mediator bosons in QCD carry a (color) charge by themselves, which induces self coupling. The corresponding terms can easily be identified by expanding and rearranging the kinetic term of the gluon field strength tensor $\mathcal{L}_{\mathrm{QCD}} \supset \frac{1}{4}\mathrm{tr}\left[G_{\mu\nu}G^{\mu\nu}\right]$.

The self-coupling of the gluon fields has further consequences with regards to the *effective charge* or *coupling* of the theory that is related to the β-function, $\beta(g) = \mu \frac{\mathrm{d}g}{\mathrm{d}\mu}$ of the theory,

which encodes the dependency of the "coupling constant" on the energy scale μ. In case of the non-abelian gauge theory QCD, the situation is different because the β-function is negative (antiscreening) [82, 83], which means that the effective strong coupling becomes small a short distances, leading to the concept of *running coupling*. This is a characteristic feature of the strong force, which gives rise to *asymptotic freedom*, which will be further discussed in Section 1.2.2.

Asymptotic freedom in QCD results in small couplings for sufficiently large energy scales (short distances) that allows to consider quarks and gluons "quasi-free" particles. In this limit, perturbation theory is appropriate to compute the matrix element of the respective process down to a characteristic *energy scale* $\Lambda_{\mathrm{QCD}} (\sim \mathcal{O}(100\,\mathrm{MeV}))$ where non-perturbative dynamics dominates. For small energies however, the coupling becomes large and perturbation theory is no longer applicable; therefore, predictions in this domain mostly rely on phenomenological model such as, e.g., the (Lund) string model [84]. Thus, the name *non-perturbative* regime. This phenomenon where the underlying partons can not be isolated from the hadrons anymore is known as *confinement*.

1.2 Perturbative QCD

As already stated above, the Standard Model of particle physics is the most accurate fundamental theory of the smallest building blocks of matter and their interaction by the fundamental forces of nature except gravity. The previous section provided a brief outline of the Standard Model's history and introduced some most basic concepts as well as its particle content. This section directly joins the previous one and serves as a basic introduction into the theoretical framework of perturbative QCD. Furthermore, the very important phenomenon of the running coupling constant and its physical and practical implications are discussed in more detail.

1.2.1 Renormalization and Running Coupling

Historically – very soon after its initial success – QFT faced a serious problem: the "naïve" computation of higher-order terms in the perturbative expansion, such as loop diagrams in the correlation functions, results in divergent integrals over the particle's (intermediate) momentum. As a consequence, most of the terms in the perturbative expansion are indeed infinite, contradicting the finite observables measured in the experiments. It was a long-lasting development process until this peculiar phenomenon and its physical meaning were finally understood. Important contributions to a better understanding of renormalization in particular came from Sin-Itiro Tomonaga (朝永 振一郎), Julian Schwinger and Richard Feynman, who all were awarded the Nobel Price in 1965 *"for their fundamental work in quantum electrodynamics, with deep-ploughing consequences for the physics of elementary particles[.]"* [85].

In a nutshell, the renormalization process aims to remove divergencies in the computations of the physical observables. It usually starts with the *regularization scheme* by introducing an additional (non-physical) parameter μ that allows the isolation of the part that gives rise to divergencies. This parameter may be a cutoff (see cut-off regularization) or a modification of the dimensionality of the respective integral (see dimensional regularization). The divergencies are removed ("discarded") by the redefinition, which is the actual *renormalization step*, of the parameters, e.g. the mass, the fields or the charge, in the Lagrangian. This is an extraordinary step that requires some reflection since it changes our perspective on the actual parameters occurring in the Lagrangian 1.2–1.5. Due to the regularization procedure, the physical quantities are finite without any divergencies; however, they are now a function of the artificially introduced regularization parameter μ on which the final prediction of the actual *observable* must not depend on.

The procedure described above, consisting of regularization and renormalization, allows to obtain finite (renormalized) states if the theory (like, e.g., QCD) is inherently renormalizable. Unfortunately, there is a certain degree of arbitrariness attributed to this technique, since the individual regularization and renormalization steps are not unique. As a consequence, the resulting QFT depends on the respective renormalization scheme. Hence, regularization and renormalization result in a *family* of QFTs that depend on the specific choice of the scheme

and the regularization parameter. To obtain a consistent procedure, the resulting quantum field theory must be independent of those particular choices. This requirement induces the aforementioned the so-called *renormalization group equations* (see *Callan-Symanzik equations* in QED) that ensures scale invariance of the physical observables if the associated beta-function vanishes. If the renormalization group equations are satisfied, the different regularization and renormalization schemes are guaranteed to result in an equivalence class of QFTs.

Based on the renormalization group equation, the regularization parameter μ is absorbed into the coupling of the theory, giving rise to the famous *running coupling*, i.e., the functional dependence of α_S on μ, which is the subject of the following section.

1.2.2 Asymptotic Freedom and Color Confinement

The renormalization procedure roughly outlined in Section 1.2.1, results in a dependence of the coupling α_S on the regularization parameter μ that is governed by the underlying renormalization group equation(s). Of particular interest is the aforementioned *beta-function* of QCD

$$\beta(\alpha_S) := 4\pi\mu^2 \frac{\partial \alpha_S}{\partial \mu^2} = 4\pi \sum_{k=0}^{\infty} \beta_k \left(\frac{\alpha_S}{4\pi}\right)^{k+2} \approx -\frac{\alpha_S^2}{4\pi}\left(\frac{11}{3}N_c - \frac{2}{3}N_f\right), \tag{1.7}$$

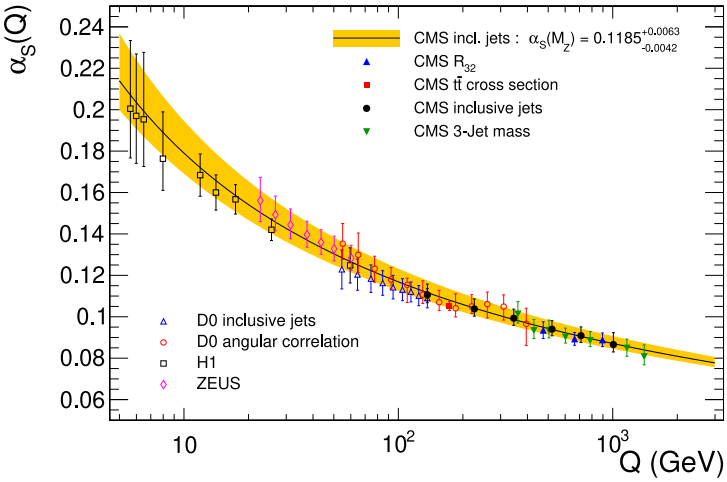
which encodes the information of how the "coupling constant" "runs" with the "energy scale" of the interaction μ. The approximation of Equation 1.7 to $\mathcal{O}(\alpha_S^2)$, with $\beta_0 = \frac{2}{3}N_f - \frac{11}{3}N_c$, is the so-called *one*-loop beta-function (one loop in the perturbative expansion), whereby N_c denotes the number of colors and N_f the number of quark flavors in the theory. For three colors $N_c = 3$ and $N_f = 6$ quark flavors – as observed in nature –, the beta-function is negative $\beta_0 < 0$, which is a characteristic feature of QCD contrary to QED where β_0 is strictly positive. Solving the differential Equation 1.7 for the integral boundaries $[Q, \mu]$ within the one-loop approximation and rewriting the resulting expression in terms of the energy scale $\Lambda_{\mathrm{QCD}} := \mu^2 \exp \frac{4\pi}{\beta_0 \alpha_S(\mu^2)}$ gives

$$\alpha_S(Q^2) = -\frac{1}{\frac{\beta_0}{4\pi}\log\left(\frac{Q^2}{\Lambda_{\mathrm{QCD}}^2}\right)}. \tag{1.8}$$

Due to $\beta_0 < 0$, the coupling α decreases with the energy scale Q^2. This phenomenon is known as *asymptotic freedom* since the coupling vanishes in the limit of very large energies $\lim_{Q^2\to\infty} \alpha_S(Q^2) = 0$; therefore, QCD appears to be a *free* theory in the ultraviolet limit (interestingly, according to Equation 1.7, theories based on the SU(3) are only asymptotically free if $N_f < \frac{11}{2}N_c$). Figure 1.1 shows several measurements of the strong coupling constant α_S for different energy scales Q^2 at various experiments that nicely illustrate the aforementioned phenomena. The current world average value of α_S evaluated at the mass of the Z boson $Q^2 = M_Z$ is given by $\alpha_S(Q^2 = M_Z^2) = 0.1172 \pm 0.0059$ [47]. The characteristic scale Λ_{QCD} is, by construction, invariant under the renormalization group and corresponds roughly to the energy scale at which $\alpha_S(\Lambda_{\mathrm{QCD}}) \sim \mathcal{O}(1)$ where perturbation theory is no longer applicable. Hence, Λ_{QCD}, which is in the order of (light) hadron masses, determines the energetic boundary between the perturbative and the non-perturbative regime in QCD. So, perturbative QCD is able to provide meaningful results if $E \gg \Lambda_{\mathrm{QCD}}$.

Another characteristic property of QCD, besides asymptotic freedom, is the so-called *color confinement* that accounts for the experimental evidence of quarks only existing within bound states that are a color-singlets under $\mathrm{SU}(3)_\mathrm{c}$ transformations, i.e., "abstract" rotations in color space. This is the reason why particles with fractional electric charge have never been *directly* observed in any experiment up to the present day. Color confinement is phenomenologically well-established and unquestioned; however, it still is remains a hypothesis in QCD because until now there is no general mathematical proof of this property for non-abelian gauge theory based on first principles in QFT[6]. Due to the lack of precise theoretical description, color

[6]The proof of confinement is equivalent to show that the quantum Yang-Mills theory exists and has a mass gap [86], which is one of the Millenium Problems advertised by the Clay Mathematics Institute and awarded with US$1 million for a righteous solution to the problem. Give it a try!



Plot 1.1: The strong coupling α_S (full line) and its total uncertainty (band) as a function of the respective choice for the scale Q. The four analyses of α_S from CMS [30, 31, 32, 33] are shown together with results from the H1 [34] [35], and D0 [36, 37] experiments at the HERA and Tevatron colliders. Figure and caption taken from proceeding report [38] "α_S determinations from CMS (status and plans)", Klaus Rabbertz (on behalf of CMS), KIT, Karlsruhe, Germany, Figure 1.

confinement needs to be approximately described by a potential $V(r)$ (which can actually be computed in lattice QCD (see, e.g., Ref. [87])) between two quarks in a color-singlet state

$$V(r) = -\frac{4}{3}\frac{\alpha_S}{r} + kr, \tag{1.9}$$

with $V(r) \to \infty$ for $r \to \infty$ (confinement) and $V(r) \to -\frac{4}{3}\frac{\alpha_S}{r}$ for large energies (asymptotic freedom). This potential is extensively used in phenomenological hadronization models such as, e.g., the already mentioned Lund string model [88], which will be further explained in the next chapter in the context of hadronization models used in Monte Carlo simulation.

1.2.3 Soft and Collinear Limits of QCD

The previous section introduced the effect of asymptotic freedom at *short* distances (large energies) and color confinement in the infrared regime (low energies), both of which are a characteristic feature of QCD that follows from the dependence of the strong coupling $\alpha_S(Q^2)$ on the momentum transfer Q^2, i.e., the running coupling which in turn follows from the self-coupling of the gluon fields. Perturbative QCD takes advantage of this property by performing a perturbation expansion of, e.g., the cross section around the strong coupling if the characteristic momentum transfer Q of the process under consideration is in the appropriate domain $Q \gg \Lambda_{\mathrm{QCD}}$. The term *perturbation expansion* refers to a power series that divides a problem, which can not be solved exactly, into several (usually infinitely many) subproblems that in turn can be solved analytically or numerically. The quality of the approximation is determined by the number of terms that are included in the power expansion. Formally, the perturbative expansion of the cross section of a given process $i \to f_n$ with n particles in the $f$inal state is given by

$$\mathrm{d}\sigma^{if_n} = \sum_{k=0}^{\infty} \alpha_S^{k+n} \mathrm{d}\widetilde{\sigma}_k^{if_n}, \tag{1.10}$$

whereby $\widetilde{\sigma}_k^{if_n} := \sigma_k^{if_n}/\alpha_S^{k+n}$ denotes kth expansion of the cross section σ^{if_n}, factorized for the coupling α_S. For the method to be valid and give meaningful results, the expansion coefficient (i.e, the coupling) must be sufficiently small ($\alpha_S < 1$) such that the contributing corrections to the matrix element decrease for higher-orders in α_S. In practice, however, the perturbative

series according to Equation 1.10 has to be terminated prematurely.

It is instructive to study some general properties of the matrix element in QCD in the soft and collinear limit (this will become very important in the next chapter in the context of parton showers). The cross section, which is proportional to the squared amplitude, of the process $i \to f_n$ can be computed according to

$$\mathrm{d}\sigma^{if_n} = \mathrm{d}\Phi_n \left|\mathcal{M}^{if_n}\right|^2 , \tag{1.11}$$

with n partons in the final state. Equation 1.11 comprised the kinematic information of the available differential phase space $d\Phi_n$ as well as the information of the underlying physical process that is encoded in the matrix element $\mathcal{M}^{if_n}$. If the nth particle is a collinear and soft gluon – i.e., radiated "Bremsstrahlung" – the cross section according to Equation 1.11 can be simplified (in anticipation of the factorization theorem to be introduced in Chapter 2.4.1) according to

$$\lim_{\substack{\theta\to 0\\ E_n\to 0}} \mathrm{d}\Phi_n \left|\mathcal{M}^{if_n}\right|^2 = \mathrm{d}\Phi_{n-1} \left|\mathcal{M}^{if_{n-1}}\right|^2 \frac{\alpha_S C_i}{\pi} \frac{\mathrm{d}\theta^2}{\theta^2} \frac{\mathrm{d}E_n}{E_n}. \tag{1.12}$$

The factorized cross section according to Equation 1.12 has a non-integrable divergence if the radiated gluon is very soft, i.e., $E_n \to 0$ and/or the radiation angle is very small $\theta \to 0$. This apparent contradiction to the finite observables measured in the experiments is explained by the circumstance that the partonic cross section is not *an* actual physical observable. Due to the color confinement explained in Section 1.2.2, quarks and gluons are not observed as freely propagating particles but as a compound states, i.e., hadrons that are a singlet under rotations in color space. This problem can be solved by "absorbing" the unpleasant divergencies in the (renormalized) *parton distribution functions* (PDFs), which are the subject of the following section.

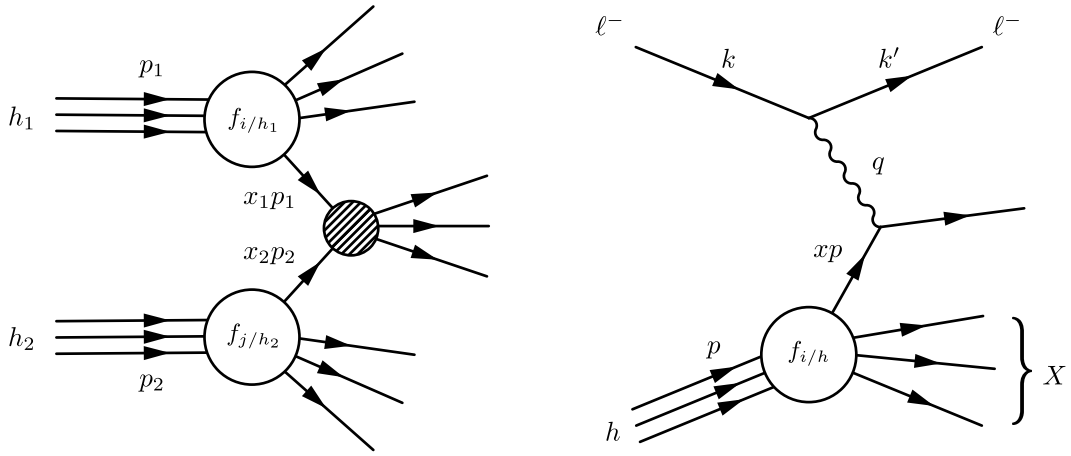
1.2.4 Factorization Theorem and DGLAP Equations

In general, a QCD process always(!) includes both short- and long-distant behavior. Therefore, perturbation theory alone is not sufficient to provide reliable predictions. Especially at lepton-hadron and hadron-hadron colliding experiments, most QCD processes live in the low energetic regime where the formation of color singlet states, i.e., bound hadrons takes place. The *factorization theorem* divides the cross section into a hard process – that comes with large momentum transfer Q^2 and hence can be computed with perturbation theory (see Section 1.2.3) – and a long-distant part that describes processes with low momentum transfer, which is mostly based on empirical models. Figure 1.2 provides a visualization of the factorization theorem for the example of hadron-hadron interaction (1.2a) and deep inelastic scattering (1.2b) through hadron-lepton interaction. In Figure 1.2, the structure of the hadrons (such as protons) is described by the Parton Density Functions (PDFs), while the hard subprocess is given by the scattering cross section. According to the factorization theorem, the differential cross section $\mathrm{d}\sigma^{h_1h_2\to f}$ of a physical process $h_1h_2 \to f$ (read: interaction between hadron h_1 and h_2 with the final state f) is given by the convolutional integral

$$\mathrm{d}\sigma^{h_1h_2\to f} = \int_{[0,1]}\int_{[0,1]} \mathrm{d}x_1\mathrm{d}x_2 \sum_{i,j} f_{i/h_1}(x_1,\mu_F) f_{j/h_2}(x_2,\mu_F) \mathrm{d}\sigma^{ij\to f}\left(x_1, x_2, \left(\frac{Q}{\mu_F}\right)^2\right), \tag{1.13}$$

with x_1, x_2 being the *momentum fraction* carried by the respective parton, and $\sigma^{ij\to f}$ being the partonic cross section and μ_F denoting the so-called *factorization scale* that can be thought of the scale that separates long- and short-distance contributions. Emission below the factorization scale (long distance effect with small energy scales Q^2) are described by the PDFs. Equation 1.13 divides the computation of the cross section into two parts: the hard scattering cross section $\mathrm{d}\sigma^{h_1h_2\to f}$ at some order in perturbation theory and non-perturbative contribution that accounts for the complex internal structure of the hadrons involved in the interaction. Loosely speaking, the computation of the cross section has been separated in an "analytic" and an "empirical" part. Equation 1.13 also provides experimental instruction on how to measure the

PDFs in Deep Inelastic Scattering (DIS).



(a) Hadron-hadron interactions *(b)* Deep inelastic scattering

Fig. 1.2: Visualization of the factorization theorem for hadron-hadron interactions (1.2a) and deep inelastic scattering (1.2b) with the respective flavor-dependent PDFs $f_{i/h}$.

The precision of the cross section computation according to the factorization theorem according to Equation 1.2 depends on the uncertainty of the hard scattering cross section and the PDFs. While the former depends on the order of perturbation theory, i.e., the number of terms that are included in the perturbative expansion of the cross section, the latter is dominated by experimental uncertainties in the measurement of the PDFs. Therefore, a precise measurement of $f_{i/h}(x, \mu_F)$ is indispensable and utterly important to obtain precise predictions of the cross section at hadron-hadron colliders.

The PDFs as introduced above are used to provide a non-perturbative description of the internal structure of hadrons. More specifically, at leading-order, the parton distribution function $f_{i/h}(x, Q^2)$ of a hadron h corresponds to the probability to find a parton of type i (a quark of a certain flavor or a gluon) with a longitudinal momentum fraction x of the compound object at an energy scale of Q^2. While (at present) the PDFs itself cannot be derived from first principles in QFT directly, their evolution with the energy scale Q^2, which can be thought of as the energy transfer between the hadron and its scattering partner, is accurately predicted by the so-called DGLAP (Dokshitzer–Gribov–Lipatov–Altarelli–Parisi) equations [89, 90, 91], which follows from the requirement that the observable structure functions are *in*dependent of the non-physical factorization scale. The DGLAP equations, which describe the PDFs' dependence on μ_F and momentum fraction x, are given by

$$\mu_F^2 \frac{\partial f_{i/h}(x, \mu_F^2)}{\partial \ln \mu_F^2} = \sum_j \frac{\alpha_S}{2\pi} \int_x^1 \frac{\mathrm{d}z}{z} P_{ij}(z) f_{j/h}\left(\frac{x}{z}, \mu_F^2\right), \tag{1.14}$$

where P_{ij} are the (leading-order) spin-averaged, regularized *splitting functions/kernels* that describe the probability of a daughter parton i splitting from a parent parton [5]. A pictorial representation of the DGLAP equations is given in Figure 1.3.

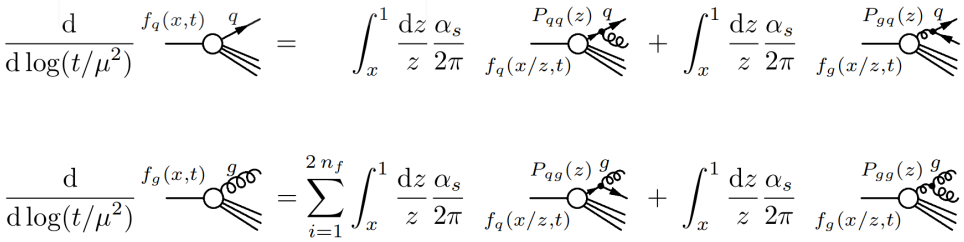
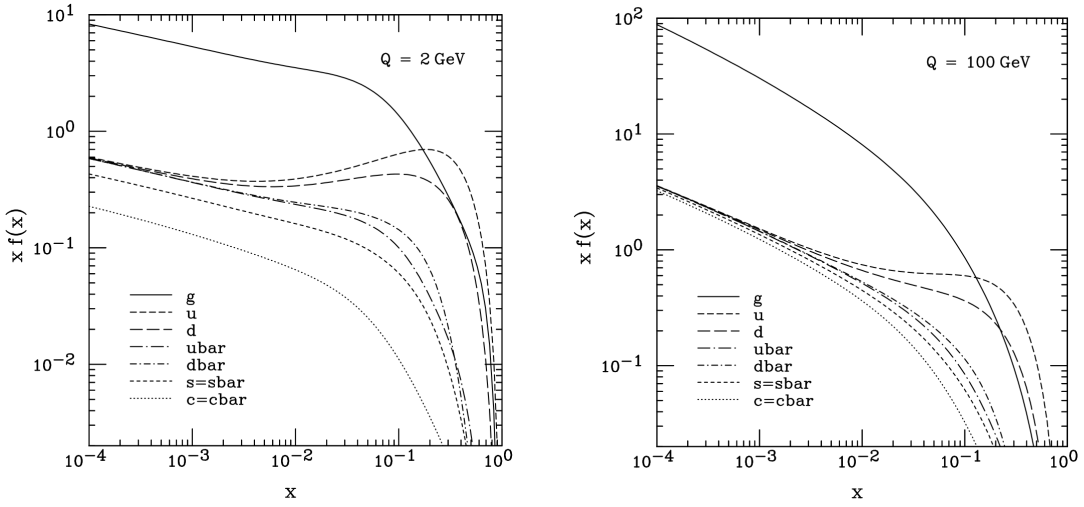


Fig. 1.3: Pictorial representation of the evolution of the parton distribution functions via the DGLAP equations (adapted from Ref. 5, Fig. 1, p. 3).

With the DGLAP equation(s) at our disposal, the PDFs measured at some lower energy scale Q^2 – as it has been extensively done at the HERA and the Tevatron collider in DIS – can be evolved to higher energy scales required by more recent colliders like the LHC. In this sense, the PDFs are *universal*.

Figure 1.2 gives an example of two parton distribution functions (CTEQ6M) published by



(a) $Q = 2\,\mathrm{GeV}$ *(b)* $Q = 100\,\mathrm{GeV}$

Plot 1.2: Parton distribution functions (CTEQ6M) for different partons at the energy scale $Q = 2\,\mathrm{GeV}$ (*left*) and $Q = 100\,\mathrm{GeV}$ (*right*) (adapted from Ref. 39, Fig. 1, p. 8).

the CTEQ Collaboration [39]. The two plots illustrate the PDFs for different energy scales $Q = 2\,\mathrm{GeV}$ and $Q = 100\,\mathrm{GeV}$ for different partons in the hadron. The up-type quarks are the dominant distribution for small values of Q and large momentum fractions x. For smaller values of x, however, the distribution is mostly dominated by gluons; hence, they carrying small momentum fractions. The distribution changes significantly for higher energy scales. For higher values of Q, the distribution becomes progressively flavor-independent for small values of x, a flavor symmetry that is not present in the PDF for $Q = 2\,\mathrm{GeV}$.

The parton distribution functions are, broadly speaking, experimentally determined by fitting a large number of measured cross sections in deep inelastic scattering in a (Q^2, x)-grid. The precise and well controlled experimental conditions in deep nuclear-lepton scattering (see Figure 1.2b) cross section measurements along with the accurate predictions in perturbative QED provide an ideal environment for probing the internal structure of hadrons.

1.3 QCD Phenomenology

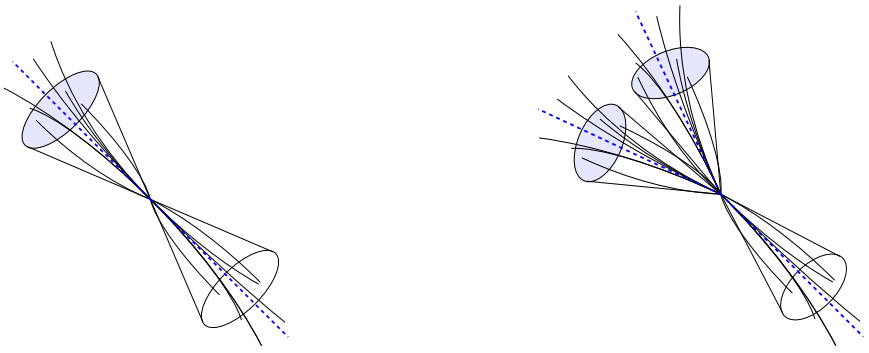
The previous section 1.2 reviewed the theoretical basics of perturbative QCD and how it can be used to compute the hard-scattering matrix elements for given processes as well as

the showering approximation to simulate all dominant collinear QCD radiation based the factorization theorem (this very important subject will be further developed in the upcoming chapter in the context of Monte Carlo event generators). As it was shown, the cascade of splittings caused by the initial parton results in a large number of additional particles that all carry a fraction of the primary parton's four-momentum. To reconstruct the four-momentum of the original parton, it is therefore necessary to measure and reconstruct the additional particles that are a consequence of this showering process. However, a particle detector is by no means a sterile environment. Usually, there are several interactions involved that cause pile-up contamination of the environment as well as higher particle multiplicities in the final state. As a result, the difficulty lies in the identification of active regions in the detector that are assumed to originate from the same process and the same initializing particle. This is – as one can imagine – a very sophisticated task. Over the years many different concepts and techniques have been developed to solve this problem in a reasonable manner. The purpose of this chapter is therefore to provide a short introduction into the extensive field of jet physics, i.e., a phenomenological approach to QCD whereby it is limited to the main theoretical aspects of this area.

1.3.1 Jets in Particle Physics

The abstract object that particle physicists refer to as *jet* is the remaining experimentally accessible signature of quarks and gluons that are produced in high-energy particle collisions, but escape direct detection in the experiment. Therefore, in many instances, the reconstruction of jets is the only way to gain insight into the underlying process of an event involving the strong force. Furthermore, hadron colliders such as the LHC are dominated by QCD process, which is why the reconstruction of jets often is the only way to extract structures and make sense out of an event in the first place. As a result, studying jets and its properties is utterly important and the key element for many studies – in particular for the search of new physics beyond the Standard Model.

Figure 1.4 shows to different jet typologies as they are steadily observed in experiments. Each solid line in Figure 1.4 represents the path or track of a particle that originates from



(a) *(b)*

Fig. 1.4: A two-jet event with clear separation of the final state particles (*left*); a three-jet event with significant overlap among radiation from two different sources. The dashed, blue line indicated the direction of the respective underlying parton (*right*).

the interaction point of the event. The tracks of charged particles are measured in the Inner Detector regions, which are usually equipped with semiconductor detectors, while the energy deposition of the particles is measured in the so-called calorimeter (derived from Latin *calor*, "heat", and Greek *metron* (μέτρον), "measure") of the detector. The left Figure in 1.4 is a somewhat simplified illustration of a two-jet event, i.e. two separated collection of particles in the final state to balance energy-momentum. The right hand side of the Figure, however, shows a three-jet event with three final state particles. As seen in this topology, there is the possibility of significant overlap between the particles measured in the detector, making the assignment

to the underlying process difficult. This becomes even worse for higher jet multiplicities in the final state. For many years, theorists and experimentalists alike worked on efficient methods that allow the reconstruction of jets through clustering particles. In the course of these efforts, a variety of different algorithms have been developed to fulfill this task. It is therefore evident that the definition of a jet is based on the respective jet reconstruction algorithm that are used since different reconstruction methods will result in non-identical jets with different particle content.

It should be noted, that the general concept of a particle jet (as outlined above) is very intuitive by no means unproblematic. The underlying picture of jet physics is often a one-to-one correspondence with a particle jet that has been reconstructed in the detector and an underlying (outgoing) particle/parton in the calculation of the hard subprocess. This assignment, however, is – strictly speaking – ill defined, because the jet is (as a measure) a color-neutral object, while the underlying parton is not a singlet under SU(3) transformations. A jet, therefore, is a *semiclassical* concept. This problem will become important once the concept of quark- and gluon-initialized jets (or jet flavor in general) are discissed in Chapter 7.

1.3.2 Infrared and Collinear Safety

Generally speaking, jet reconstruction is the task to group particles into orthogonal categories that are assumed to originate from the same parton of the underlying hard scattering process[7]. To organize the particles into groups, the jet reconstruction algorithm usually comes with a distant measure, as well as recombination scheme that specifies how particles are (re-)combined into a single four-vector for the next iterative step [92].

There exists a large variety of different jet algorithms since, historically, each experiment used its own definition of a jet. In 1990, there was an attempt to specify a set of criteria based on experimental and theoretical considerations that each jet reconstruction algorithm should meet. This set of rules became later known as the "Snowmass Accord", which was the first step *"[t]oward a standardization of jet definitions[.]"* [93]. The CDF Collaboration at the Tevatron, Illinois, USA, was the first collaboration that tried to implement this "accord" [94].

In the Snowmass accord from 1990 the Authors proposed *"[s]everal important properties that should be met by a jet definition[.]"* [93]. Among other things, the jet definition should *"[y]ield[.] finite cross sections at any order of perturbation theory"* as well as *"[...] a cross section that is relatively insensitive to hadronization"* (Ref. 93, p. 6, property 4 & 5). Those properties are guaranteed if the definition of a jet is *infrared* and *collinear* (IRC) safe. Formally, infrared safeness can be defined as:

> *An observable [$\mathcal{O}$] is infrared safe if, for any parton configuration,adding an infinitely soft parton does not affect the observable at all (Ref. 95, Def. 4, p.5).*

$$\mathcal{O}_{m+1}(..., k_i, ...) \to \mathcal{O}_m(..., k_i, k_j, ...) \quad \text{if} \quad k_i \to 0[.] \tag{1.15}$$

Similar, one defines collinear safeness:

> *An observable [$\mathcal{O}$] is collinear safe if, for any parton configuration, replacing any massless parton by an exactly collinear pair of massless partons does not affect the observable at all (Ref. 95, Def. 5, p.5).*

$$\mathcal{O}_{m+1}(..., k_i, k_j, ...) \to \mathcal{O}_m(..., k_i + k_j, ...) \quad \text{if} \quad k_i \parallel k_j[.] \tag{1.16}$$

As described in Section 1.2, the parton in the hard interaction undergoes numerous soft and collinear splittings as part of the fragmentation process that finally leads to stable hadrons in the final state. A consistent jet definition should therefore be insensitive to those effects, i.e., the radiation of additional soft and collinear partons shall not change the jet. Apart from that, fixed-order calculations in perturbative QCD soft radiation and collinear splittings are accompany by divergent matrix elements already at tree-level. Those divergences are guaranteed to cancel with higher order loop diagrams that enter the calculation with an opposite sign.

[7] A comprehensive introduction into several approaches to the reconstruction of jets along concrete examples will be provided later in Chapter 4 in the second Part of this report.

2. Event Generation/Simulation in HEP[1,2]

Until now, the focus was largely on theoretical and conceptional considerations. However, in physics – as in natural sciences in general –, each theory must prove itself valid in the experiment; otherwise, it will be buried in the vast cemetery of disproven theories.

In this chapter, both threads, theory and experiment, are wed together to provide an introduction into the broad topic of *event generators* along with their practical applications in experimental high-energy particle physics. A full description of this extensive subject is far beyond the scope of this work; therefore, it is limited to merely convey an idea of the underlying principles that are common to most event generators.

The title of this chapter already reflects parts of its structure. Starting from the Monte Carlo method (2.1), the very general concept of Monte Carlo integration (2.1.1) and simulation (2.1.2) is introduced (which will also be relevant for the subsequent chapters), as both methods are fundamental and frequently used in high-energy particle physics. The next section (2.2) then builds a bridge to the subject matter of the previous chapter by introducing the idea behind of Monte Carlo event generators (2.2.1) in the context of high-energy particle physics. Furthermore, special attention is devoted to the fixed-order computation of the matrix element of the underlying hard subprocess (2.3) as well as the simulation of parton showers utilizing the factorization theorem (2.4) that have been introduced in the previous chapter. The next section (2.5) provides an introduction into the most common phenomenological models for the simulation of the (non-perturbative) formation of color-singlet states by means of hadronization. As most of the states produced in the previous hadronization step are energetically excited and unstable, the next section (2.5.1) very briefly explains the subsequent decay to final state particles that are measured in the detector. The last element of the simulation chain (2.6) is the simulation of the specific measuring apparatus that is used to measure and reconstruct the particles in the event.

2.1 The "Monte Carlo" Method

Generally speaking, the Monte Carlo method – a term first introduced by the Greek-American physicist Nicholas Metropolis (Νικόλαος Μητρόπουλος) and an innuendo to the eponymic Monégasque city – is, in its broadest sense, a class of numerical algorithms to solve complex mathematical problems employing probabilistic principles. The application of this method can be broadly divided into Monte Carlo *integration* (2.1.1) and *simulation* (2.1.2).

It was first made public in a theoretical essay by Nicholas Metropolis and the Polish-American scientist Stanłslaw Ulam in 1949, mainly to take *"[...] a statistical approach to the study of differential equations [...]"* [96]. However, the proposed method was already invented several years before by Ulam when he took part in the nuclear weapons program at the Los

[1]Essential parts of the succeeding text passages have been adopted (almost) literally from my Master's thesis "*Towards a Data-Driven Simulation of QCD Radiation with Generative Models utilizing Machine Learning Methods*", 2019, Universität Heidelberg, Christof Sauer [49].

[2]This chapter may safely be skipped provided that the reader is already familiar with the basic concepts (like parton showers, hadronization, factorization etc.) behind event simulation in HEP.

[2]Essential parts of the succeeding text passages have been adopted (almost) literally from my Master's thesis "*Towards a Data-Driven Simulation of QCD Radiation with Generative Models utilizing Machine Learning Methods*", 2019, Universität Heidelberg, Christof Sauer [49].

Alamos National Laboratory, USA (where he was involved in the Manhattan Project during World War II). The Hungarian-American scientist John von Neumann, who was involved in the same project at that time, quickly recognized the importance of the novel method suggested by his colleague. In a first practical application of the method, Ulam and von Neumann studied – under strict secrecy self-evidently – the problem of neutron diffusion, which is essential to build and design nuclear weapons [97].

Although the Monte Carlo method has a turbulent history that goes back to a period of geopolitical tension and military confrontation, it has become an indispensable tool in a wide spectrum of scientific applications: from industrial engineering; the simulation of physical, chemical and biological processes of all different kinds; to economics and finance. Everywhere, people "play dice".

2.1.1 Monte Carlo Integration

Formally, Monte Carlo computations are equivalent to an approximate integration in higher-dimensional space. Consider, for instance, the d-dimensional integral

$$I = \int_{[0,1]^d} \mathrm{d}\boldsymbol{x}\, f(x) = \int_{[0,1]} \mathrm{d}x_1 \cdots \int_{[0,1]} \mathrm{d}x_d\, f(x_1, \ldots, x_d), \tag{2.1}$$

whereby $f : [0,1]^d \to \mathbb{R}^d$ is a real-valued function over (for reasons of simplicity and without loss of generality) the normalized hypercube $[0,1]^d = \Pi_{i=1}^d [0,1]$. Equation 2.1 may be interpreted as an expectation value $\mathbb{E}_{X\sim\mathcal{U}^d(0,1)}\left[f(X)\right]$ of the function f over the d-dimensional uniform probability distribution $\mathcal{U}^d(0,1)$, with X being an vector of *i.i.d.* random variables $X \sim \mathcal{U}^d(0,1)$, meaning $X_i \in [0,1]$ with $i \in \mathbb{N}^{\leq d} \setminus \{0\}$.

The Monte Carlo approximation of Equation 2.1 is then given by

$$S_n = \frac{1}{n} \sum_{i=1}^{n} f(x_i), \tag{2.2}$$

where $\{x_i\}_{i=1}^n$ are n independent samples of the random variable X. The convergence of the Monte Carlo approximation according to Equation 2.2 to the Integral 2.1 is guaranteed by *the law of large numbers*, i.e., the sample mean of a sequence of variables $\overline{S}_n = \frac{1}{n}\sum_{i=1}^n S_i$ converges to the expectation value $\overline{S}_n \to \mu := I$ for $n \to \infty$ *iff* the first two moments $\mathbb{E}[\overline{S}_n] = \mu \in \mathbb{R}$ and $\mathbb{E}\left[(S_i - \mu)^2\right] = \sigma^2 \in \mathbb{R}^{>0}$ exist. The circumstance that the sample mean converges to a fixed number justifies the approximation of Equation 2.2 by means of Equation 2.1.

The statements above vindicates the estimation of integrals by means of the Monte Carlo approximation. However, what are the benefits of Monte Carlo integration compared to other numerical integration methods available (e.g., rectangular integration or Simpson's rule)? The advantage lies in the statistical nature of the method that not only connects it to the law of large numbers but also to *the central limit theorem*[3] in probability theory.

According to the central limit theorem, the sample mean $\overline{S}_n$ of a sequence $\{S_i\}_{i=1}^n$ of with $\mathbb{E}[S_n] = \mu$ and $\mathrm{Var}(S_i) = \sigma^2$ approximately follows a normally distribution $\mathcal{N}(\mu, \sigma^2/n)$. This implies that the standard score $\sqrt{n}\frac{\overline{S}_n - \mu}{\sigma}$ converges to a stochastic variable that is normal distributed with zero mean and unity variance, i.e., $\sqrt{n}\frac{\overline{X}_n - \mu}{\sigma} \overset{n\to\infty}{\sim} \mathcal{N}(0,1)$.

To relate the statements above with Monte Carlo integration, consider the approximation error $S_n - I(f)$; according to the central limit theorem this quantity is (approximately) normal distributed $S_n - I(f) \dot{\sim} \mathcal{N}(\mu, \sigma^2/n)$. Then, the following statement is true within the limits of finite statistical precision: $P\left(a\frac{\sigma}{\sqrt{n}} < S_n - I(f) < b\frac{\sigma}{\sqrt{n}})\right) \approx \Phi(b) - \Phi(a)$, with $\Phi(\cdot)$ being the cumulative distribution function. From this simple consideration, it can be concluded that the approximation error of the Monte Carlo method behaves like $\mathcal{O}\left(1/\sqrt{n}\right)$. The statement holds – and this is the decisive aspect! – *regardless* of the dimensionality $d \in \mathbb{N}$ of the integral according to Equation 2.1. This property, i.e., the approximation error being independent of

[3] For a proof of the central limit theorem see Ref. [98].

the dimensionality of the integral to be calculated, makes the Monte Carlo method superior when it comes to evaluate *higher dimensional* integrals (for $d \leq 2$ there are numerical methods that give lower or similar errors with $\mathcal{O}\left(1/n^{1/d}\right)$, e.g., the general trapezoid approximation).

The line of arguments above explain the success of Monte Carlo integration in many scientific fields, which usually deal with high-dimensional integrals. The same argument applies to high energy particle physics due to the appearance of high-dimensional phase-space integrals in, e.g., cross section computations. This makes the Monte Carlo method (integration) a natural method of choice in particle physics and hence in the event generators discussed in Section 2.2.

2.1.2 Monte Carlo Simulation

As it has been mentioned, the code name "Monte Carlo" is derived from the city with the same name in Monaco, situated along the French Riviera. It is well known for its numerous casinos, which makes Monte Carlo the world's largest gambling centre. The term is by no means arbitrarily chosen: Nicholas Metropolis suggested the name as an allusion to Ulam's uncle who used to borrow money from its relative, only to squandering it for gambling [99].

A typical problem in statistical physics (as well as in many other fields) is to estimate the mean value of some function h of a random variable $\boldsymbol{x}$ with respect to a (usually unknown) probability density function $f(\boldsymbol{x})$ (PDF)

$$\langle h \rangle \coloneqq \mathbb{E}_{\boldsymbol{x}\sim f(\boldsymbol{x})}\left[h(\boldsymbol{x})\right] = \int_{\Omega} \mathrm{d}\boldsymbol{x}\, h(\boldsymbol{x}) f(\boldsymbol{x}). \tag{2.3}$$

A typical example in experimental particle physics would be, for instance, the expected mean energy $\langle E \rangle$ deposition in some region and part of the detector system or a cross section. The probability of some particle to deposit energy in the interval $[E, E + \mathrm{d}E]$ is given by $f(E)\,\mathrm{d}E$. However, the PDF might be a very complicated function that is *a priori* unknown. This is usually the case. The Monte Carlo method provides a procedure to estimate the unknown PDF, f, through simulation of the actual (physical) processes assumed to be involved and hence is referred to as *Monte Carlo simulations*. The objective is to numerically mimic and reproduce the basic processes as accurately as possible to get a good estimation of the underlying PDF that have generated the data. This requires precise theoretical and experimental models that accounts for the stochastic nature of the respective process under consideration.

Monte Carlo simulations are used extensively in particle event generators to estimate the expected number of events in some regions of phase space, and therefore it is an indispensable tool in HEP.

2.2 Physics of MC Event Generators

From a philosophical perspective, theoretical concepts in physics are studied concerning their ontological assumptions and implications in nature. Hence, the interplay between theory and experiment in physics is an essential concept – one is not possible without the other. From an epistemological point of view, the agreement of theoretical predictions and corresponding observations in nature does not only allow for a description of the process under consideration of a specific language, but also reveals a deep insight in the actual structure of the surrounding reality [100]. This fundamental concept applies to all physics questions and fields. Indeed, one could say that for a theory to be called physical it needs to describe measurable processes. A theory that protects itself by being non-verifiable and hence non-falsifiable is not valid in a physical sense; it is part of metaphysics.

To apply those concepts to particle physics, it is necessary to compare experimental observations with theoretical predictions. Monte Carlo event generators have been developed to satisfy exactly this need. Their purpose is to incorporate the current knowledge of elementary particle physics into a tool that allows to make predictions of known processes. However, they are also used to study and predict processes beyond the current Standard Model based on theoretical considerations. Thus, Monte Carlo event generators are the bridge that links theory and experiment in the broad field of elementary particle physics.

This section intends to provide a brief introduction to Monte Carlo event generators and their applications. By doing so, it refers to the physical fundamentals introduced in Chapter 1.

2.2.1 The Bigger Picture

As state above, the objective of MC event simulation is to give accurate predictions of some physical *observable* $\mathcal{O}$ that can in the end be compared with data obtained in an experiment. Generally, this observable is computed from the reconstructed four-momenta $p_1, \ldots, p_n$ in the final state f, i.e., $\mathcal{O}_f := \mathcal{O}_f(p_1, \ldots, p_n)$[4], with n being the number of "particles" (this may also be a jet or missing transverse energy) in the final state. Due to the underlying statistics of quantum mechanics, the probability of a specific final state configuration to occur is given by the differential cross section $\mathrm{d}p_f(p_1, \ldots, p_n) \propto \frac{\mathrm{d}\sigma_f(p_1,\ldots,p_n)}{\mathrm{d}^3p_1 \ldots \mathrm{d}^3p_n}$. The cross section σ of a physical process is a *detector independent* quantity; however, it is related to the number of observed events N in an experiment/particle collider through the integrated luminosity $\mathcal{L}$ as per $N(t) = \sigma \int_t \mathrm{d}t' \mathcal{L}'$. Therefore, the frequent repetition of an experiment allows to measure the expectation value $\langle \mathcal{O}_f \rangle$ of the observable $\mathcal{O}_f$ based on the likelihood and the integrated luminosity according to

$$\langle \mathcal{O}_f \rangle_{\mathcal{L}} = \mathcal{L} \sum_{i \in I_f} \int_{\mathcal{V}} \mathrm{d}^3p_1 \ldots \mathrm{d}^3p_n \, \frac{\mathrm{d}\sigma_f(p_1, \ldots, p_n)}{\mathrm{d}^3p_1 \ldots \mathrm{d}^3p_n} \mathcal{O}_f(p_1, \ldots, p_n), \tag{2.4}$$

(compare this with Equation 2.3). The dimensionality of the complicated integral in Equation 2.5 increases with the number of final state particles in an event. As a result, one relies on Monte Carlo methods to get a reasonable estimation of the expectation value (see the comments regarding Monte Carlo integration). The Monte Carlo approximation of Equation 2.5 according to Equation 2.2 is therefore given by:

$$\langle \mathcal{O}_f \rangle_{\mathcal{L},N} \approx \mathcal{L} \frac{|\mathcal{V}|}{N} \sum_{j=0}^{N} \sum_{i \in I_f} \mathrm{d}^3p_1 \ldots \mathrm{d}^3p_n \, \frac{\mathrm{d}\sigma_f(p_1, \ldots, p_n)}{\mathrm{d}^3p_1 \ldots \mathrm{d}^3p_n} \mathcal{O}_f(p_1, \ldots, p_n), \tag{2.5}$$

with $|\mathcal{V}|$ being the integration volume and I_f denoting the index set of final state particles. Based on the statements in Section 2.1.1, the sample mean $\langle \mathcal{O}_f \rangle_{\mathcal{L},N}$ converges to the true expectation $\langle \mathcal{O}_f \rangle_{\mathcal{L}} = \lim_{N \to \infty} \langle \mathcal{O}_f \rangle_{\mathcal{L},N}$ for infinite statistics. Following Equation 2.5, the expectation value for each observable may be computed based on the measured differential cross section.

Now, to get a prediction for $\langle \mathcal{O}_f \rangle^{\mathrm{MC}}_{\mathcal{L},N}$ that can be compared with the experiment, the probability, i.e., the differential cross section of a process, must be accurately predicted by theory. Within the framework of pQCD (or perturbation theory in general), the precision of the prediction is improved by including higher-order corrections into the expansion of the matrix element. This is known as *fixed-order expansion* in the simulation chain of MC event generators, which is the subject of the following section.

2.2.2 Event Simulation and Event Topology

The purpose of Monte Carlo event generators is to simulate high energy collisions in all process steps up to the stable final-state particles that are measured in the detector. This process chain and its individual component follows physical principles and should therefore mimic the actual process that produced the event in the real world.

The generation of an event starts with the computation of the matrix element at fixed order of perturbation theory for the underlying hard subprocess with some dedicated matrix-element-generator like POWHEG or MADGRAPH5_aMC@NLO. This very first step only computes the partonic final-state without any hadronization involved (*parton-level events*). To simulate the emissions of QCD radiation, the information of the parton-level event is further processed by a shower Monte Carlo, such as HERWIG6, HERWIG++, PYTHIA6, PYTHIA8 or SHERPA,

[4] Of course, there is a much larger class of observables also including other properties.

that simulates a cascade of showering particle up to the hadronization scale where color-less hadrons are finally formed (see "color confinement" in Chapter 1.2.2), which may further decay to stable final-state particles. With the simulation of the so-called *particle-level events*, the event generation process is finished. However, due to the imperfection of measuring devices at our disposal and the resulting limitations in the reconstruction of particles, the simulation of detector effects is utterly important to get realistic predictions that can be compared to the measurement with the actual detector. The simulation of detector effects is based on Monte Carlo too (see Section 2.1.2). A well-known tool for detector simulations is GEANT4 that actually *"[...] is a toolkit for the simulation of the passage of particles through matter[.]"* [101]. After all those steps, the generated *reco/detector-level events* are comparable with the measurement in the respective detector and only this detector.

Following now is a short description of the individual steps of the event generation process and how they are related to the theoretical fundamentals introduce in this chapter.

2.3 Hard Subprocess

At modern particle colliders like, e.g., the LHC, the momentum transfer between the colliding particles (or partons in case of hadron-hadron accelerators) is usually quite large. Due to the asymptotic behavior of QCD at large energy scales and the running of the strong coupling constant as a consequence thereof, the *hard subprocess* can be computed using perturbation theory, whereby the term "hard" implies processes associated with large momentum transfer. The usage of perturbation theory for those processes allows to use a pictographical representation of the individual terms of the perturbative expansion using a set of Feynman diagrams $\{\mathcal{F}_k\}_k$ and to account for non-trivial interference effect between different processes yielding the same final state.

The hard subprocess is one ingredient in the already introduced factorization theorem (see Equation 1.13) that allows to factorize the calculation of the cross section at hadron colliders into different components. The parton-level cross section $\hat{\sigma}^{ii\to f}$ is the cross section for a specific final state f produced via the interaction of parton i and j in the respective colliding hadrons. This cross section depends on the available phase space of the final state Φ_f, the factorization μ_F and renormalization μ_R scale, the (initial-state) spin averaged squared matrix element $\left|\bar{\mathcal{M}}^{ij\to f}\right|^2$ as well as the parton flux, with the matrix element being given by a sum over all terms in the perturbative expansion that contribute to the process up to a certain order in perturbation theory $\mathcal{M}^{ij\to f} = \sum_i \mathcal{M}_i^{ij\to f}$.

2.4 Parton Shower Simulation

The calculation of the hard subprocess is followed by the simulation of additional radiation. This step is necessary because fully perturbative calculations of the cross section in QCD are usually only available to NLO precision (up to NNLO for some observables) due to the factorial increase of correction terms to the perturbative expansion of the scattering amplitude. This poses a problem since, due to the running coupling constant, higher-order terms are enhanced in some region in phase space and hence can not be discarded. This issue is addressed by all-order *resummation* techniques and/or *parton shower* models.

Since the hard subprocess usually is associated with high energies of the final-state particles at parton level, the accelerated particles that carry a color charge tend to emit radiation in form of gluons (similar to the emitted photons in case of accelerated electric charges). In contrast to QED, the non-Abelian nature of QCD causes self interactions between the gluons (see Figure 1.1), which carry a color charge by themselves. This causes QCD radiation to further emit radiation and so forth, causing a cascade of splittings until the energy of the radiated particles reaches the hadronization threshold where the particles become "soft"(=low energies) and the formation of color singlet hadrons sets in.

In high energy particle physics, there are different schemes available that try to provide an approximate modeling of the process described above based on the parton shower philosophy. The parton shower sum the leading contribution to all orders in perturbation theory, but only

in the limit of soft and collinear splittings, otherwise, the underlying approximation fails and diverges. This means in particular, that the showering algorithms preferably add soft and collinear radiation to the partons from the hard subprocess. On the one hand, this is beneficial since the computation of the matrix element "fails" for soft and collinear particles; hence, the parton shower algorithm allows to fill regions in phase space that are inaccessible for the matrix element. On the other hand, the emission of hard radiation with wide angles is suppressed. This is also related to the problem of matching and merging between matrix element and parton shower since both create and overlap in the phase space and hence give rise to double counting of events.

2.4.1 Jet Factorization and Parton Emission

Is it instructive to start the motivation of parton showers by considering collinear enhancements encountered in electron-positron annihilation $e^+e^- \to q\bar{q}$, keeping in mind that one of the striking features of parton showers is their property of being *independent* of the underlying process.

The leading-order or Born-level cross section $\sigma_{q\bar{q}}$ for $e^+e^- \to q\bar{q}$ can easily be computed based on Feynman rules in electroweak theory. To increase the number of particles in the final state, one needs to consider higher-order corrections to the leading-order term; at NLO, the $q\bar{q}$ system is accompanied by the emission of an additional gluon $e^+e^- \to q\bar{q}g$. This gives rise to a three-particle phase space that can be kinematically conveniently parameterized by the opening angle between the gluon and its mother quark θ as well as the energy fraction z defined as the ratio between the energy of the gluon and the initializing quark prior to the emission. The cross section for this process (neglecting contributions from non-divergent terms) is given by

$$\frac{1}{\sigma_{q\bar{q}}} \frac{\mathrm{d}\sigma_{q\bar{q}g}}{\mathrm{d}\cos\theta \mathrm{d}z} \approx C_\mathrm{F} \frac{\alpha_\mathrm{S}}{2\pi} \frac{2}{\sin^2\theta} \frac{1+(1-z)}{z}, \tag{2.6}$$

whereby $C_\mathrm{F} = 4/3$ is the color factor of quarks, which are elements of the fundamental representation of the QCD symmetry group. The NLO cross section according to Equation 2.6 has a divergence associated to the emission of very soft gluons ($z \to 0$) in the event independent of the opening angle with respect to the quarks. Furthermore, there are two angular divergences: one for the emitted gluon being collinear with the quark ($\theta \to 0$) and another one for the gluon being back-to-back with the quark ($\theta \to \pi$) or – equivalently – being collinear to the anti-quark. The collinear divergences are related to the propagators and the interference of the two diagrams entering the calculation.

With the application of yet another approximation $\frac{2}{\sin^2\theta} = \frac{1}{1-\cos\theta} + \frac{1}{1+\cos\theta} \approx \frac{1}{1-\cos\theta_1} + \frac{1}{1-\cos\theta_2}$, it is possible to separate the cross section according to Equation 2.6 into two disjoint terms each one of which is related to either one of the two aforementioned collinear divergences. This separation into two probability distributions is convenient because it allows for a simultaneous treatment of the gluon being associated to either one of the two quarks. Applying this approximation to Equation 2.6 and expressing it in terms of θ^2 (for reasons that will become apparent very soon), the cross section for the process $e^+e^- \to q\bar{q}g$ can be summarized as

$$\sigma_{q\bar{q}g} \approx \sigma_{q\bar{q}} \sum_j C_\mathrm{F} \frac{\alpha_\mathrm{S}}{2\pi} \frac{\mathrm{d}\theta^2}{\theta^2} \mathrm{d}z \, \frac{1+(1-z)}{z}, \tag{2.7}$$

whereby j represents the different partons. On closer inspection of Equation 2.6, it becomes apparent that the NLO cross section naturally factorizes into the Born-level contribution and a factor that accounts for the additional gluon and hence can be thought of as the probability for the additional emission of such a particle. Despite the Expression 2.7 being derived for the specific example of electron-positron annihilation, it can be shown that the factorization of the hard sub-process holds valid for *any* collinear emission of a parton j of flavor i with a momentum fraction z. This is a general feature of the squared matrix element in the limit of collinear radiation and therefore commonly referred to as *collinear factorization theorem*.

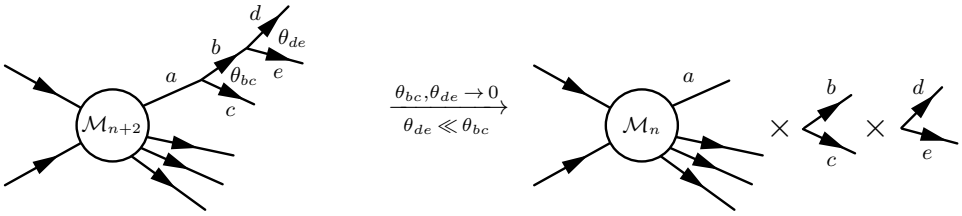


Fig. 2.1: The collinear factorization theorem for the example of a matrix element $\mathcal{M}_{n+2}$ of a given process with $n+2$ external legs and a final state a that branches (splits) two times with ordered splitting angles θ.

The collinear factorization theorem states that the $(n+1)$-parton differential cross section with an additional soft and collinear gluon in the final state can be *factorized* into the n-parton differential cross section prior to the splitting multiplied by an emission probability

$$\mathrm{d}\sigma_{n+1} \approx \mathrm{d}\sigma_n \frac{\alpha_S}{2\pi} \frac{\mathrm{d}\theta^2}{\theta^2} \mathrm{d}z \, \mathrm{d}\phi \, P_{i,jk}(z,\phi). \tag{2.8}$$

whereby the (Dokshitzer-Gribov-Lipatov-Altarelli-Parisi) *splitting functions* $P_{i,jk}(z,\phi)$ give the distribution of the fraction of energy of parton i carried by j. The factorization of the cross section into a lower-order term multiplied by a sequence of emission probabilities (as schematically illustrated in Figure 2.1) already indicates the iterative nature of the parton shower branching algorithm based on it.

Before continuing with the formal development of the parton shower formalism, it is worth to pause for a moment and reflect the importance of the seemingly inconspicuous term $\mathrm{d}\theta^2/\theta^2$.

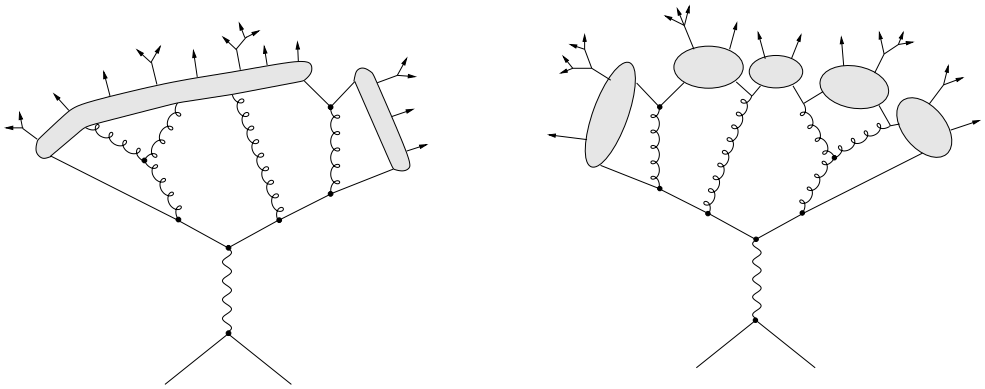
The factorization approximation according to Equation 2.8 already suggests an iterative approach to calculate the contribution due to multiple emissions to the cross section that can be considered as higher-order corrections. For reasons of simplicity, the following considerations only take the branching or emission of a gluon from a quark into account.

2.5 Jet Hadronization

The parton shower formalism described in detail in the previous section allows to increase the parton/particle multiplicity in an event by approximating higher-order corrections to the hard subprocess (2.3) through the resummation of dominant logarithms in each order in perturbation theory. Due to the infrared and collinear singularities present in the cross section, those leading logarithm corrections are connected to the emission of soft, i.e., low energetic gluons and the splitting of a parton into collinear daughters, e.g., $g \to gg$. Based on the factorization of the cross section in the collinear limit, the associated algorithm consecutively adds emissions to the main branch and performs parton splittings up to some infrared cut-off $\Lambda_{\mathrm{QCD}} \sim \mathcal{O}(100\,\mathrm{MeV})$ that corresponds to the "confinement scale" at which, due to the energy dependence of the strong coupling constant, non-perturbative effects start to contribute or even dominate. From this moment forth, a perturbative treatment of the evolution of the event is not possible anymore and one usually makes use of effective models to describe the formation of color-singlet hadronic jets phenomenologically. Two of the most commonly used hadronization models, the so-called "(Lund) string model" as well as the "cluster-based model", are subject of this section.

Before taking a closer look at the aforementioned hadronization models, it should be noted that the two terms "hadronization" and "fragmentation" model are often used interchangeably while strictly speaking denoting non-identical concepts.

Following the definition according to B. R. Webber (one of the creators of the HERWIG event generator and its cluster-based hadronization model), the term *factorization "[...] refer[s] [to] inclusive hadron spectra, for which factorization 'theorems' are available[.] [...] [that] allow[s] predictions to be made without any detailed assumptions concerning hadron formation."* [102]. This statement refers to the fragmentation functions $D_p^h(z, p_{\mathrm{T}})$, which describe the probability



(a) String model *(b)* Cluster model

Fig. 2.2: Visual illustration of the two commonly used hadronization model: the string (*left*) and the cluster (*right*) model (adapted and modified from Ref. 6, Fig. 5.14, p. 189 and Fig. 5.15, p. 190).

of finding a hadron h with an energy fraction z of the parton p with a transverse momentum p_{T}. Like the parton distribution functions, the fragmentation functions are not perturbatively calculable; however, the scaling/evolution, i.e., the z dependence, is described by the DGLAP equation

$$\mu \frac{\partial}{\partial \mu} D_p^h(z, p_{\mathrm{T}}) = \sum_j \int_z^1 \frac{\mathrm{d}z'}{z'} P_{ji}\left(z, \alpha_{\mathrm{S}}(z)\right) D_p^h(z, p_{\mathrm{T}}), \tag{2.9}$$

with P_{ji} denoting once again the Altarelli-Parisi splitting functions. Of course, the fragmentation functions itself are inaccessible, but the summed integrand $F_p^h(z, p_{\mathrm{T}}) = \sum_i \int_z^1 \frac{\mathrm{d}z'}{z'} f_i(p_{\mathrm{T}}) D_i^h(z, p_{\mathrm{T}})$, with $f_i(p_{\mathrm{T}})$ being the fraction of parton i at transverse momentum p_{T}, can be measured experimentally.

In contrast to fragmentation – once again quoting B. R. Webber –, *hadronization "refer[s] specifically to the mechanism by which quarks and gluons [are] produced in hard processes form the hadrons that are observed in the final state."* [102]. Contrary to fragmentation, which still contains perturbative elements through the evolution equations of the fragmentation functions, hadronization is an inherently non-perturbative process that completely relies on effective models or results from lattice QCD.

Within the context of this thesis, the term hadronization always refers to the semantic content stated in the last paragraph, i.e., the transition from the partonic final state to the actual hadronic final state that corresponds to the configuration of color-singlet states as "seen" by the detector.

(Lund) String Model

Due to the non-abelian nature of the $SU(3)$ gauge group and the resulting charge carrying gauge bosons of QCD, the beta function is negative, which causes the gauge coupling of the interaction to decrease in magnitude with an increasing energy scale; a phenomenon known as *asymptotic freedom* (see Chapter 1.1.3). In case of long-distant physics (associated with low energies), however, the coupling strength increases and eventually becomes ill-defined if $\alpha_{\mathrm{S}} \sim 1$ whereupon perturbation theory is no longer applicable.

The *string-based* family of hadronization models is based on the linear limit of the *static* $q\bar{q}$ potential as it can be calculated in lattice QCD from the Wilson loop. The phenomenological potential that describes the "force" between a quark and an anti-quark as a function of their relative distance r is given by

$$V_{q\bar{q}}(r) = V_0 - \frac{a_{q\bar{q}}}{r} + k_{q\bar{q}} r, \tag{2.10}$$

whereby $a_{q\bar{q}}, k_{q\bar{q}} \in \mathbb{R}$ are constants that need to be determined experimentally or calculated in lattice QCD. According to Equation 2.10, for sufficiently large spatial separations, the potential energy between the quarks grows linearly $V_{q\bar{q}}(r) = V_0 + k_{q\bar{q}}r$ with their distance giving rise to a linear confinement of the two partons. The potential energy of the $q\bar{q}$ system can be provisional pictured as a "string" or a "color (flux) tube" connecting the particles in space. This observation forms the basis of the string hadronization model that generally assumes that the Coulomb term which is dominant at short distances, can be neglected in favor of a linearly rising energy potential.

Among the earliest approaches towards a phenomenological description of the formation of hadrons was an (very entertaining!) article by Kenneth A. Johnson titled *"The Bag Model of Quark Confinement"* [103] in which he modeled the confinement of partons *"analogous to [...] bubbles in a liquid"*. One of the first string-based formation models in today's sense is the one proposed by X. Artru G. Mennessier in their trend-setting paper *"String model and multiproduction"* published in the year 1974. The suggested model, however, had a few shortcomings, especially the prediction of an nonphysical mass spectrum that required the application of a priori mass cuts in order to match the distributions measured in data.

To address the problems of the hadronization models common at that time, the development of the *Lund string model* started in the year 1977, with the framework being available six years later in 1983 [104], which – to this day! – is one of the most successful hadronization models and still widely used today by many state-of-the-art event generators.

The vivid idea behind the string-based approach is illustrated in Figure 2.2a. In case of a $q\bar{q}$ pair, the two particles quickly moving apart from each other with the potential energy increasing linearly. To put things into perspective, the "tension" of the "flux tube" is $k_{q\bar{q}} \sim \mathcal{O}(1\,\mathrm{GeV/fm})$, which defines a characteristic length of $1 - 5\,\mathrm{fm}$ (corresponding to the order of hadrons) of the string at which the system splits into two new color-singlets [105]. This simple picture becomes more complicated if one includes additional partons like, for instance, a gluon as a NLO correction to the $q\bar{q}$ matrix element. For this reason, one usually works in the already mentioned large N_c limit to ensure that each parton in the system only comes with exactly one unique color partner to which it is connected by a tube. With this approximation, the procedure closely follows the one described in case of the simple $q\bar{q}$ system. The method is applied consecutively to the newly emerged parton systems till a final configuration of (usually excited) hadrons has been formed.

Cluster-Based Model

An alternative family of hadronization models is based on the observation that for sufficiently low energy scales compared to the hard subprocess the emitted partons in the parton shower algorithm can be naturally clustered in colorless groups of quarks anti-quarks, hence, commonly referred to as *cluster-based* hadronization models. This discovery was made by D. Amati and G. Veneziano and published in 1979 in their article *"Preconfinement as a Property of Perturbative QCD"* [106]. The term *preconfinement* in this case (not to be confused with "confinement") refers to the observation of perturbative shower emissions in QCD to form the aforementioned color-singlet "clusters" of partons, with the property to produce an universal invariant mass distribution that only depends on the scale of the evolution as well as Λ_{QCD} and hence asymptotically independent of energy or the production process. The cluster-based model therefore, simply speaking and summarizing, relies on the observation that color-singlet pairs tend to be close in phase space in the large N_c approximation in which their is a well-defined color connected partner for each final state parton.

A clear motivation for the cluster-based hadronization model is given by the invariant mass spectra of light meson states, i.e., color-neutral, bounded quark-antiquark states. In this spectrum, the lightest meson states appear as narrow resonances, while the heavier ones (with masses above roughly $1.5\,\mathrm{GeV}$) are broader, which suggest *a continuum of overlapping states* [107]. The already mentioned preconfined color-anticolor pairs are then projected directly onto this continuum of meson states, which gives rise to clusters. Those mesons are usually excited and further decay into well-known stable hadron states.

One major advantage of the cluster-based approach to hadronization compared to the string

model introduced in Section 2.5 is the significantly reduced number of parameters needed to tune the model (in principle, all properties are fully determined by the mass spectrum of the clustered color-singlet states). However, despite working well for most hadrons, it was quickly realized that the cluster-based approach tends to produce hadrons that are too soft. This problem was solved by introducing additional parameters to the model.

At this stage, it is worth pausing for a moment to highlight the differences the *string-* and the *cluster-based* model that are conceptually and philosophically quite difference. The former one focuses on the non-perturbative aspect of the confinement of partons, while the latter one stresses on the perturbative phase space of the shower evolution [107].

2.5.1 Decay to Final-State Hadrons

The hadrons produced at the end of the hadronization procedure according to the (Lund) string model or the cluster-based approach (or alternative models not mentioned at this occasion) are referred to as *"primary" hadrons* that are color-singlet states, i.e., being – in principle – observable. However, most of the generated hadrons are in an excited energy state and therefore highly unstable given the time scales considered relevant for a measurement. As a consequence, they quickly decay in to lower energy configurations (mostly pions and kaons) with lifetimes long enough in order to be detected in an experiment. This subsequent chain of consecutive decays to stable hadrons must be simulated posterior to the hadronization step outlined in Chapter 2.5. This part of the event generation sequence is often an underestimated ingredient to the full simulation chain; however, its modeling usually has a *significant* on the final spectra and particle yield.

For instance, the process $B^{*0} \rightarrow \gamma B^0$ involves electromagnetic decays (involving the emission of a photon), weak mixing of B mesons, weak decay (involving a neutrino), a strong decay of hadrons into hadrons etc. Given the energy scales of hadrons, the decay to stable particles is an inherently non-perturbative process which does – just as the hadronization step – rely on phenomenological models as well as information gathered in plenty of measurements. This information is usually encoded in particle summary tables that summarize the current knowledge given the available experimental measurements. However, given the vast amount of possible decay channels, the branching fractions of many processes are only known with a limited precision, which is why in many cases the Monte Carlo must be tuned to data to obtain reliable results.

2.6 Detector Simulation and Reconstruction

A detector that has been build to measure the final-state particles produced in hadron-hadron interactions does not measure the particles directly but their "traces" left in the material due to energy deposition in calorimeters or ionization in (Silicon) pixel cells (hits and tracks). From the measured energy, transverse momentum, track curvature, position in η–ϕ etc., the four-momentum of the particle(s) may be reconstructed. However, the measurement of the detector is not perfect since the detector has a finite granularity as well as limited acceptance in its materials. Furthermore, a detector does not cover the entire region $(\eta, \phi) \in \mathbb{R} \times [0, 2\pi)$ and therefore potentially misses some decay products.

To account for the detector effects, the actual physical processes occurring in its subcomponents are simulated with Monte Carlo methods (see Section 2.1.2). This includes, e.g., the simulation of energy deposition, ionization and much more. After this step the generated data is valid only for this particular measuring device since those simulations depend on the actual detector architecture (number of layers, material etc.) that is used in the experiment

The last step is the identification and/or reconstruction of particles in the final state for further analysis.

Part II

Methodology

With the first part of this thesis being concluded, all theoretical foundations necessary for understanding the physical background and motivation of the analysis have been established. This introduction was rather general and only in a broader sense related to the measurement(s) presented in this thesis. The following second part now ties with the previous one by introducing analysis-specific methods and content that are directly (or indirectly) relevant for the studies presented here.

The first Chapter 3 of this part provides an introduction of the experimental setup (i.e., the LHC and the ATLAS experiment) that was used to record the proton-proton collision data, which has been analyzed in the scope of this dissertation.

With the accelerator and the detector being introduced, the next Chapter 4 explains how physical objects (e.g., leptons, jets etc.) are derived based on the raw collision data. The main focus of this chapter is on tracks and jets constructed thereof, as this (besides the calorimeter clusters) is the only information directly used in this analysis. Furthermore, ATLAS' calibration scheme for jets introduced, which is one of the main sources of uncertainty in this measurement.

Chapter 5 briefly introduces the concept of (detector) unfolding, which is used to obtain a cross section measurement at particle level corrected for detector and acceptance effects.

In Chapter 6, there is a change of subject by providing a very general introduction into the vast field of machine learning, as one measurement of the triple-differential multijet cross section is done as a function of a neural network score-based observable.

Finally, the last Chapter 7 of this second part introduces the method that is used to disentangle quark- and gluon-related contributions in a model independent way from the final hadronic cross section measurement to obtain the underlying probability distributions and fractions of quark/gluon-initialized jets.

3. Experimental Setup[1]

The measurement described in the thesis at hand is based on proton-proton collision events that have been recorded with the ATLAS detector over a period of four years of data-taking between 2015 and 2018 (inclusive) at the world's largest particle physics laboratory at CERN. Conducting experiments at this level of complexity requires a *tremendous* amount of organization, communication and cooperation between thousands of scientist around the world from dozens of nations who all share a common goal and passion: to gain a better understanding of how nature works at the most fundamental level of material being. Therefore, all insights gained within the scope of this analysis are based on the indispensable groundwork done by a multitude of people who should be acknowledged at this point once again. Today's science always is a joint effort as already mentioned in the acknowledgments of this thesis.

The purpose of this chapter is to provide an overview of the experimental apparatus (ATLAS detector) as well as the research facility at which the data has been recorded that forms the basis of the analysis outlined hereinafter. The first section (3.1) of this chapter conceptually introduces the LHC accelerator complex as well as the particle physics laboratory CERN at which the experiment is located and operated. By doing so, the section is limited to the most important aspects without going too much into detail. Furthermore, it introduces some important accelerator concepts that will provide helpful to understand the results reported in this thesis. Following (3.3), the actual experimental apparatus, i.e., the ATLAS experiment/collaboration as well as the detector, is introduced along with its individual components – whereby the main focus is on the detector parts that are most relevant for this measurement.

3.1 The Large Hadron Collider at CERN

The Large Hadron Collider (LHC) is the world's *largest* superconducting hadron-hadron accelerator located at the Centre for European Nuclear Research (CERN, French: *Conseil Europén pour la Recherche Nucléaire*), which is situated near the French-Swiss border in the vicinity of Geneva, Switzerland. At the time of writing this report (last quarter of 2022), the collaboration has twenty-three member states (of which twelve are founding Members of the organization) and several other nations around the world that hold an associated membership or an observer status, respectively. The main focus of CERN comprises the development and construction of high-energy accelerator-based particle physics experiments (of which the LHC is the largest for the time being) as well as their maintenance and further improvement along with the supply of computing infrastructure like, e.g., the (Worldwide) LHC Computing Grid ((W)LCG).

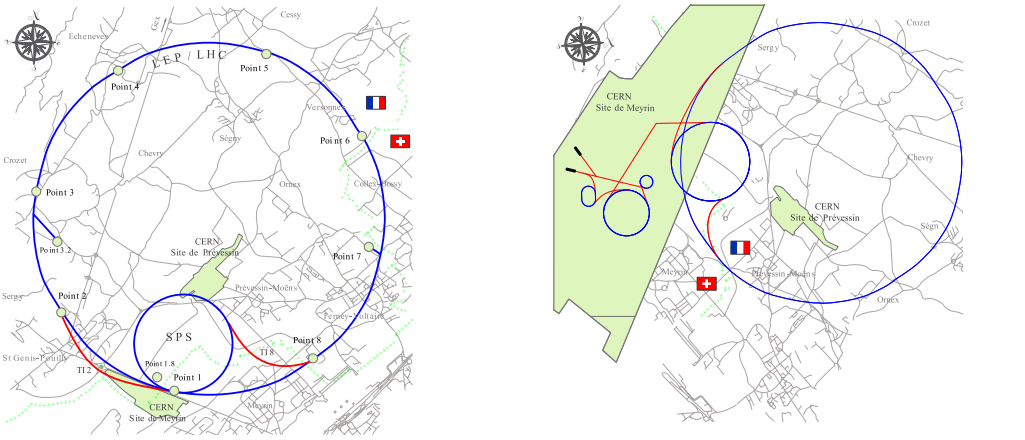
The LHC is a two-ring proton-proton synchrotron accelerator with a circumference of roughly 27 km that started running in November 2009 and currently operates with a nominal centre-of-mass collision energy of $\sqrt{s} = 13\,\text{TeV}$, which makes it the highest energy particle accelerator in the *world*. It supplies four main experiments – lined up along the accelerator tunnel like pearls on a string – at dedicated collision points as shown in Figure 3.1a below. As it can be seen in Figure 3.1a, the accelerator ring is subdivided into *octants*, each one fulfilling

[1] This chapter may safely be skipped provided that the reader is already familiar with the Large Hadron Collider as well as the ATLAS experiment.

their own purpose.

Going clockwise and starting from the first (1) collision point, the first collaboration one encounters is the ATLAS (A Toroidal LHC ApparatuS) experiment, conveniently located at the Meyrin site of CERN. This side serves as a collision point for proton-proton, proton-lead, and lead-lead collisions; accordingly, the ATLAS experiment studies all of them with the main focus being on proton-proton-related physics; however, collision point two (2) is reserved for proton-lead and lead-lead collisions which is why ALICE (A Large Ion Collider Experiment) build its nests at this cozy place close to the Jura Mountains. The third (3) access point does not hold any larger particle physics experiment, but serves for monitoring and beam cleaning purposes to remove particles that stray from the desired beam trajectory. The fourth (4) octant provides another control point that hold the Radio Frequency (RF) cavities which are responsible for accelerating the LHC beams to the operational centre-of-mass energy. Point number five (5) is, once again, a collision point for proton-proton, proton-lead, and lead-lead collisions at which the CMS (Compact Muon Solenoid) experiment is located. Service point six (6) and seven (7) are responsible for beam dumping and beam cleaning, respectively. Last but not least, the final octant (8) is a collision point for proton-proton collisions exclusively where the LHC*b* (Large Hadron Collider *beauty*) experiment is doing its studies. The two transfer lines TL2 and TL8 in Figure 3.1a establish a connection between the Super Proton Synchrotron (SPS), which serves as a pre-accelerator and final injector, and the LHC.

The LHC accelerator ring is located in the 26.7 km tunnel that was formerly constructed for



(a) Geographic location of the LHC accelerator ring including the SPS and access points at the French-Swiss border (dashed line) and the CERN area.

(b) Simplified illustration of the LHC acceleration and injection chain. The components are *not* drawn to scale nor does their position match the exact location geographically.

Fig. 3.1: Schematic representation of the LHC ring and the CERN research facility (*left*) including a simplified illustration of the acceleration and injection chain (adapted and modified from Ref. 7, Fig. 1, p. 4).

the Large Electron-Positron (LEP) collider at CERN, which operated between 1989 and 2000. Contrary to LEP and Tevatron at the Fermi National Accelerator Laboratory, United States (that both collide opposite charged particles), the LHC is designed to work with hadron-hadron collisions instead; hence, it can not use the same magnetic field to accelerate the counter-rotating beams. Due to space limitations in the tunnel system, however, it is not possible two support separate rings of superconducting bending magnets for the two counter-rotating beams. This constraint motivates the so-called "2-in-1" concept used for the superconducting niobium-titanium LHC bending magnets according to which the two pipes are contained in the very same cryostat with the opposite running beams being held in their particular orbits by two coupled magnetic fields with opposite polarity. Due to the aforementioned serious technological challenges, the magnet system of the LHC is of considerable complexity consisting of 1232 main superconducting dipoles magnets with field strengths close to $\mathcal{O}(10\,\mathrm{T})$ for bending and tightening the particles' trajectories. Taking the entire injection chain (with the LHC as the

final step) into account, the entire acceleration system comprises more than 50 different *types* of finely tuned magnets.

The process of accelerating the protons (or heavy ions) to the nominal collision energy is done via a chain of successive pre-accelerators that gradually increase the energy of the colliding beams. A simplified illustration of the main components of the LHC injection chain is shown in Figure 3.1b. The protons[2] are obtained from Hydrogen atoms that are exposed to an intense electric field for ionization. The first element in the actual chain is the "new" linear accelerator LINAC 4, which replaced the old LINAC 2 accelerator in 2020. The linear accelerator accelerates the protons to an energy of roughly 50 MeV before they are injected into Proton Synchrotron Booster (PSB). In the SPB, the protons are further accelerated up to an beam energy of 1.4 GeV and injected into the Proton Synchrotron (PS), which is CERN's first synchrotron and the subsequent element in the acceleration chain. At the end of the PS, the protons reach an energy of approximately 26 GeV. The penultimate step in the injection chain is the Super Proton Synchrotron (SPS) that accelerates the protons to an energy of 450 GeV in a 7 km ring. It is worth mentioning that the SPS already started operating in 1976 and was involved in several groundbreaking discoveries such as, e.g., the discovery of the $W^{\pm}$ and Z bosons in the year 1983 at the UA1 (Underground Area 1) experiment [108]. Following the SPS, the particles are finally injected into the main LHC ring where they are accelerated to the final nominal energy that depends on the respective year-dependent run of the machine. The entire acceleration chain described in this paragraph takes roughly 20 minutes for the particles to reach the final energy of the LHC.

At the time of commissioning during the first Run (2011–2012), the LHC reached a centre-of-mass energy of 7 TeV and 8 TeV, respectively. During the second Run (2015–2018), the collision energy was increased "significantly" to $\sqrt{s} = 13\,\text{TeV}$, which is close to the designed energy of the LHC of $\sqrt{s} = 14\,\text{TeV}$. Once the particle bunches are circulating in opposing orbits at the nominal centre-of-mass energy, they are deflected to collision course using specific quadrupole magnets placed at the four collision points highlighted in Figure 3.1a.

The result of the previously described injection is not a continuous stream of protons in the LHC ring, but an ensemble of hadrons referred to as (proton) *bunches*, whose final geometry depends on the accelerator-specific parameters such as the oscillation frequency of the longitudinal electromagnetic field defined by the RF cavities, which determines the spacing and the maximum amount of bunches being present in the ring at the same time.

This gives rise to another very important parameter – besides the collision energy – that characterizes the performance of an accelerator: the so-called (instantaneous) *luminosity* $\mathcal{L}$. The luminosity can be defined as a constant of proportionality between the number of events N in a given interval of time Δt and the physical, i.e., detector-independent cross section σ of a given process. This relation is captured in Equation 3.1 below

$$\frac{\mathrm{d}N}{\mathrm{d}t} = \mathcal{L}\sigma. \tag{3.1}$$

Based on the observation that the luminosity connects a detector-*dependent* and a detector-*independent* quantity, it is obvious that $\mathcal{L}$ must depend on the bunch-crossing frequency, the cross section of the two colliding bunches as well as the number of particles per bunch. This relation is described by the following formula

$$\mathcal{L} = \frac{N_{1,2}^2 n f}{4\pi\sigma_x\sigma_y} \cdot S, \tag{3.2}$$

whereby S denotes a geometric factor that accounts for deviations from a assumed Gaussian distribution of the transverse beam widths $\sigma_{x,y}$, and $N_{1,2}$ being the number of particles per bunch, which are assumed to be identical here. The number of protons for each bunch in the LHC is roughly $\mathcal{O}(10^{11})$ particles that are temporally spaced by 25 ns. Consequently, each bunch crossing results in *many* simultaneous proton-proton interactions at each interaction

[2]The acceleration and injection chain for heavy ions is basically the same as for protons with the difference of starting at a different linear accelerator LINAC 3 instead of LINAC 4

vertex whose final state particles are together reconstructed in what is referred to as a (collision) *event*. The average number of interactions for each bunch crossing is usually denoted by the Greek letter μ; the larger the average number of interactions μ, the more challenging the selection of interesting events becomes.

Aiming for high luminosities is desirable in order to increase the sensitivity to signals originating from new physical phenomena, which are usually associated with a low cross section, i.e., low production rates. This can be seen by rewriting Equation 3.2 in terms of the *integrated* luminosity $L := \int \mathrm{d}t\, \mathcal{L}(t)$ that is a measure of the number of collected data and hence often quoted on many results. The design luminosity of the LHC is $\mathcal{L} = 10^{34}\,\mathrm{cm}^2 s^{-1}$.

3.2 LHC and HL-LHC Timeline

The LHC successfully operates since 2008. In this time, it provided an answer to the important question whether the Higgs boson actually exist or not. With its existence finally being (officially) confirmed in July 2012, the Standard Model (see Chapter 1) is a consistent model of particle physics. However, based on experimental evidence, it is unlikely that the current SM does provide a complete description of nature – as a matter of fact: it does *not*. One of the biggest unanswered questions is, to name just one, the origin of Dark Matter whose existence is supported by several *independent* observations in astronomy (it should be noted that there are alternative explanations for the origin of Dart Matter besides the particle-based hypothesis like, for instance, modified gravity or superfluid Dark Matter). Other fields of studies involve extensions to the current Standard Model like, for instance, supersymmetry, or a mechanism that explains the generation of neutrino masses. Many of those theories predict the existence of new fundamental particles that scientist were hoping to find evidence for at the LHC. In all honesty, however, it should be noted that despite the undeniable success of the LHC – that exceeded its technical expectations –, no hints of new (fundamental) particles besides the Higgs have been observed. Of course, the high collision energy of the LHC allowed for plenty constraints on the mass of hypothetical particles and also excluded a large range of theories through "null result"[3], but no indication of new physics beyond the Standard Model was found – so far.

However, the operation of the LHC is far from over as illustrated in Figure 3.2 that shows the aimed scheduled of the LHC and its foreseen upgrade(s): the High-Luminosity Large Hadron Collider (HL-LHC). Already during the completed Run 1 and Run 2, the LHC showed

Fig. 3.2: Foreseen timeline of the LHC and the HL-LHC (adapted from Ref. [8]).

a performance far beyond expectations. Despite huge delays due to the global situation starting in 2020, Run 3 of the LHC started (almost) according to schedule, beginning 2022 after a two year lasting shutdown of the machine. Compared to Run 1, the amount of data that will be collected is expected to be roughly $\mathcal{O}(10^2)$ larger with an increase of the luminosity by an order of magnitude. Furthermore, several upgrades and extensions to the detector are included compared to the previous Run.

The third Run of the LHC will end in the year 2023/4. Due to the very harsh environment during Run 3 of high radiation dose (mainly due to increased pile-up), most detector components of the individual experiments will be at the end of their life cycle and need to be completely replaced for the follow-on project, i.e., the HL-LHC that was announced as the top priority

[3] "Nill results" are by no means worthless; however, they provide very little feedback to constrain theories.

already in 2013 and is supposed to start operating from the end of 2027. The centre-of-mass energy can not be further increased due to the radius of accelerator ring and the upper limit on the field strength of the bending magnets, but the luminosity is intended to rise by yet another order of magnitude to increase the potential for discoveries of rare processes. This environment presents scientist and engineers with completely new challenges in terms of hardware and software. Certainly interesting times to be a particle physicist!

3.3 The ATLAS Experiment

The ATLAS detector is a hermetic general-purpose particle detector located at the experimental cavern at access point 1 (see Figure 3.1a showing the different access points of the accelerator complex) at the LHC ring. The detector design is coaxial with respect to the beam pipe and cylindrical in shape with a coverage of almost the entire solid angle 4π around the interaction point at the center of the detector. It extends about 44 m in length and 25 m in height and weights about 7000 tonnes, which makes it the largest (yet not the heaviest) particle detector at the LHC.

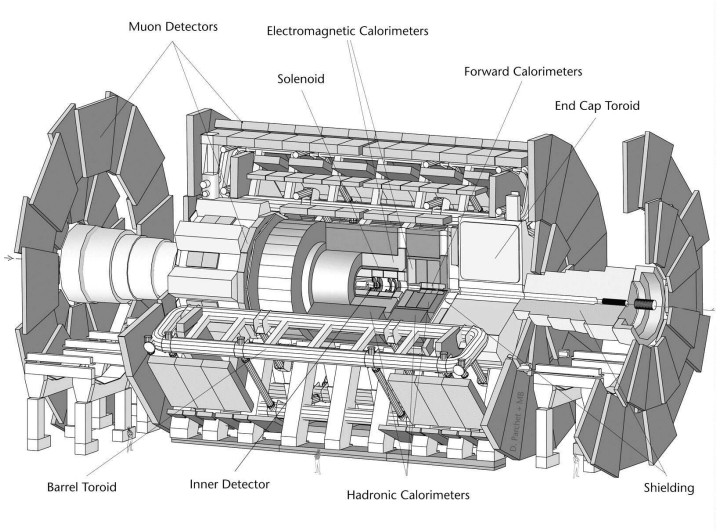


Fig. 3.3: View on a computer-generated schematic representation of the whole ATLAS detector including its individual sub-components (taken from Ref. [9]).

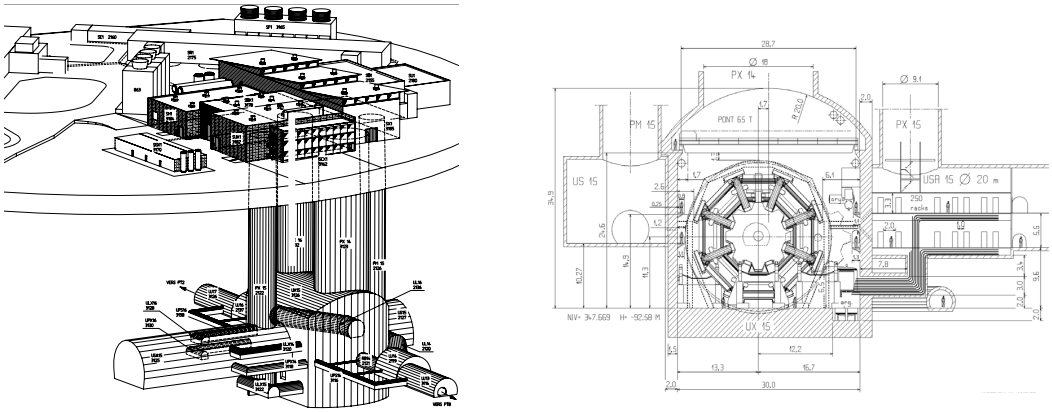
ATLAS uses a right-handed coordinate system with its origin at the nominal Interaction Point (IP) in the centre of the detector and the z-axis along the beam pipe. The x-axis points from the IP to the centre of the LHC ring, and the y-axis points upward. Cylindrical coordinates $(r, \phi) \in \mathrm{R} \times [0, 2\pi)$ are used in the transverse plane, with ϕ being the azimuthal angle around the z-axis. The pseudorapidity η is defined in terms of the polar angle θ as $\eta = -\ln\tan(\theta/2)$.

The ATLAS detector is build in cylindrical layers parallel to the beam line as well as several disks perpendicular with respect to the beam. A simplified view of the whole ATLAS detector and its individual components is provided by Figure 3.3.

Each subcomponent of the detector is designed for the identification and the reconstruction of specific types of particles or dedicated to certain fundamental interactions. The innermost layer – which is closest to the interaction point – serves as a tracking detector and is engulfed with a solenoid magnetic field of 2 T to bend the particles' trajectories for momentum reconstruction in the pseudorapidity range $|\eta| \leq 2.5$ (this will be relevant for this analysis and its event/object selection as all jet substructure observables are reconstructed from tracks matched to calorimeter-based jets). Following the inner detector, the subsequent layers are the electromagnetic and hadronic calorimeters for energy measurement (and particle identification) of the absorbed particles. For the calorimeters of the ATLAS detector, a distinction is made between the

barrels, parallel to direction of the beam, and the end-caps that are perpendicular to the beam; together, they cover a pseudorapidity range of $|\eta| < 4.9$. The outermost layer of the ATLAS detector is a muon spectrometer taking account of the fact that muons are Minimum Ionizing Particles (MIPs), which deposit little to none energy in the calorimeter system. The muon spectrometer is placed in a separate toroidal magnetic field of 0.5 T for the barrel and 1 T in the end-caps that allows for the reconstruction of the muons' momentum within a range of $|\eta| < 2.7$. Besides the aforementioned main detector components, the ATLAS experiment is equipped with luminosity detectors, beam conditions monitors as well as the ATLAS Forward Proton Detectors (AFP and ALFA) also known as "Roman Pot" detectors (not present in Figure 3.3).

The detector itself is located in the cavern UX15 at point 1 roughly 100 meters below ground in the neighborhood of the main CERN site very close to the village of Meyrin, Switzerland. A schematic view of the underground ATLAS installation is sketched in Figure 3.1a, which shows the actual cavern that houses the detector as well as installation shafts for surface access. Cavern UX15 is almost completely filled out by the detector (which deprives ourselves of a



(a) *(b)*

Fig. 3.4: The ATLAS experimental Area and cavity at access point 1 (*left*, taken from Ref. 10, Fig. 9, p. 9) and a cross section view of the ATLAS cavern including the outer magnet system (*right*, taken from Ref. 11, Fig. 10.1, p. 161).

nice overall perspective of the measuring device unfortunately) as can be seen in Figure 3.4b, which shows a cross section of the detector mounted in the cavern. The control room(s), which monitor the functionality of the ATLAS detector and performs some basic data quality checks during operation of the LHC, is not located in the cavern but in building SCX1/3162 on ground level. In Figure 3.4a, the two outgoing arrows (labeled as "VERS") indicate the direction of the beam pipe pointing towards point 2 (PT2) and point 8 (PT8), respectively.

3.3.1 Brief History of the Project

Driven by the unified theory of the electroweak interaction by Glashow, Salam and Weinberg [52**?**] and the estimation of the weak mixing angle, $\sin^2\theta_W$, based on data of muon neutrino and anti-neutrino scattering from the bubble chamber Gargamealle (see, e.g., early calculations by Carl H. Albright [109]), the mass of the electroweak bosons was expected to be in the order of something like $\mathcal{O}(10\,\text{GeV})$, which required future colliders capable of probing such energies. While the $W^\pm$ and Z bosons have already been discovered at the S$p\bar{p}$S in 1983, the detailed exploration of the electroweak energy scale was one of the main motivations for constructing the aforementioned LEP collider at CERN. It was realized that in order to reach higher energies, a proton-proton collider was needed – the idea of the LHC was born and the discussion of possible detector concepts started.

The first concepts of the detectors at the future LHC have been presented and discussed at

the ECFA-CERN workshop in Lausanne, Switzerland in 1984 [110]. Based on studies presented at this workshop, experimental and theoretical physicists from many Institutes worked together to develop future detector concepts. Based on those attempts, the ATLAS Collaboration was launched and presented their concept for a *"[...] General-Purpose pp Experiment at the Large Hadron Collider at CERN"* [111] in the Letter of Intent (LoI) in the year 1992. The LoI was submitted to the CERN LHC experiments committee that inspected the proposal with respect to its scientific yield as well as its financial and technical feasibility. Some ideas presented in the initial concept had to be adjusted and further developed; however, just two years later, with the submission of the technical proposal in 1994 [112], the basic concept of the measuring device was settled with the final approval of the project by the LHC Experiments Committee in 1997.

With all organizational obstacles being out of the way, Institutes around the world began building on the detector components. Just one year later, the excavation of the cavern for the ATLAS detector began and was completed in 2003 with the installation of the detector starting soon after. On October 4, 2008, the construction of the ATLAS detector was officially completed, which was celebrated by a big ceremony at CERN's magnet testing facility.

With the construction of the ATLAS detector being completed, it was about time to test the functionality of the apparatus with some first collision events recorded on November 23, 2009 with an energy of 0.9 TeV. This "baptism of fire" was a great success and the very first time that real particle tracks from the interaction point of the detector had been reconstructed. This moment of great awe was commented by the ATLAS spokesperson, Fabiola Gianotti, as follows: *"[t]his is great news, the start of a fantastic era of physics and hopefully discoveries after 20 years' work by the international community to build a machine and detectors of unprecedented complexity and performance."* First collision events from proton-proton interactions at a centre-of-mass energy at 7 TeV followed just one year later – the operation of the detector for physics analyses had begun.

3.3.2 General Requirements and Design Criteria

The design concept of the general purpose ATLAS detector (rudimentary outlined in the previous section and further detailed in the following ones) reflects the physical objectives defined by the collaboration as well as the constraints set by the LHC accelerator. The latter ones basically determine the design luminosity and the interaction rate, compelling strict limits on the electronics and sensors in terms of speed and radiation hardness, ATLAS' complicated trigger system and pile-up suppression due to, e.g., inelastic events occurring simultaneously to the hard interaction.

Aside from constraints set by the LHC accelerator, the design of the ATLAS detector is driven mainly by concrete physical processes and new phenomena expected to be observed at the electroweak energy scale, which served as a benchmark.

First and foremost, the main physics goal of the ATLAS (as well as CMS) experiment was the discovery of the Higgs boson to fix the aforementioned inconsistency in the Standard Model that predicted all particles to be mass less. Therefore, the expected properties of the Higgs bosons was an important benchmark process to set constraints on the design of the ATLAS detector. Since the mass of the Higgs was a free parameter at that time – covering a wide range of possible energies –, mass resolution was of utter importance. Hadronic decay modes are dominated by the overwhelming QCD/multijet cross section and therefore were not considered as potential discovery channel for the Higgs boson. The Higgs decay into leptons and photons on the other hand provides a clean signature for discovery. This requires the ATLAS detector to be equipped with a high resolution electromagnetic calorimeter with dedicated electron and photon identification. Furthermore, the calorimeter was supposed cover almost the entire solid angle around the interaction point to catch potential signals originating from Vector Boson Fusion (VBF) WW and ZZ events anticipated to be important for Higgs masses close to 1 TeV. The only hadronic final state of the Higgs decay expected to be experimentally accessible is $H \to b\bar{b}$ due to b(-jet)-tagging allowing for a certain degree of discrimination with respect to the QCD background. Since b-tagging fundamentally relies on the identification and measurement of secondary, displaces vertices, precise vertex reconstruction was a crucial requirement in

terms of setting constraints on the inner part of the detector for the reconstruction of charged particles' trajectories.

Supersymmetry was (and to some extend still is to this very day) another important motivation for the construction of the LHC and the ATLAS detector, but one should not be too vain to admit that the expectations were *undoubtedly* disappointed in this regard. In general (whereby different supersymmetric models sometimes differ significantly), the decay of supersymmetric particles was expected to be accompanied by a large imbalance of energy in the detector reconstructed as missing transverse energy $E_{\mathrm{T}}^{\mathrm{miss}}$, which also is a signature expected for events involving Dark Matter particles – provided that they exist. A precise measurement of transverse momentum imbalance in an event requires an almost *hermetic detector* to reconstruct all final state object associated with the hard interaction as well as an electromagnetic and hadronic calorimeter system with high energy resolution.

Some physicists hoped to find signatures of extra spatial dimensions, which could manifest itself through effects from quantum gravity. Roughly speaking, the energy scale at which quantum gravity would provide non-negligible contributions would decreases with the number of "hidden" spatial dimensions, which allows to set a limit on the number of dimensions for energies probed by the LHC. Due to the expected properties of gravitons, signatures from quantum gravity are expected to be, once again, a large missing transverse momentum in the event as well as the evaporation of small and short-living black holes due to Hawking's radiation [113] – to this very day, no traces of such exotic objects have been found.

There are plenty more motivations from physics like, for instance, forward electron signatures, hadronically decaying tau leptons and – not forgetting! – signatures expected from heavy ion collisions. In summary, it can be stated that the ATLAS detector should meet the following requirements in order to fulfill its physical objectives:

- a fine-granularity *Inner Detector* system for (partial) particle identification and efficient charged particle track reconstruction up to high p_{T} for precise lepton measurements as well as high b-tagging efficiency;
- a superb *electromagnetic calorimeter* for precise energy and position measurements along with particle identification based on electromagnetic showers;
- a *hadronic calorimeter* system for the measurement of hadronic jets and missing transverse energy;
- a dedicated *muon spectrometer* for high-precision tracks measurements and identification of muons;
- finally, the detector should be almost *hermetic*, i.e., full coverage of the entire solid angle to capture all event remnants;
- a trigger system with high efficiency for physical processes of interest and capable of triggering particles at low thresholds.

The criteria mentioned above suggest a rather traditional detector design with a forward–backward symmetric cylindrical geometry of alternating detector layers. The following sections provide a short introduction and description of each one of the aforesaid detector components.

3.3.3 The Inner Detector (ID)

The Inner Detector (ID) of the ATLAS experiment is, as the name already suggests, the innermost sub-component of the detector closest to the interaction point. It consists of three different systems of sensor technologies, i.e., the *Pixel Detector* (PD), the *Semi-Conductor Tracker* (SCT) and the *Transition Radiation Tracker* (TRT), all of which are placed in a solenoid magnetic field of 2 T that is oriented parallel to the beam line. A sectional view of the barrel region of ATLAS' Inner Detector is shown in Figure 3.5. The system (as a whole) measures the trajectories of *charged* particles with a transverse momentum p_{T} above the nominal threshold of 500 MeV (lower thresholds are possible though) inside the pseudorapidity range $|\eta| < 2.5$. A precise reconstruction of tracks is crucial for precise angular information

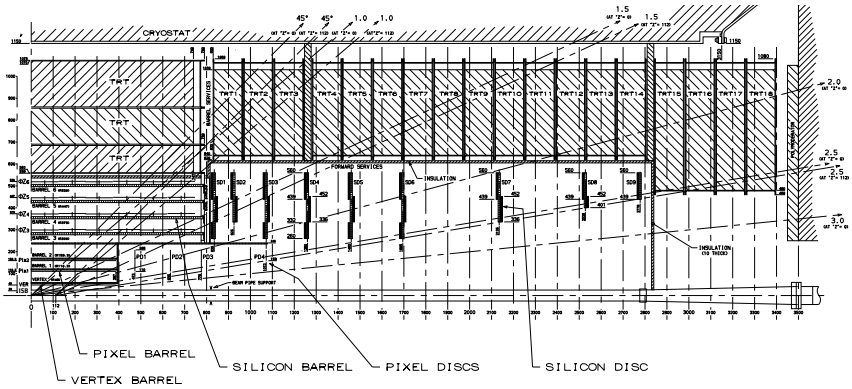


Fig. 3.5: View on a cross section of the ID engineering layout through the beam axis of the ATLAS Inner Detector system in ascending order: Pixel Detectors, Semiconductor Tracker and Transition Radiation Tracker(has been adopted from the TDR of the ID from Ref. 12, Fig. 1.1, p. 6 and was modified to account for the extended ISB).

as well as the reconstruction of primary and secondary vertices for, i.e., the identification of b-hadrons as well as track-based pile-up suppression and the soft term of the missing transverse energy. The designed tracking resolution of the ATLAS detector is given by

$$\frac{\sigma_{p_\mathrm{T}}}{p_\mathrm{T}} = 0.05\,\% \cdot p_\mathrm{T} \oplus 1\%, \tag{3.3}$$

with the resolution decreasing for high-p_T jet due to the reduced curvature of the trajectories. Furthermore, it allows for electron identification within the pseudorapidity range $|\eta| < 2.0$ in the TRT as well as by combining track information from the ID with characteristic energy clusters in the electromagnetic calorimeter (see track-to-cluster matching).

As discussed later, the multi-differential multijet cross section measurement described in this thesis is based on jets obtained from matching tracks to topo-cluster jets. Therefore, the Inner Detector is of particular importance for this analysis and hence deserves a discussion in adequate detail.

The following Sections provide a quick overview of the individual sub-components that build the ID of the ATLAS detector. For a very detailed description of the Inner Detector, the interested reader is referred to the technical design report of the sub-detector [12] (which leaves nothing to be desired) provided by the ATLAS Collaboration.

The Pixel Detector (PD)

The Pixel Detector is the innermost component of the Inner Detector, just a few millimeters apart from the beam line at its closest point up to a radial distance of 122.5 mm in the barrel and 150.0 mm in the end-caps. It consists of four cylindrical barrel layers and three disk-shaped end-cap layers that together extend over a pseudorapidity range of $|\eta| < 2.5$.

The Pixel detector is assembled from modules based on high-granularity semiconductor sensors as well as front-end electronics for signal processing. The signal generation is based on charge ionization due to charged particles passing through the active material of the sensor that results in a measurable electric signals. If the measured voltage drop is above a pre-defined threshold, a *hit* is registered.

Due to its proximity to the interaction point and the considerable dose of radiation as a consequence thereof, the bias voltage for the drift fields varies significantly to counteract the effect of radiation damage and increased leakage current. At the end of Run 2, this resulted in a bias voltage close to the design maximum operating voltage of 600 V!

In total, the Pixel Detector consists of 1744 sensors each one of which holds 47232 individual pixels (46080 independent read-out channels due to combination of pixels required by the read-out system) with a thickness of 250 μm. The segmentation of the Pixel Detector results

in a excellent position resolution of 10 μm in the transverse (r, ϕ)-plane as well as 115 μm in the longitudinal z-direction.

The system outlined in the previous paragraphs describes the Pixel Detector as it was during the first Run of data taking from 2011 to the end of 2012. Throughout the long shutdown (lasting from the end Run 1 to 2015), the Pixel Detector got extended by an additional pixel layer that is referred to as Insertable B-Layer (IBL). Contrary to the other sub-systems of the ID, the IBL does not have end-cap regions. With this "new" layer at hand, the Inner Detector allows measurements very close (33 mm) to the interaction point and hence significantly improves interaction vertex reconstruction for the benefit of reducing pile-up and improved b-hadron identification.

The Semiconductor Tracker (SCT)

The subsequent sub-component of the Inner Detector following the Pixel Detector system is the so-called SemiConductor Tracker (SCT). The SCT is a semiconductor microstrip detector (contrary to the innermost component that uses pixel instead of strips) that extends over a radial distance from 299 mm to 560 mm with respect to the beam line and covers a pseudorapisity range of $|\eta| < 2.5$, whereby the barrel module only extends to $|\eta| < 1.5$.

The operating principle is the same as for the Pixel Detector and works by measuring the voltage drop due to the collection of ionized charges in an external electromagnetic field when exciting electrons to the conduction band of the semiconductor.

Because of the strip design of the SCT and due to the larger detection area that needs to be covered, the sensor size of the SCT is larger compared to the ones used for the Pixel Detector. The entire SCT consists of 4088 sensors with a size of 126 mm $\times$ 80 μm that distribute over 4 layers in the barrel region of the detector and over 9 end-cap disks perpendicular to the beam line. In addition, each sensor is double-layered, with an angle of 207/162 μrad to get rid of ambiguities in hit reconstruction and to provide 2-dimensional positional information. This results in a position resolution of 17 μm in the (r, ϕ)-plane and 580 μm in the z- and r-direction respectively.

The Transition Radiation Tracker (TRT)

The Transition Radiation Tracker is the third and the outermost sub-component of the Inner Detector system that combines two technologies; namely, measurement of particle-dependent transition radiation, i.e., radiation emitted if charges passes through inhomogeneous media, as well as tracking detectors for the reconstruction of positional information. Unlike the Pixel Detector and the SCT – which are based on semiconductor material –, the TRT consists of polyimide, gas-filled drift tubes (also referred to as "straws") with a diameter of 4 mm that cover a pseudorapidity range of $|\eta| < 2.0$ rather than $|\eta| < 2.5$ as it was the case for the aforementioned sub-detectors.

As it was the case for the other components (with the exception of the IBL), the TRT comprises a barrel region that encompasses a radial distance of $r \in [563\,\text{mm}, 1066\,\text{mm}]$ with a longitudinal range of $|z| < 712$ mm as well as an end-cap region with $r \in [644\,\text{mm}, 1004\,\text{mm}]$ and $|z| \in [848\,\text{mm}, 2710\,\text{mm}]$. The barrel part of the TRT consists of a total of 52,544 straws each one being oriented alongside to the beam line and 1440 mm in length. The end-cap on the other side is made up of 122,880 radially aligned straws with a length of 370 mm. Each individual straw is made from a 35 μm thick mantle of Kapton that is coated by an aluminum and a graphite-polyimide film, with the latter one being responsible for good electrical conductivity as well as protection of the gas content of the tubes. Inside each straw tube, there is gold-tungsten wire with a diameter of 30 μm that acts as an anode with an electric potential of roughly 15, 000 V with respect to the outer sheath.

The gas that fills the tubes consists of 70 % of the noble gas argon[4] (Ar), 27 % carbon dioxide (CO_2) and 3 % oxygen (O_2). The composition of the gas is chosen to optimize the

[4]During Run 1, xenon was used instead, but it got replaced by argon before Run 2, which is why the TRT can not be used for electron identification anymore.

absorption of photons originating from the transition radiation when the particles traverse through inhomogeneous media.

As it was the case for the Pixel Detector and the SCT, if a charged particle passes through the straw tubes, the gas inside the recipient is ionized and the excited electrons drift towards the wire anode while the ions/holes drift to the straw wall. Due to the very large potential difference between anode and cathode, the signal gets amplified by a factor of 20,000 due to an avalanche of induced electrons.

The emission of transition radiation is triggered (with a certain probability) by a 3 mm polypropylene-polyethylene fiber "mattress" which is placed between the straws and through which charged particles passing through. The transition radiation is a consequence of the difference between the index of refraction between the two media which creates a correlation between different Lorenz factors γ and the energy of the emitted photon. This relation can be used for particle identification as γ will be different for different kinds of particles due to mass differences between those.

3.3.4 The Calorimeter Systems

After the particles traversed through the tracking detectors, they enter the calorimeter system of the ATLAS detector, which is responsible for precise energy reconstruction and measurement. Contrary to the Inner Detector, the calorimeter does measure the energy of *all* particles that undergo either electromagnetic and/or strong interactions and therefore charged and neutral particles alike – with the exception of the muon as it is a minimally ionizing particle at the energy scale of the LHC. Generally, since neutrinos solely interact via the weak interaction, they leave the detector without any trace, but causing a net momentum in the transverse direction, which will be discussed later in this thesis when talking about the reconstruction of physical objects. The particles originating from the hard interaction vertex are not directly

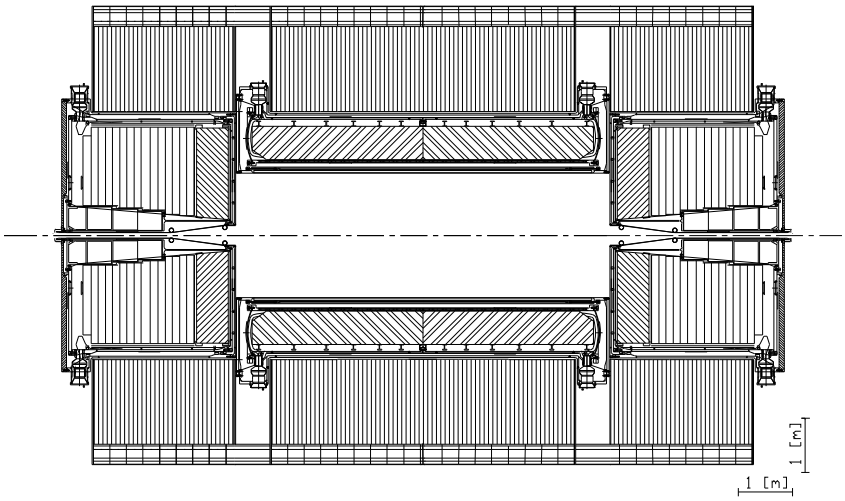


Fig. 3.6: View on a computer-generated schematic representation of the ATLAS calorimeters (taken from Ref. 13, Fig. 1.4, p. 9).

reconstructed by the calorimeters, but the cascade of secondary particles they initialize – generally referred to as *particle shower*. The characteristics of this shower strongly depend on the underlying interaction (electromagnetic or hadronic) that caused the cascade, which is why the ATLAS detector is equipped with two calorimeter systems each one dedicated to one specific type of shower.

Contrary to the inner detector, the energy resolution of the calorimeters improves with increasing energy, *inter alia*, due to the collimation of showers. Generally, the (relative) energy

resolution of the calorimeter can be expressed in terms of the following Equation

$$\frac{\sigma(E)}{E} = \frac{a}{\sqrt{E}} \oplus \frac{b}{E} \oplus c, \tag{3.4}$$

whereby $a \in \mathbb{R}$ denotes the stochastic term that depends on the detector material; $b \in \mathbb{R}$ is the noise term that corresponds to the noise level in the detector (which significantly depends on the pile-up condition of the respective Run) and the constant term $c \in \mathbb{R}$ that dominates at high energies and depends on detector layout like, for instance, the absorption/radiation length, cracks between transitions of different detector modules, dead material etc. The constant term can be reduced by means of calibration corrections as will be described later in Chapter 4.

The ATLAS calorimeter consists two *sampling* calorimeters, i.e., they are made of alternating layers of absorber material to induce particle showers followed by some active material to measure the energy of the initialized cascade. The sampling design is a specific choice that allows for very compact calorimeters with well controlled showers; however, only the active material components of the sampling calorimeters allow for energy reconstruction, which gives rise to energy loss in the absorbing material. This effect can be corrected based on simulations and test beam measurements, but it comes with an additional uncertainty. Figure 3.6 shows both calorimeters in the ATLAS detector with the Inner Detector (described in Section 3.3.4) being in its center.

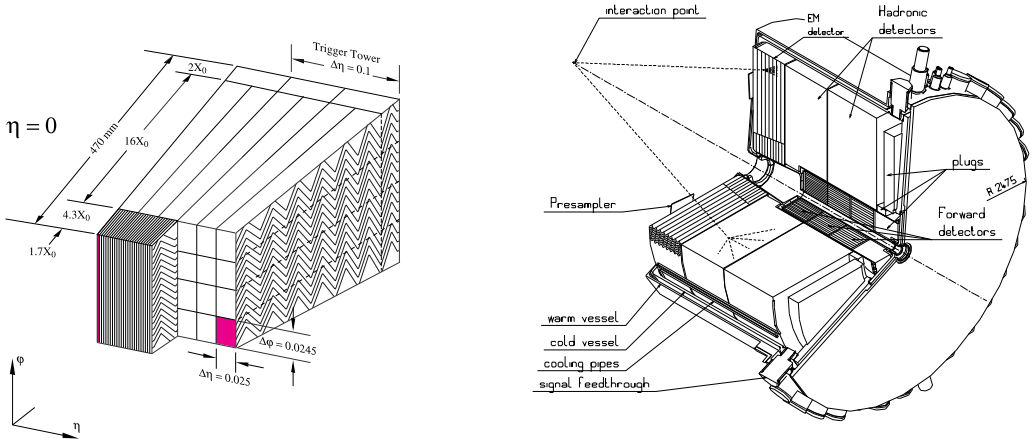
Consequently, the two calorimeters of the ATLAS detector are named based on the interaction they are primary sensitive to, i.e., the *electromagnetic calorimeter* for the energy reconstruction of showers originating from electrons or photons (muons are, as already mentioned, MIPs at the LHC and taus hadronize before they reach the calorimeter system) and the *hadronic calorimeter* that measures the energy of all kind of hadronic showers involving strong interactions. Together, the entire calorimeter system of the ATLAS detector, including the (extended) barrel as well as the end-cap and forward regions, cover a pseudorapidity range of $|\eta| < 4.9$, making the ATLAS detector almost hermetic (4π-detector).

The two following sections provide a quick overview of both calorimeters used at the ATLAS detector starting with the electromagnetic calorimeter and followed by the hadronic calorimeter. For a comprehensive discussion of the ATLAS calorimeter system, the reader is referred to the corresponding TDR [114] as before.

The Electromagnetic Calorimeter

The Electromagnetic calorimeter of the ATLAS detector consists of three sub-components: the presampler, the barrel and the end-cap calorimeter, that use Liquid Argon (LAr) as passive material to initialize the electromagnetic shower due to its radiation-hardness. The active material that measures the energy of the shower is lead because of its high density and short radiations length, which contains the shower dept of the electromagnetic calorimeter. Figure 3.7 shows a segment of the barrel module of the electromagnetic calorimeter with the granularity of the different layers in (ϕ, η)-space.

The electromagnetic calorimeter is optimized for showers originating from either electrons(positrons) or photons. Since electrons emit photons via Bremsstrahlung at high energies and photons convert to electron-positron pairs, both shower types are fundamentally connected. In the progress of the particle cascade, the energy of the initial particle that started the shower distributes over an increasing number of particles such that the energy scale of the individual particles within the shower reduces, which gives rise to other interactions with the detector material. For instance, at lower energies, the wavelength of the photons increases due to the interaction with the atomic kernels through Compton scattering. At even lower energies, photons are converted to low energy electrons by means of the photoelectric effect, which are captured and therefore no longer available to further develop of the shower. The electromagnetic shower finally terminates. The penetration dept of the described process strongly depends on the radiation length of the material. As already mentioned, this is one reason for using lead as an active material since it constraints the shower to the dept of the electromagnetic calorimeter. If a particle traverses through the absorbing material of the calorimeter, the LAr gets ionized and the excited electrons drift towards the copper electrodes



(a) *(b)*

Fig. 3.7: Sketch of a barrel module of the electromagnetic LAr calorimeter showing the detector granularity in (ϕ, η)-space for different layers including tower cells used by the L1 trigger of the ATLAS experiment (*left*, adapted and modified from Ref. 14, Fig. 9, p. 9) as well as a view on a computer-generated schematic representation of the end-cap (*right*, taken from Ref. 15, Fig. 1.1, p. 20).

in a strong electromagnetic field. This process takes about 450 ns, which is much longer (18×) than the average distance between two bunches of 25 ns; hence, not the entire signal is read out, but a signal shaper is applied to create a pulse with much shorter read-out time.

The design energy resolution of the electromagnetic calorimeter (barrel and forward component) is

$$\frac{\sigma^{\text{ECAL}}(E)}{E} = \frac{10\,\%}{\sqrt{E}} \oplus 0.7\,\%, \tag{3.5}$$

whereby the noise term has been neglected as it significantly depends on the pile-up conditions which changes from Run to Run [115].

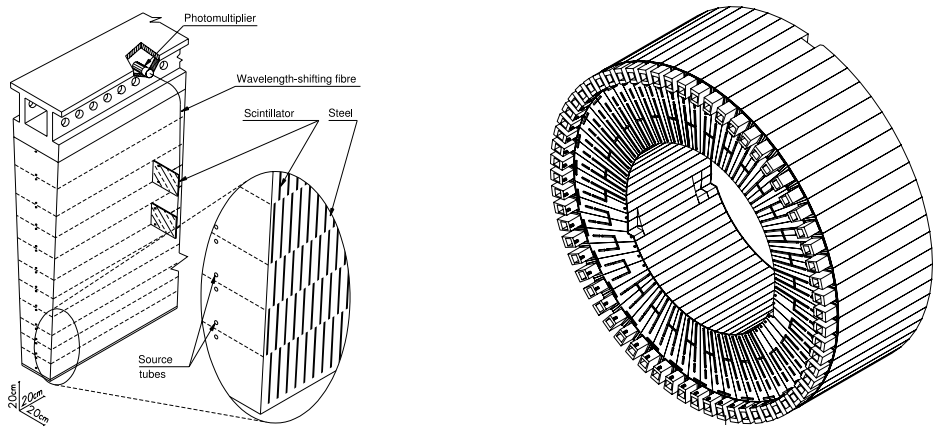
The Hadronic Calorimeter

ATLAS' hadronic calorimeter system is an essential component for the energy measurement of hadronic showers and thus crucial for the reconstruction of jets. It is placed outside of the electromagnetic calorimeter and uses different types of sampling calorimeters depending on the pseudorapidity region due to the accompanying varying demands regarding the radiation hardness of the sensors. Contrary to the electromagnetic calorimeter introduced in the previous section, the hadronic calorimeter also measures energy losses due to nuclear interactions and is therefore sensitive to strong interactions. This results in a *considerably* more complicated phenomenology of hadron-initialized showers compared to those caused by the interaction of, e.g., electrons or photons with matter. It is not just that hadronic showers always contain electromagnetic components to some extend, but the presence of strong interactions give rise to nuclear interactions like, for instance, nuclear spallation, recoil effects with the atomic nucleus, neutron capture, decay to stable hadrons etc. The strong interaction of a single hadron with matter usually results in the production of may secondary particles giving rise to hadronic cascades. Since the initial energy distributes over many secondary particles, the energy scale of hadronic showers is lower, which results in significant event-by-event fluctuation compared to electromagnetic showers. Those low energy hadrons may generate low energy photons and neutrons due to nuclear *de*excitation that are absorbed as nuclear binding energy or recoil events; energy losses due to the latter one will remain undetected. Generally, a significant fraction low-energy nuclear interactions are not reconstructed, with the consequence being a reduced energy resolution of hadronic calorimeters. The high-energy interactions of

hadronic showers, on the other hand, usually leads to the production of neutral pions which predominantly decay via $\pi^0 \to \gamma\gamma$, thus giving rise to electromagnetic showers. Similarly, the photons emitted through the process of deexcitation may initialize electromagnetic showers too.

Like the electromagnetic calorimeter, the hadronic calorimeter consists of a barrel and a forward detector. The barrel region is a cylindrical shaped detector that comprises the central barrel and two extended barrels in opposite ends of the central one. The hadronic calorimeter surrounds the electromagnetic calorimeter and is the outermost component of ATLAS' calorimetr system. This arrangement is necessary since, *firstly*, the hadronic showers contain, as mentioned above, electromagnetic components too, and, *secondly*, the absorption length, i.e., the characteristic energy loss distance for hadrons, is usually significantly larger compared to the radiation length of electromagnetic showers. The latter point is the reason for hadronic calorimeters being thicker compared to their electromagnetic counterpart.

The *barrel part* of the hadronic calorimeter along with one of the `HCAL` modules is shown



(a) *(b)*

Fig. 3.8: Sketch of a central barrel module (*left*, taken from Ref. 15, Fig. 1.1, p. 20) and an entire section (*right*, taken from Ref. 16, Fig. 1.3, p. 6) of the hadronic Tile calorimeter showing the alternating layers of steel absorber plates and plastic scintillating tiles connected to photomultipliers by means of wavelength shifting fibers.

in Figure 3.8. It is a *Tile sampling calorimeter* with the central barrel part covering a pseudorapidity region of $|\eta| < 1.0$ and the extended barrels increasing (with an overlap) this range up to $0.8 < |\eta| < 1.7$. The individual modules are made from alternating steel absorber plates followed by polystyrene scintillating tiles (the active material), with the latter one sending the scintillation light via wavelength shifting fibers (to convert from the ultraviolet to the visible range) to photomultipliers that convert the received light into electrical signals. A particle traversing through the calorimeter gives rise to a shower if it interacts with the steel absorbers, with the plastic scintillators producing photons that are converted in to electrical signals as mentioned above. The entire Tile calorimeter contains approximately $420,000$ plastic scintillators connected to 9500 photomultiplier tubes with a total weight of 2900 tonnes; making it the heaviest component of the entire ATLAS detector.

The *end-cap* part of the hadronic calorimeter (HEC) covers a pseudorapidity range of $1.5 < |\eta| < 3.2$ and is based on the LAr technology due to its intrinsically increased radiation-hardness to ensure smooth operation at larger pseudorapidity closer to the beam with significantly increased radiation dose. The absorber plates are orthogonal to the beam axis and made from Copper instead of steel/iron with a thickness of 25 mm and 50 mm, respectively. The entire hadronic end-cap is contained in the same cryostats as the electromagnetic calorimeter end-cap.

The last part of the hadronic calorimeter is the *Forward Calorimeter* (FCal) that consists of two end-caps and covers a pseudorapidity range of $3.1 < |\eta| < 4.9$. This module is exposed to

a *very* high degree of radiation dose, which is reflected in its composition and its structure. It uses two longitudinal sampling layers with Copper as an absorber to improve the measurement of electromagnetic interactions and Thungsten for hadronic showers. For all modules, LAr is used as an active material accounting for the challenging radiation environment.

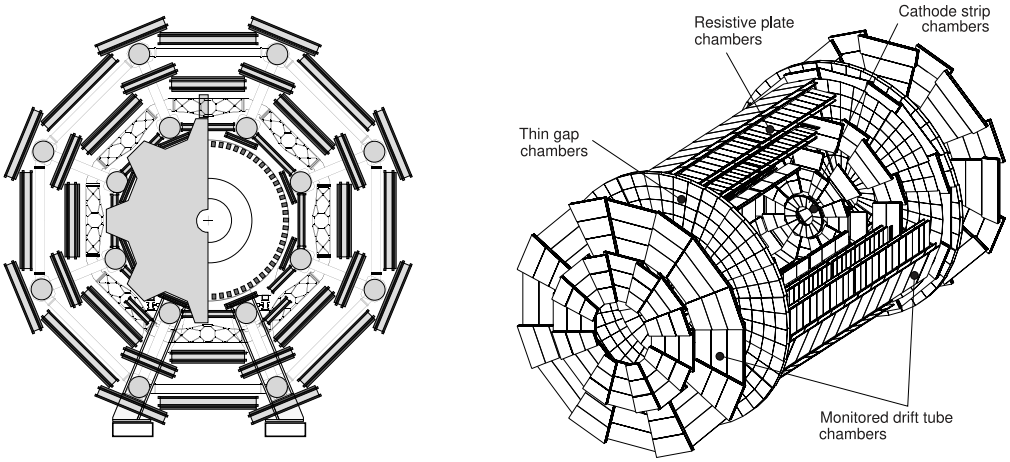
The design energy resolution of the hadronic calorimeter is different for the Tile, HEC and FCal

$$\frac{\sigma^{\texttt{Tile/HEC}}(E)}{E} = \frac{50\,\%}{\sqrt{E}} \oplus 3\,\%, \qquad (3.6) \qquad \frac{\sigma^{\texttt{FCal}}(E)}{E} = \frac{100\,\%}{\sqrt{E}} \oplus 10\,\%, \qquad (3.7)$$

with the noise being neglected again. As one can see, the resolution of the hadronic calorimeter considerably decreases in the forward region of the detector.

3.3.5 The Muon Spectrometer

The last component of the ATLAS detector is a spectrometer solely dedicated to the measurement and detection of muons. As already mentioned in Section 3.3.2 and Section 3.3.8, the muon played an important role in the discovery (or exclusion) of the Higgs boson as it provides one of the cleanest signatures with very high background suppression. Figure 3.4



(a) *(b)*

Fig. 3.9: Cross section view of the muon spectrometer of the ATLAS detector including its support structure (*left*, taken from Ref. 17, Fig. 1.2, p. 4) and a perspective view of the detector component, showing barrel region and end-cap modules of the spectrometer (*right*, Ref. 18, Fig. 1.1, p. 18).

shows the cross section of the spectrometer including its support structure it rests on, as well as an angled view of the detector that shows the barrel part and the end-caps. The barrel consists of three cylindrical tracking chambers around the beam axis; the end-caps (as well as the transition regions on both sides) are planes perpendicular to the beam line.

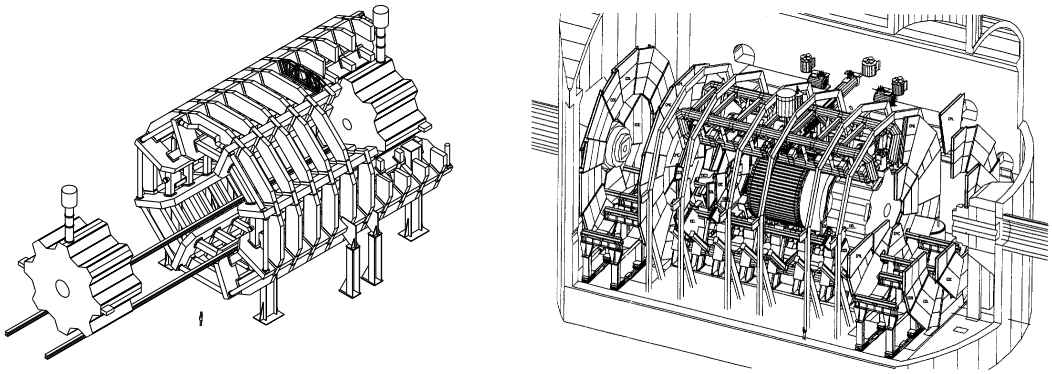
As mentioned above, the information of the muon spectrometer is used by the first-level trigger; consequently, this detector component is equipped with its own trigger system. The high-precision track and position reconstruction is done by using a range of different detection techniques such as the Monitored Drift Tubes (MDTs) that measure the bending of the tracks in the three air-core toroids (see Section 3.3.6). Furthermore, depending on the pseudorapidity, the muon spectrometer further uses Cathode Strip Chambers (CSCs) with fine granularity to account for high event rates and higher doses of radiation close to the beam line.

The muon spectrometer does not make use of any methods for particle identification as all particles produced in the interaction are assumed to be absorbed completely by preceding detector components.

3.3.6 ATLAS' Magnetic System

To reconstruct the momentum and charge of charged particles passing through the detector, the trajectories *must* be bend. For this purpose, the ATLAS detector is equipped with two types of superconducting magnet systems – the so-called *solenoidal* and *toroidal* magnets –, which needs to operate at a temperature close to absolute zero in an effort to achieve the field strength necessary.

Figure 3.10 shows an overview of the ATLAS magnetic system. Starting from the innermost component of the magnetic system, Figure 3.10b exposes the powerful *Central Solenoid magnet* (CS) that encloses in Inner Detector of the ATLAS experiment. It has a longitudinal extension along the beam line of around 5.6 m with a diameter of roughly 2.5 m (by the way, all numbers related to the magnetic system are taken from the official "ATLAS Fact Sheet" [116]) and a total weight of 5 tonnes. The magnet archives a field strength of up to 2 T with a casing thickness of just 4.5 cm to reduce the interaction with particles to a minimum. The magnet itself consists of over 9 km Niobium-titanium superconductor wires into fortified aluminum strips.



(a) *(b)*

Fig. 3.10: The ATLAS magnet system and its components. The superconducting air-core toroid magnet with the left-hand end-cap toroid being retracted from its operating position (*left*, taken from Ref. 13, Fig. 1.7, p. 12), as well as, a perspective view on ATLAS' magnet system including the innermost central solenoid magnet enveloping the ID (*right*, taken from Ref. 19, Fig. 2.2, p. 16 (TDR)).

The central solenoid is followed by the *toroid magnet* that is further subdivided into a *barrel* and *end-cap* region as shown in Figure 3.10a. This is an incredible system with flabbergasting dimensions.

The barrel toroid consists of a series of eight large coils with a length of approximately 20 m and an outer diameter of the assembled system of 20 m (which makes it the largest toroidal magnet constructed so far), providing a magnetic field up to 3.5 T. The eight coils are build from 56 km of superconducting wire made of (super)conducting Aluminium, Niobium-titanium and Copper. The entire structure is arranged radially and symmetrically around the beam axis.

The *End-Cap Toroid* (ECT) is shown in Figure 3.10b, being retracted from its operating position. The purpose of the two ECTs is to extend ATLAS' magnetic field to forward particles close to the beam line. Both end-caps have an axial length of roughly 5 m and an outer diameter of 10.7 m with a bore in the center for the beam line. The two components are made from the same material as the CBT but with a total weight of 240 tonnes and a magnetic field strength of 4 T.

The operation of the entire magnet system in terms of cooling and current supply ($8-21$ kA) is quite energy-intensive. The cooling is managed in a dedicated refrigeration plant located in a separate cavern close to the main one UX15 as can be seen in Figure 3.4a.

3.3.7 LUCID, ALFA and ZDC Experiments

The ATLAS experiment is equipped with a series of smaller experiments/detectors that monitor beam conditions, provide luminosity determination and measure spectator beam remnants (which is of particular importance for detection of diffractive processes).

The *LUminosity measurement using a Cherenkov Integrating Detector* (LUCID) is the dedicated luminosity monitoring detector of the ATLAS experiment [117]. To meet the doubled luminosity and the increase in the number of interactions per bunch-crossing for Run 2, LUCID and its entire electronics have been completely rebuilt; as a consequence, the detector was renamed to LUCID-2 to account for this fact. The detector itself consists of two modules that are placed roughly 17 m symmetrical with respect to the interaction point of the ATLAS experiment. A total number of 16 photomultipliers build the heart of the system that are used to detect Cherenkov radiation. The photomultipliers are calibrated using radiactive sources, and based on the integrated Cherenkov radiation, the luminosity can be determined. The LUCID group also provides systematic uncertainties that are dominated by calibration errors and Run-to-Run systematic errors; currently, the uncertainty on the luminosity is roughly 2.1 %, which is a significant improvement compared to Run 1. For the measurement described in this thesis, the uncertainty on the luminosity is irrelevant since relative rates are measured ,with the luminosity being fully correlated between the nominator and the denominator.

The *Absolute Luminosity For ATLAS* (ALFA) is another interesting sub-detector system of the ATLAS experiment that is located approximately 240 m from the interaction point. It uses Roman Pots for the measurement of the *total* proton-proton cross section by counting the number of forward protons. This measurement is based on the *optical theorem* $\sigma_{\text{tot}} = \frac{4\pi}{k}\text{Im}f(\theta = 0)$ in QFT that relates forward scattering amplitude $f(\theta = 0)$ to the total cross section σ_{tot} of a process.

Furthermore, the so-called *Zero-Degree Calorimeter* (ZDC) is installed very forward at around 140 m at both sides from ATLAS' interaction point. Its main task is to detect very forward neutral particles – primarily (spectator) neutrons and photons – in proton-proton as well as heavy-ion collisions. Recently, the detector has gained popularity in the search for light-by-light scattering predicted by quantum electrodynamics.

3.3.8 "Data Flow" – Trigger and Data Acquisition

The ATLAS Trigger and Data AcQuisition (TDAQ) system is responsible for all the *online* processing of the collision data such as the selection and storage of events containing signatures considered interesting for physics analyses. It is a highly complicated, multi-component system that has undergone considerable changes from Run 1 to Run 2 to meet the significant increase of instantaneous luminosities and bunch crossings in a given time frame.

With the design parameters of the LHC, roughly one billion events occur within the time frame of a second. Given the enormous amount of calorimeter cells, pixels etc., this corresponds to an immense amount of data that would need to get permanently stored on disk. Due to bandwidth limitations and capacity limits regarding the analysis as well as the storage of data, the total event rate must be significantly reduced from $\mathcal{O}(10^9\,\text{kHz})$ to something in the neighborhood of $\mathcal{O}(1\,\text{kHz})$. For this purpose, the ATLAS (Run 2) trigger is a two-stage system that utilizes a fast hardware-based component as the first level in the trigger chain followed by a software-based level for further rate reduction and event selection.

The Level-1 (L1) trigger is an entirely hardware-based system that is build from custom electronics optimized to achieve the best performance. As an input, the L1 trigger uses reduced information from the calorimeter (L1Calo) and the muon spectrometer (L1Muon); the Inner Detector is not taken into account in the first step of the trigger decision as tracking is not possible within the strict time budget of the L1 trigger. The calorimeter part of the L1 trigger reads signals from the calorimeter with a reduced granularity. The hardware of the L1Calo trigger is optimized to identify patterns in the calorimeter expected for electrons, photons, taus, jets as well as an estimate of the missing transverse energy in an event. Since muons are MIPs at the LHC, the dedicated L1Muon trigger searches for characteristic signals in the muon spectrometer of the ATLAS detector. This is necessary since muons provide one of the cleanest selection of interesting events like, e.g., $H \to ZZ^* \to \mu^+\mu^-\,\ell^+\ell^-$ with $\ell = e, \mu$. In

the meanwhile, during analysis of the reduced-granularity calorimeter information, the analog signals of all sub-components of the detector are processed and buffered. Depending on the L1's decision, the information in the event buffer is passed to the next trigger step or discarded. If the L1 trigger identifies one of the previously mentioned objects, it defines Regions-of-Interest (RoIs) in terms of η and ϕ and checks the results against entries in the predefined L1 trigger menu. If *at least* one of these criteria is matched, the event is accepted by the L1 trigger and the information temporarily stored in the buffer along with the RoIs is passed to the next step in the trigger chain.

If the event is accepted in the first level, the full detector information is forwarded to the second stage, i.e., the software-based High-Level-Trigger (HLT) using approximately 20 k commercial CPUs; this trigger aims to reduce the event rate to manageable $\mathcal{O}(1\,\mathrm{kHz})$. Before the event is finally analyzed by the CPU-intensive algorithms similar to the ones used in offline analyses, some devoted fast algorithms are used to estimate if this event is likely to pass the HLT criteria. The HLT checks against a list of customizable criteria, the so-called *trigger menu*, that reflect the physical goals of a given data-taking period. If the current event is accepted by the HLT, the data is transferred to the CERN computing center Tier-0 for permanent storage.

4. Object Reconstruction and Jet Physics

The previous chapter provided an introduction into the ATLAS detector, which is used to measure the experimental signature of particles traversing through the measuring apparatus originating from the interaction point. The information read out from the instrument are a variety of electronic signals like, for instance, energy depositions in the calorimeter system, hits in the Inner Detector, trigger information etc. Hence, particles are not measured directly but rather implicitly via their interaction(s) with dedicated sub-components of the detector. To draw conclusions about the particles involved in the event, the feedback provided by the detector must be interpreted and connected to the actual physical objects that gave rise to the measured signatures in the first place. This is a highly non-trivial task that is generally referred to as *(physical) object reconstruction*. The objective is to use the detector information to reconstruct the four-momentum of the object (which might be a fundamental lepton, the experimental signature of a parton in form of a particle jet or a composed quantity like the missing transverse momentum that accounts for momentum imbalance in the transverse plane) in the event from which higher-level observables are derived.

This chapter covers the reconstruction of different physics objects. The measurement outlined in this thesis mostly relies on jets that, consequently, will be the central theme of this chapter. However, for the sake of completeness, a quick overview of other important physical objects is provided as well.

The first section (4.1) provides a brief introduction into the reconstruction and identification of electrons and photons at ATLAS. Since the reconstruction of the muon differs significantly, it gets its own section (4.2) that quickly introduces muon reconstruction and identification as well as the different definitions of muons and working points at the ATLAS experiment. To leave no family member unmentioned, some very shallow comments regarding the reconstruction of taus are provided. With the leptons and the photon being discusses, the subsequent section (4.4) is devoted to jet physics at the ATLAS experiment. This is a very broad topic and quite relevant for the studies presented in this report; hence, the material is divided into several parts (whereby concepts introduced in Chapter 1.3 are used extensively) starting with the experimental signatures of jets in the detector as well as the objects they are constructed from (4.4.1), followed by common jet finding and reconstruction algorithms (4.4.2), and concluded by an introduction to jet calibration (4.4.3) and some comments to jets build from charged constituents (4.4.4).

4.1 Electrons and Photons

Electrons as well as photons are fundamental particles that are produced in may physics processes at the LHC. Often, they provide a very clean signature in the final state and hence allow for a pure selection of channels involving those particles. In case of this analysis – which is exclusively based on jets in the final state –, electrons and photons play a comparatively minor part (they only enter the measurement implicitly through overlap removal in the jets). However, they still contribute due to detector mismeasurement effects and, following, the possibility of misidentifying an electron or a photon with a jet. This effect, however, is vanishingly small compared to the overwhelmingly large cross section of multijet events. Of course, hadronically decaying taus may contribute as well. On the other hand, jets originating from QCD contribute significantly to falsely identified electrons and photons also known as *fakes*.

The identification and reconstruction of electrons and photons is based on measurements done in the Inner Detector (inner tracking system) and the electromagnetic LAr calorimeter of the ATLAS detector, which has already been described in Chapter 3.3.4.

The first reconstruction step is identical for electrons and photons alike; it starts in the electromagnetic calorimeter on cell level, i.e., considering individual cells of the calorimeter that have registered an energy deposition above a certain threshold. Those cells are the input to the so-called *sliding window algorithm* that uses a fixed-size collection of 5×5 calorimeter cells, with the objective to identify a region that maximizes the amount of energy within the sliding window. If a maximum has been found, the algorithm tries to identify a track (which – of course – only exists in case of an electron) from the ID with a momentum measurement compatible with the energy deposition in the localized region in the calorimeter. If this procedure turns out to be successful, the electron could still be caused by photon conversion due to the interaction of a photon with the detector material. This is checked by searching for the presence of an associated conversion vertex. If no conversion vertex is found, the reconstructed particle is labeled as an electron candidate; otherwise, if there is a conversion vertex, the particle is a photon candidate.

After electron and/or photon candidates have been identified, the localized clusters inside the calorimeter are rebuild using adjusted window sizes, whose extension in η and ϕ depends on whether the candidate have been reconstructed in the barrel region ($(3 \times 7)_e$ and $(3 \times 5)_\gamma$) or the end-cap(s) ($(5 \times 5)_{e,\gamma}$) of the calorimeter. The reason for using larger window sizes for electrons than for photons is a consequence of their increased interaction probability in the upstream material, as well as because of bending due to the magnetic field, which may result in broader showers in case of the electron. Based on the sliding window, the basic kinematic properties of the electron/photon candidate is calculated. The angular components η, ϕ are calculated as the energy-weighted barycenter of the individual cells in the window while the energy (the particles are assumed to be massless) are the sum of energy depositions inside the individual layers as well as presampler of the calorimeter to correct for upstream energy losses.

As already mentioned, as a result of the overwhelmingly large multijet cross section, the number of electron and photon candidates is completely dominated by fakes. This problem is accounted for by using a set of variables that capture differences between jet- and electron/photon-initialized showers like, to name but a few examples, the shower shape, the quality of the tracks, information provided by the TRT (see Chapter 3.3.3) that serve as input for cut-based as well as multivariate techniques. The classification score is optimized for different efficiency that give rise distinct working points, e.g., a *tight* electron with an electron identification efficiency of roughly 64 % and a jet rejection of approximately 10^5 [118].

4.2 Muons

Precise muon reconstruction and identification is of utter importance for many measurements and in searches for particles beyond the current Standard Model. For instance, many theories and extensions to the SM predict the existence of heavy resonances that predominantly couple to the top quark, whose decay gives rise to a hadronic or a leptonic final state. A heavy particle decaying into $t\bar{t}$ therefore may become noticeable in a fully hadronic, a semi-leptonic or a fully leptonic final state. The fully hadronic final state is completely dominated by QCD multijet events, and the fully leptonic one is kinematically under constraint[1] due to the presence of two neutrinos in the final state. Furthermore, the fully-leptonic channel is limited by statistics due to the small branching ratios of the decaying W bosons. The semi-leptonic decay channel (one W decaying hadronically the other one leptonically) has a very clean signature and is usually the one that is used in searches. In this channel, muons are often used to significantly reduce contributions background to account for the aforementioned problem of QCD jets faking electrons (see Section 4.1). The reliable identification of muons, therefore, directly results in

[1] Actually, the resulting system of kinematic equations in case of the fully-leptonic final state results in a quartic polynomial that can be solved analytically if all masses entering the equations are fixed [119]. However, the solution to this polynomial is not unique and the efficiency in selecting the correct solution is low which decreases mass resolution.

reduced background and improved mass resolution.

This is just one example; there are plenty of other cases where muons are crucial in physics analysis like, for instance, $Z \to \mu^+\mu^-$.

As it was already outlined in Chapter 3.3, the reconstruction of muons is very different from the one of electrons or photons described in the previous section. Like for the electron and the photon, the muon reconstruction and identification algorithms make use of tracks reconstructed in the Inner Detector; however, due to the minimal ionization nature of the muons at the energy scales of the LHC, muons deposit little to no energy in the calorimeter system of the ATLAS detector. For this reason, the ATLAS detector is equipped with a dedicated muon spectrometer (see Chapter 3.3.5) for the reconstruction and identification of muons originating from the central collision point as well as for the rejection of cosmic muons reaching the detector despite being 100 m below ground.

There are several definitions of muons at the ATLAS experiment depending on which information provided by the detector enters the reconstruction. In general, a muon candidate is given by a set of five parameters that correspond to the five independent parameters $(d_0, z_0, \phi, \theta, q/p)$ obtained in a track fit.

If a muon originates from the collision point, its track follows a helix trajectory due to the strong solenoid electromagnet field in the ID. With the innermost layers of the ID being very close to the beam line, the impact parameter d_0, z_0 as well as ϕ, θ can be precisely determined. After the calorimeter system, the muon candidate enters the muon spectrometer in the outer layer of the ATLAS detector, with its own toroidal magnetic field as described in Chapter 3.3.6. Since the muon spectrometer is the outermost layer of the ATLAS' detector sub-components, there is a long "lever arm" which results in an improved q/p resolution compared to the ID.

The actual muon four-momentum depends on the reconstruction algorithm that has been used to reconstruct the muon. At ATLAS, there are four of such distinct reconstruction algorithms that thus result in four definitions of muons: *Combined (CB) muons* from a global fit of tracks from the ID and the muon spectrometer; *Segment-Tagged (ST) muons* from a fitted track in a segment of the ID and the muon spectrometer; *Calorimeter-Tagged (CT) muons* obtained from a fitted ID track and energy deposition[2] in the electromagnetic calorimeter; and *Extrapolated (ME) muons* that solely use tracks from the muon spectrometer.

Which types of muons are used strongly depends on the respective analysis and the selection Working Point (WP) that has been chosen. At ATLAS, four of such working points for muons exist: the *loose* WP that uses *all* muons defined above and hence maximizes reconstruction efficiency; the *medium* WP (which is ATLAS' default WP for muons) only uses CB and ME muons; the *tight* WP only makes use of CB and ME while maximizing purity; and the *high-*p_{T} WP for, as the name suggests, high p_{T} muons with an optimized momentum resolution.

This basically finishes the muon part. There are some more steps necessary not described in this overview like the calibration of the muon's momentum scale for the measurements in the ID and the muon spectrometer to mach the measurements in data using the invariant mass spectrum of the $\mu^+\mu^-$ system originating from a Z or a J/ψ decay. However, compared to the calibration of jets, the procedure for muons is rather straightforward.

A nice introduction into muon reconstruction and identification at the ATLAS experiment can be found in the following note Ref. [120], which also formed the content-related basis of this section.

4.3 Tau lepton

Wait ... there are three families of leptons – where is the tau? No one is left behind. With a mass of roughly 1.78 GeV, the tau lepton is by far the heaviest lepton known, which opens the window for the decay into light hadrons (predominantly pions). This is an extraordinary property – absolutely unique in the Standard Model! Of course, the tau may also decay into

[2]Despite muons being MIPs at the LHC, there might be some trace of a passing muon in the electromagnetic calorimeter of the detector. This energy deposition is usually small compared to, e.g., an electron or a photon, but nevertheless may be used in muon reconstruction, provided that the energy deposition is isolated.

lighter particles in the lepton family (involving two neutrinos!), but the branching fraction $\Gamma(\tau \to \text{hadrons}) \approx 68\,\%$ for hadrons dominates. From an instrumentalist's point of view, however, this feature goes hand in hand with a lot of complications when it comes to reconstruct and identification of the tau lepton in the experiment, which arise due to the fact that taus give rise to hadronic showers, and because of its extremely short life time $\tau_\tau = (290.3 \pm 0.5)\,10^{-15}\,\text{s}$ [121]. Due to these distinct features of the tau lepton, the treatment is very different compared to electrons and muons.

It was mentioned already that the tau may also decay leptonically into lighter leptons; usually, it is not possible to distinguish electrons and muons originating from tau decay from the ones starting from the collision point (or only with very low efficiency).

For the hadronic decay channel, however, several techniques have been invented to discriminate hadronic showers from top decays and showers being initialized by final state partons in the underlying event. Essentially, hadronic tau decays can be categorized into *one-pronged* or *three-pronged* events (of course, higher-pronged decays exists, but their contribution is rather small such that they are irrelevant for almost all practically applications), depending on the number of (charged) pions produces in the decay. Those different decays result (on average) in different shower shapes with different substructure inside the produced jets.

Due to the large mass of the tau and the comparatively small mass of the final state particles being produced, the decay products are usually highly boosted. For this reason ATLAS uses small $R(= 0.4)$ anti-k_t jets build from *locally* calibrated topo-clusters for tau candidates. To get a tau enriched region of phase space, multivariate-based methods are used, whereby the efficiency of tau identification is somewhat low compared to the other members of the lepton family.

4.4 Jets

This section opens the wide range of topics covering jet-related physics, which is of great importance for the research presented in the thesis at hand. Since the analysis described in this document measures the relative rates of dijet events, jets will be covered in much more detail compared to the objects introduced in the previous sections, which play a subordinate role in this study.

The basic concept of a hadronic shower and particle jet as a theoretical and phenomenological, well-defined (see IRC-safety in Chapter 1.3.2) approach to QCD was already introduced in Chapter 1.3.1. This section therefore focuses on the experimental aspects of jets, i.e., the input used for the reconstruction that yields different jet definitions, the reconstruction of jets in an event for different configurations of commonly used finding algorithms, the calibration of the energy scale and energy resolution of jets to particle level as well as the ghost- or ΔR-association for track-to-jet matching and, finally, the measurement of jet substructure information, which is the subject of this analysis.

4.4.1 Inputs to Jet Reconstruction

Jets are usually associated with a collimated spray of particles of electromagnetic or hadronic origin; hence, the most important detector component for the reconstruction of jets is the calorimeter system of the detector (lengthily described in Chapter 3.3.4). The experimental signature of jets in the calorimeter is a spatially distributed energy deposition (in all three spatial dimensions) that typically exceeds the size of a single cell of the calorimeter except for very boosted jets. This energy deposition can be subdivided into the transverse profile, i.e., the arrangement of cells in the $\eta - \phi$-plane, and the longitudinal energy deposition, namely, the organization of cells along the radial direction that corresponds to the penetration dept of the jet. Both, the transverse and the longitudinal shower profile, are important quantities with regard to the discrimination of electromagnetic and hadronic jets.

Solely based on calorimeter information, two important objects can be defined that serve as an input for jet reconstruction, i.e., calorimeter *towers* and calorimeter *cluster*

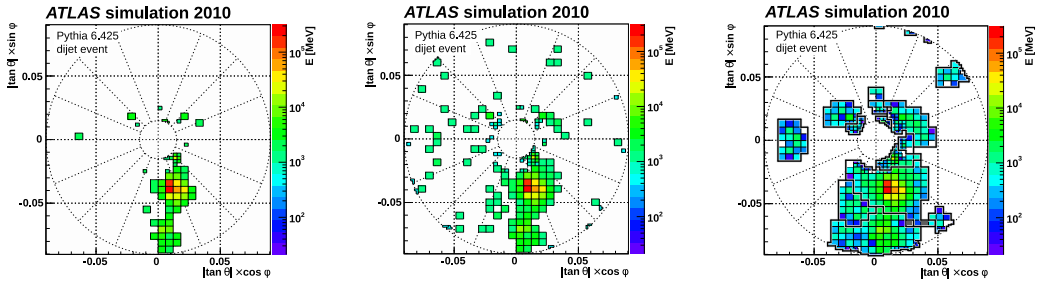
Calorimeter Towers

Calorimeter towers are the most simplest jet object available at ATLAS. They play a fundamental for triggering events containing jets as well as missing transverse energy. For this object, signal cells with sufficient activation are collected and the energy flow along the longitudinal direction is project onto a two dimensional $(\eta, \phi) \in (-\pi, \pi] \times (-5, 5]$ grid with a granularity of $\eta \times \phi = 0.1 \times 0.1$[3]. The tower algorithm simply collects signals cells and adds the energy, based on the fraction the energy deposition of overlapping cells as the granularity of the calorimeter changes. If the projected cell is smaller (or equal) to the bin size of the tower, its energy is simply added to the total tower energy. However, the cells may overlap such that the energy must be shared between towers. This is done by a simple geometrical reweighting of the energy based on the overlap [123].

The calorimeter towers can be subdivided into electromagnetic and hadronic towers depending on whether cells from the electromagnetic calorimeters or the entire ATLAS calorimeter are used. The former ones are primarily used as an input to the sliding window cluster finder to define Regions of Interest (RoI) for the higher-level trigger or the topological clustering algorithm described in the following section; the latter one is usually used for jet reconstruction. Both objects (and hence the jets build from it) are on the electromagnetic energy scale, which is one motivation for jet calibration introduced in one of the following section.

Calorimeter Clusters

Contrary to the previously introduced calorimeter towers, *topological cell clusters* are three dimensional objects that are reconstructed following a seed-based clustering algorithm. The resulting objects are referred to as *Topological Clusters* (TCs) and are a basic ingredient to many jet objects – including this very analysis that uses jets build from (locally) calibrated TCs as a basis to derive "track-based" jets! The underlying motivation behind the topological



(a) Seeding step *(b)* Growing step *(c)* Merge/split step

Fig. 4.1: The different steps of ATLAS' topo-clustering algorithm for a simulated dijet event in the `FCAL0` module of the forward calorimeter. The first Figure 4.1a highlights all cells with sufficient energy-to-noise ratio to seed a topo-cluster object. In the next step 4.1b, neighboring cells are collected if their energy deposition is significant. The last Figure 4.1c shows the final topo-clusters black outlined after the last merging step and splitting of local energy maxima (the figure has been adopted from Ref. 20, Fig. 6, p. 19).

cell clustering algorithm is the identification and reconstruction of characteristic showers in the calorimeters. Even if the initial particle is highly boosted (such that the collation angle of the decaying particles is in the order of the detector resolution), TCs are still strongly correlated with the momentum flow of the initial particle that gave rise to the shower cascade. The cluster algorithm itself is straightforward and will be covered in the following paragraph.

[3]It should be mentioned that the aforementioned scheme has been updated for the Phase-1 (Run 3) upgrade of the ATLAS detector. The readout electronics as well as the granularity of the LAr calorimeter have been updated which lead to the introduction of so-called "super cells" allowing to use shower shape information already at the very first trigger level [122].

The three main steps of the clustering algorithm are illustrated in Figure 4.1. As already mentioned, the topo-clustering procedure is a seed-based clustering algorithm that starts with a so-called *seed cell*, which is characterized by an energy deposition significantly above the noise level of the detector (see Figure 4.1a). As such, it is naturally noise and pile-up suppressed, which is of particular importance in an environment with significantly underlying activity like the LHC.

Each cell in the calorimeter with an energy significance of $E^{\mathrm{EM}}_{\mathrm{cell}}/\sigma^{\mathrm{EM}}_{\mathrm{noise,cell}} > S(=4)$, whereby the noise term is a squared sum of electronic noise and contributions from pile-up $\sigma^{\mathrm{EM}}_{\mathrm{noise,cell}} = \sigma^{\mathrm{electronics}}_{\mathrm{noise}} \oplus \sigma^{\mathrm{pile-up}}_{\mathrm{noise}}$, is able to seed a topo-cluster. In the following step, all cells with $E^{\mathrm{EM}}_{\mathrm{cell}}/\sigma^{\mathrm{EM}}_{\mathrm{noise,cell}} > N(=2)$ being directly adjacent to the seed cell are collected into the so-called proto-cluster (including the seed cell itself). It is important to keep in mind that the clustering of seed cells is done in *all* directions, i.e., the significance criterion is checked for cells even in different detector modules. As such, a resulting topo-cluster may span modules of the same calorimeter. This effect can nicely be observed in Figure 4.1b where all of a sudden growing cells appear without a seed cell being present in Figure 4.1a. In this case, the seed cell is located in a different module (e.g., `FCAC1`). This procedure is iteratively repeated; in case of two proto-clusters sharing the same neighboring cell, the two are merged into one cluster. The collection of cells stops until the last set of neighboring cells with significance above $E^{\mathrm{EM}}_{\mathrm{cell}}/\sigma^{\mathrm{EM}}_{\mathrm{noise,cell}} > P(=0)$. The significance levels S, N and P have been optimized for the ATLAS experiment and are also known as the so-called "420 scheme". One additional step is applied that is of particular importance for jet substructure and to avoid biases in the jet finding algorithm – this step is referred to as *cluster splitting*. Generally, the aforementioned clustering algorithm results in large topo-clusters with more that one local energy maxima (seeds). The topo-clusters are split and the energy is shared between the resulting new clusters based on a simple geometrical *re*weighting.

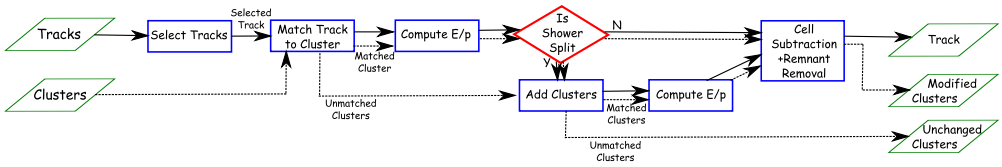
The output of the clustering algorithm is a three-dimensional collection of cells that can be used as an input to any jet finding algorithm (see Chapter 4.4.2) for jet reconstruction. However, before using those objects for the reconstruction of jets, they must be corrected to account for the non-compensating nature of the ATLAS calorimeter. Non-compensating means that the reconstructed energy of two different types of particles with identical initial energy may differ, resulting in different calorimeter responses. This problem is addressed by dedicated calibration schemes of which there are two currently available at ATLAS: the Local Cluster Weighting (LCW) and the EM scheme – with the latter one being more common and thus used in this analysis. Both analysis schemes are based on a classification of topo clusters as being of electromagnetic or hadronic origin. This can be done at high efficiency by means of shower shape variables [124].

ATLAS' Particle-Flow (PF) Jets

During the first Run of data-taking, all hadronic jets at ATLAS have been reconstructed from either calorimeter cluster or inner detector tracks solely. Track-based jets, to name an application, have been used in the calculation of the soft term in the missing transverse momentum while calorimeter-based clusters are jet object of choice for most analyses due to ATLAS excellent calorimeter system. In the final calibration of the jet energy scale that included the full Run 1 data set, it was found that the inclusion of track matched to the TC-based jets considerably improved the energy resolution of the jet (see Ref. [40]). Generally, the advantage of combining information from both detector components (inner detector and calorimeter) can be understood by the complementary behavior of the calorimeter and the inner tracker in terms of the energy resolution of particles: in case of low-energetic charged particles, the large curvature of the reconstructed tracks results in a very high energy resolution compared to the one obtained from the calorimeter; on the other hand, as the curvature (and hence the energy resolution) of tracks reduces inversely with the momentum of the particles, the energy resolution of the calorimeter is usually better in case of high-energetic particles. Hence, combining information from the tracker and the calorimeter of the detector results in an optimal resolution over a large range of energies. Furthermore, the utilization of tracks allows for a better identification of particles that are likely to be due to underlying activity in

the event, which is particular important for experimental setups with high luminosities.

Combining measurement information from different subsystems of the detector results in the risk of double counting energies; there are different approaches to solve this problem such as weighting the inputs to the algorithm accordingly. At ATLAS, *[a] cell-based energy subtraction algorithm is employed to remove overlaps between the momentum and energy measurements made in the inner detector and calorimeters, respectively.* ([40]). The individual steps of ATLAS' particle flow algorithm are shown in Figure 4.1. The algorithm starts from a collection



Plot 4.1: 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[.] (the figure as well as its caption have been adopted and modified from Ref. 40, Fig. 2, p. 10).

of tracks and calorimeter cluster cells. In order for the tracks being possible candidates, they must fulfill certain quality criteria (see Section 6.6 in Ref. [40] and/or Chapter 9.2.1 that introduces the track selection for this analysis). Once a suitable, well-measured collection of jets has been selected, jets are matched to the selected calorimeter clusters by means of the ΔR metric. The next step is to estimate the expected energy deposition of the charged particle (given by the reconstructed track) in the calorimeter of the detector, which depends on the track's momentum as well as on the position of the calorimeter clusters. The expected energy deposition is mostly derived from the simulation of $\pi^{\pm}$ events in Monte Carlo. Internally, the algorithm performs an hypothesis test by evaluating the probability that the energy of the charged particle has been deposited in more than just one single topo-cluster. Depending on the outcome of this step, more tracks and topo clusters are added until a consistent configuration gas been found. In each step, the expected energy deposition of the charged particle is *subtracted* (hence the naming convention) from the energy that was actually measured in the calorimeter. For a more detailed explanation of the individual steps, the interested reader is referred to the following publication in Ref. [40].

4.4.2 Jet Finding and Jet Reconstruction

For a long time, the standard jet reconstruction algorithms have been so-called cone algorithms, based on the geometric definition of, well ..., a cone (the first jet algorithm has been developed by Sterman and Weinberg in 1977 [125]). Even though this definition might appear straightforward, it still allows for numerous different implementations of reconstruction algorithms. However, with the Snowmass Accord (see Chapter 1.3.2) as guidance regarding meaningful jet definitions, most cone algorithms seek to extremize the hadronic energy flow through a cone with a fixed radius R, with $R^2 = (\Delta\eta)^2 + (\Delta\phi)^2$, in $\eta - \phi$ space. According to this definition, the transverse energy of the jet is simply given by the scalar sum of the constituent's transverse energy within the cone

$$E_{\mathrm{T}} = \sum_{i \in \{R_i < R\}} E_{\mathrm{T},i}, \tag{4.1}$$

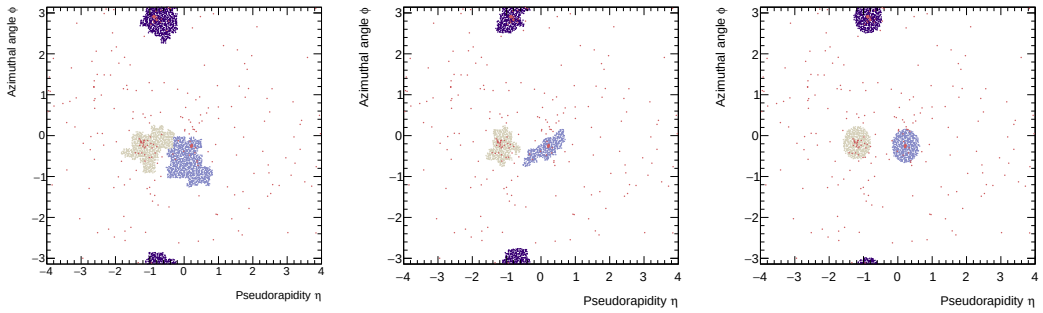
while the pseudorapidity and azimuthal angle of the jet is the weighted sum of the constituent's position in η–ϕ-space

$$\eta = \frac{1}{E_{\mathrm{T}}} \sum_{i \in \{R_i < R\}} E_{\mathrm{T},i}\eta_i, \qquad \phi = \frac{1}{E_{\mathrm{T}}} \sum_{i \in \{R_i < R\}} E_{\mathrm{T},i}\phi_i. \tag{4.2}$$

Those very simple definitions [126] already allow to reconstruct jets.

Most cone algorithms are seed-based, e.g., the particle with the largest momentum, and make use of an iterative procedure to successively combine particles until stable configurations are found. Within the group of iterative cone algorithms, the most prominent methods are the iterative cone algorithm with progressive removal (IC-PR) – which uses the hardest cell as a seed – and the iterative cone algorithm with the split-merge procedure – where all cells above a certain energy threshold are seeds. In contrast to iterative cone algorithms, fixed cone algorithms use a fixed geometry around the seed direction. Although cone algorithms are comparatively simple – or perhaps exactly for this reason –, most of them suffer a serious problem: they are soft (IC-SM) and collinear *un*safe (IC-PR) [127]. The aforementioned problem has their origin in the seed-based approach that is used by the algorithms. Therefore, the explicitly seedless cone algorithm, SISCone (Seedless Infrared Safe Cone [128]) was developed, which is the only representative of its family that is infrared as well as collinear safe (see Chapter 1.3.2). The algorithm used to identify stable pseudo-jets by SIScone is rather complicated since it involves several nested iterations and depends on several parameters. This circumstance also makes it complex regarding the implementation of the algorithm in a computer program. Hence, it would not be very conducive to discuss its details at this point (a description of the complete algorithm can be found in Ref. 128, Algorithm 2, p. 12). The problems regarding cone based algorithms were one of the main motivations for the development of so-called sequential recombination algorithms for jet clustering, which are the subject of the next section.

The second family of jet reconstruction algorithms are the so-called *sequential recombination algorithms* that have been introduced in the '80s by the JADE collaboration [129]; hence, have their origin in studies of e^+e^--collisions at PETRA at DESY, Hamburg, Germany. Within this class, there exists a large variety of different algorithms that are based on different assumptions such as the splitting functions in QCD for the k_t algorithm [130, 131], angular ordering for Cambridge-Aachen [132] and collimated jet cores as used by anti-k_t [133] algorithm. The first two are especially suited for studies regarding the substructure of jets, while the latter one is often used to study single-parton jets. Sequential recombination algorithms are inherently



(a) k_t algorithm *(b)* C/A algorithm *(c)* Anti-k_t algorithm

Plot 4.2: Jets obtained with three different sequential recombination algorithms: the k_t (*left*), the Cambridge-Aachen (*middle*) and the anti-k_t algorithm (*right*) for small-R(= 0.4) jets. The colored regions correspond to the reconstructed detector region of each jet, whereby the jet area has consistently been defined using "ghosts". The red points scattered over the detector area correspond to "hits". The jets obtained with the Cambridge/Aachen and the k_t algorithm show quite complex, non-circular boundaries. The anti-k_t algorithm, on the other hand, gives perfectly circular jet shapes since the hardest constituents of the jets are clustered first.

(by construction) infrared and collinear safe. This property makes them theoretical superior to cone based reconstruction of jets, which is why those are generally favored by theorists and experimentalist's alike. The problem was, however, that this class of algorithms suffered from a poor computational performance. This issue has mostly been resolved by the efficient

implementation of the algorithms in the widely used FastJet package [134] in C++ that is also used by the experiments at the LHC.

Generally, the three most commonly used sequential recombination algorithms can be summarized by the following distance metric

$$d_{ij} = \min\left(p_{\mathrm{T},i}^{2p}, p_{\mathrm{T},j}^{2p}\right)\frac{\Delta R_{ij}^2}{R^2}, \quad d_{iB} = p_{\mathrm{T},i}^{2p}, \tag{4.3}$$

whereby $p = 1$ correspond to the k_t algorithm, $p = 0$ the C/A algorithm and $p = -1$ correspond to the anti-k_t algorithm.

The following short subsections provide a quick overview of the three algorithms.

The k_t Algorithm

When the k_t algorithm was introduced by the JADE experiment at DESY in 1993, it was designed in accordance with studies at lepton colliders. The (modified) distance measure of the algorithm is given by

$$y_{ij} = \frac{2\min(E_i^2, E_j^2)}{Q^2}(1 - \cos\theta_{ij}), \tag{4.4}$$

whereby Q is the total energy of the event. Equation 4.4 is evaluated for each pair of particles from which the minimum $y_{\min}$ is determined. If $y_{\min}$ below some threshold value y_{cut}, then i and j are recombined into a new particle (referred to as a "pseudo-jet"); afterwards the procedure is repeated until the iteration terminates. Due to the definition according to Equation 4.4, this algorithms favors soft particles, which results in somewhat "diffuse" shapes as can be seen in Figure 4.2a.

In the limit $\theta \ll 1$, y_{ij} is reduced to $y_{ij} \overset{\theta\ll 1}{=} (\min(E_i, E_j)\theta_{ij}/Q)^2$ which is the squared normalized transverse momentum – hence the name k_t algorithm. As mentioned in Section 4.4.2, the k_t algorithm can be related to the splitting functions in QCD by considering a soft and collinear branching $k \to ij$. In this case, the differential splitting is given by $\partial_{E_i,\theta_{ij}} P_{k\to ij} \sim 1/\min(E_i, E_j)\theta_{ij}$.

Equation 4.4 has been constructed with regard to the application in lepton colliders. In hadron colliders, however, the total energy of an event is not well defined along with other complications. An adapted version of the k_t algorithm for hadron collisions is given by

$$d_{ij} = \min\left(p_{\mathrm{T},i}^{2}, p_{\mathrm{T},j}^{2}\right)\frac{\Delta R_{ij}^2}{R^2}, \quad d_{iB} = p_{\mathrm{T},i}^{2}, \tag{4.5}$$

with each particle being assigned to either a beam-jet (d_{iB}) or a final state-jet [135].

The Cambridge-Aachen (C/A) Algorithm

The Cambridge-Aachen algorithm uses a distant measure similar to Equation 4.4 and 4.5 along with a second distant measure defined as $v_{ij} = 2(1 - \cos\theta_{ij})$. The rationale behind this method was to combine the k_t algorithm with angular ordering that is related to the ordering in multiple gluon emissions. The actual algorithm is simple: find the pair of particles with minimum v_{ij} and recombine those particles to a new pseudo-jet if the corresponding condition $y_{ij} < y_{\mathrm{cut}}$ is satisfied and repeat the process accordingly. As it was the case for the k_t algorithm, the Cambridge-Aachen algorithm must be modified to meet the requirements of hadron colliders; the updated metric is given by

$$d_{ij} = \Delta R_{ij}/R. \tag{4.6}$$

Like the k_t algorithm, the Cambridge-Aachen's approach tends to result in irregular jet shapes as seen in Figure 11.2a. This problem is, *inter alia*, addressed by the next sequential recombination algorithms – the famous and widespread *anti-k_t* algorithm.

The Anti-k_t Algorithm

The so-called anti-k_t algorithm corresponds to a parameter configuration of $p = -1$ in the distance measure as stated in Equation 4.3, which results in the following expression

$$d_{ij} = \min\left(p_{\mathrm{T},i}^{-2}, p_{\mathrm{T},j}^{-2}\right) \frac{R^2}{\Delta R_{ij}^2}, \quad d_{iB} = p_{\mathrm{T},i}^{-2}. \tag{4.7}$$

According to Equation 4.7, the anti-k_t algorithm assigns a small distance to pairs involving a hard particle. The metric is therefore mostly determined by the transverse momentum of the hard particles in the jet as well as the angular separation. This property implies that soft particles inside the jet get clustered with hard ones long before they have the chance to combine among themselves. For instance, given a jet with only one high-k_t constituent and plenty of soft particles (like in case of jets initialized by gluons or light quarks), since there is no hard neighbor present, the anti-k_t algorithm just collects all activity within the jet radius parameter R. As a consequence, the anti-k_t algorithm usually results in jets with a very circular boundary as can be seen in Figure 4.2c. Generally, the key feature of this algorithm is the resilience of the resulting shape of the jet with respect to soft particles [136], a property shared with the SISCone algorithm [137].

4.4.3 Jet Energy Scale Calibration

As explained in Section 4.4.1, topo-clusters are usually calibrated with the EM (EMTC) scheme and are therefore at the electromagnetic scale, i.e., the energy of electrons and photons is correctly reconstructed, but the energy of hadrons is not. Accordingly, jets build from calibrated topo-clusters are at the electromagnetic energy scale themselves, which does *not* correspond to the actual energy scale of the jet due to the significant fraction of hadronic showers, which exhibit a different detector response. The purpose of the calibration of the *Jet Energy Scale* (JES) is to correct for differences between the truth jet energy and the reconstructed energy as a function of the jet energy E_{jet} and pseudorapidity η_{jet}. The entire JES calibration scheme follows a sequence of consecutive steps of corrections that are derived from Monte Carlo studies for the most part, but also contain adjustments from in-situ measurements where the transverse momentum of the jet is compared with the p_{T} of a well-defined reference object. Each step in the calibration chain – illustrated in Figure 4.2 – yields a multiplicative (for scalar

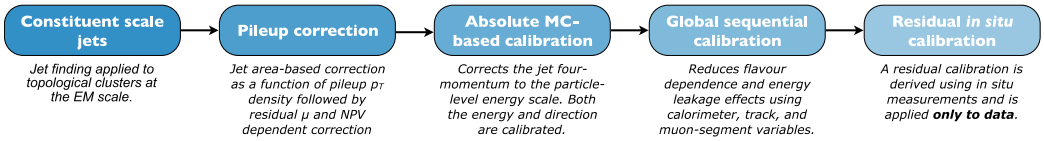


Fig. 4.2: Correction steps of the complete JES calibration sequence for jets at the EM energy scale. All corrections shown in the flowchart above are applied to the scalar component of the jet four-momentum; the origin correction is (usually) not part of the calibration chain for small-R jets [21].

four-vector components) or an additive (in case of angular components) correction $c_i \in \mathbb{R}$ and $i \in \mathbb{I} = \{\mathrm{PU}, \mathrm{JES}, \mathrm{GSC}, \mathrm{in-situ}\}$ to the four-vector of the jet according to

$$p_{\mathrm{jet}} = (c^{\mathrm{calib}} E_{\mathrm{jet}}, \eta_{\mathrm{jet}} + \Delta\eta^{\mathrm{calib}}, \phi_{\mathrm{jet}}, c^{\mathrm{calib}} m_{\mathrm{jet}}), \tag{4.8}$$

with c^{calib} being the final calibration factor that is a product of the individual corrections

$$c_{R=0.4}^{\mathrm{calib}} = \prod_{i \in \mathbb{I}} c_i = c_{\mathrm{PU}} \cdot c_{\mathrm{JES}} \cdot c_{\mathrm{GSC}} \cdot c_{\text{in-situ}}. \tag{4.9}$$

It should not go unmentioned that the above-mentioned calibration scheme is specific to so-called small-R jets with radius parameter $R = 0.4$ (as used in this measurement); their exists

an expanded scheme for large-$R(= 1.0)$ jets that accounts for the increased pile-up sensitivity of larger jet radii.

The following paragraphs provide a quick introduction in a nutshell into the individual steps of the JES calibration chain mentioned in Equation 4.9.

Origin correction: The first step is an origin correction of the reconstructed jet that recalculates the four-momentum of all reconstructed jets in the event to point to the hard-scatter primary vertex instead of the geometrical center of the detector. This operation does not change the energy of the jet but its position. This correction is done in order to improve the resolution of the pseudorapidity.

Pile-up correction: The subsequent step of the calibration sequence is a correction of the transverse momentum of the reconstructed jet to mitigate the effect of energy deposits due to particles originating from *pile-up* events.

This correction consists of two parts: *firstly*, the expected/mean contribution due to pile-up, ρ, is estimated and subtracted from the p_{T} of the reconstructed jet based on the jet area A calculated using ghost particles; *secondly*, to remove the remaining residual dependence of the jet's p_{T} on the number of primary interaction vertices N_{PV} (a proxy for the number of pile-up interactions in the event), a simple affine transformation $L_{\alpha,\beta}$ between p_{T} and N_{PV} is derived and subtracted to get a flat transverse momentum as a function of pile-up interactions.

The two corrections are subtracted from the reconstructed transverse momentum of the jet according to

$$p_{\mathrm{T}}^{\mathrm{corr}} = p_{\mathrm{T}}^{\mathrm{reco}} - \rho \times A - [\alpha \times (N_{\mathrm{PV}} - 1) + \beta \times \mu], \tag{4.10}$$

with the first term being the area-based correction and the latter two belonging to the residual correction.

Absolute MC-based calibration: This step corrects for a variety of effects that result in an mismeasurement of the reconstructed jet energy compared to the true energy of the initial hadron like, for instance, dead material, leakage, out-of-calorimeter jets and energy depositions below noise thresholds of the calorimeter. One additional obvious contribution is the deviation from the true energy scale of the jet due to the already mentioned non-compensating nature of the ATLAS calorimeter.

The energy of the jet is corrected based on the inverse average jet response function $\langle 1/\mathcal{R}^{\mathrm{EM}} \rangle$ defined via

$$c_{\mathrm{JES}} := \left\langle 1/\mathcal{R}^{\mathrm{EM}} \right\rangle_{(\eta,p_{\mathrm{T}})} = \left\langle E_{\mathrm{jet}}^{\mathrm{truth}}/E_{\mathrm{jet}}^{\mathrm{EM}} \right\rangle_{(\eta,p_{\mathrm{T}})}, \tag{4.11}$$

whereby $E_{\mathrm{jet}}^{\mathrm{truth}}$ denotes the energy of the *particle-level* jet that was matched (based on an ΔR matching criterion) to the *detector-level* jet $E_{\mathrm{jet}}^{\mathrm{EM}}$, i.e., the reconstructed jet *after* detector effects calibrated to the EM energy scale. The correction factor is a function of the transverse momentum and the pseudorapidity of the jet to account for varying responses due to different detector components and transition between modules.

After the correction of the jet energy scale has been applied, one obtains a mostly uniform response.

Global sequential calibration: The previous calibration steps corrects the energy of the jets such that it results in an response that is uniform – *on average*. On an event-on-event basis, however, the calorimeter response is sensitive to features of the shower development like the distribution of energy in the calorimeter as well as the particle content of the jet. On average, one observes a correlation between details of the shower development and the initializing particle that gave rise to the jet in the first place. This dependence on the shape of the shower and the original particle is particular pronounced for quarks and gluons that differ in terms of the number of constituents as well as the shower width due to different color factors (see Chapter 1.2.3). Furthermore, quark-initiated jets tend to come with higher momentum emissions since the energy of the initial quark is distributed over a smaller number of constituents. This results in an on average deeper shower dept of quark-initiated jets compared to those originating from gluons. These fundamental differences between quark- and

gluon-initialized jets can be used to derive a set of (simple) discrimination variables to further improve the resolution of the jet energy scale by means of the so-called Global Sequential Calibration (GSC).

The GSC corrects the dependence of the jet response for the following variables in order to reduce flavor dependence (adopted from Ref. 21, list on p. 12):

- f_{Tile0}, the fraction of energy measured in the first layer of the hadronic Tile calorimeter;
- f_{LAr3}, the fraction of energy measured in the third layer of the electromagnetic LAr calorimeter;
- $n_{\mathrm{trk}}^{>1\,\mathrm{GeV}}$, the number of tracks with ghost-associated with the jet;
- $\mathcal{W}_{\mathrm{trk}}$, the average p_{T}-weighted transverse distance in the $\eta - \phi$ plane between the jet axis and all tracks with $p_{\mathrm{T}} > 1\,\mathrm{GeV}$ ghost-associated to the jet;
- n_{seg}^{μ}, the number of muon track segments ghost-associated with the jet.

This list of variables clearly expresses the above mentioned statements regarding the difference between quark- and gluon-initialized showers into simple one-dimensional discrimination variables.

The last term, n_{seg}^{μ}, is also known as *punch-through* correction. It corrects for energy leakage when a high-p_{T} jet is not fully contained in the calorimeter, but "punches through" it and leaves traces in the outer muon spectrometer. This is a rare phenomenon that usually only occurs for the most energetic jets in the central region of the detector with the smallest radiation/absorption lengths.

For each variable listed above $\zeta \in \{f_{\mathrm{Tile0}}, f_{\mathrm{LAr3}}, n_{\mathrm{trk}}^{>1\,\mathrm{GeV}}, \mathcal{W}_{\mathrm{trk}}, n_{\mathrm{seg}}^{\mu}\} =: I$, a correction factor $c(\zeta) \propto 1/\langle R(\zeta)\rangle_{(p_{\mathrm{T}},\eta)}$ is derived successively. In this case, the response $\mathcal{R}$ is a function of the transverse momentum and a jet feature ζ in bins of pseudorapidity η to take into account different detector regions. The final correction factor due to the GS calibration is then given by $c_{\mathrm{GSC}} = \prod_{i\in I} c_i(p_{\mathrm{T}}^{i-1}, \zeta)$.

Residual *in-situ* calibration: Up to this step, all calibration corrections are based solely on Monte Carlo that aim to bring the EM jet energy scale to particle level. The fact that all correction factors being derived from Monte Carlo involves the danger of biasing the adjusted result towards mismodelings present in the simulations since MC – undoubtedly – does not provide a perfect representation of data. To address this problem, a series of *in-situ* (meaning: *"Occurring or situated in its local or natural environment, or at the locality in question"* [138]) validation techniques is applied in which the *calibrated* transverse momentum of the jet is compared to the p_{T} of a well-defined, -measured *reference object* in data.

The *in-situ* correction is based on a double-ratio of the transverse momentum measured in data and MC according to

$$\frac{1}{c_{\mathrm{in-situ}}} = \frac{\mathcal{R}_{\mathrm{data}}^{\mathrm{in-situ}}}{\mathcal{R}_{\mathrm{MC}}^{\mathrm{in-situ}}} = \frac{\langle p_{\mathrm{T}}^{\mathrm{jet}}/p_{\mathrm{T}}^{\mathrm{ref}}\rangle_{\mathrm{data}}}{\langle p_{\mathrm{T}}^{\mathrm{jet}}/p_{\mathrm{T}}^{\mathrm{ref}}\rangle_{\mathrm{MC}}}, \tag{4.12}$$

whereby $p_{\mathrm{T}}^{\mathrm{ref}}$ denotes the transverse momentum of the balanced reference object in data and MC, respectively.

The *in-situ* calibration is sub-divided into three corrections: η-intercalibration, Z/γ+jet and MultiJet Balance (MJB), all three of which are briefly introduced in the following paragraphs.

The *η-intercalibration* uses well-measured and well-calibrate jets in the (very) central region of the detector with a pseudorapidity of $|\eta| < 0.8$ to derive a residual calibration for jets in the forward region for jets with $0.8 < |\eta| < 4.5$. The p_{T} balance is quantified by the asymmetry $\mathcal{A} = (p_{\mathrm{T}}^{\mathrm{probe}} - p_{\mathrm{T}}^{\mathrm{ref}})/p_{\mathrm{T}}^{\mathrm{avg}}$, with $p_{\mathrm{T}}^{\mathrm{probe}}$ belonging to the forward jet, and relative jet response is computed accordingly $\left\langle p_{\mathrm{T}}^{\mathrm{probe}}/p_{\mathrm{T}}^{\mathrm{ref}}\right\rangle \approx (2 + \langle\mathcal{A}\rangle)/(2 - \langle\mathcal{A}\rangle)$.

The $Z+\mathrm{jet}$ *and* $\gamma+\mathrm{jet}$ *balance* method allows for the calibration of jets up to a transverse momentum of 950 GeV in the central region of the detector ($|\eta| < 0.8$) based on measurement of well-measured leptons and photons through $Z \to e^+e^-$ and $Z \to \mu^+\mu^-$. The reason for

restricting this calibration method to the central region of the detector is the statistical limitation of those events.

The *multijet balance* forms the last step of the *in-situ* calibration and serves the purpose of extending the calibration to an transverse momentum of 2 TeV. In this correction, events with at least three jets are selected with one jet being of "large" transverse momentum ($p_{\mathrm{T}} > 950\,\mathrm{GeV}$) that is balanced against system consisting of several lower-p_{T} jets. The transverse momentum of the individual lower-p_{T} recoiling jets must be adequately to be in the regime of the Z/γ+jet calibration to ensure that only fully calibrated objects are used.

Flavor-related uncertainties: The *flavor composition uncertainty* accounts for differences in the response for quark $\mathcal{R}_q$ and gluon $\mathcal{R}_g$ jets, with $\mathcal{R}_q$ and $\mathcal{R}_g$ being the responses as "measured" in Pythia. This difference is, as already stated, a consequence of the quark and the gluon being different representations of the SU(3) symmetry group of QCD, which results in different couplings and hence different phenomenology regarding their showers. Generally, the fraction of quark-initialized jets in a sample is much better determined due to constraints from measurements at LEP, which have been used to tune the Pythia event generator (*"including flavor-dependent fragmentation measurements"* [139]). Given the uncertainty σ_g^f on the fraction of *gluon*-initialized jets in a sample f_g (the corresponding uncertainty on the quark fraction f_q is negligible due to $\sigma_g^f \gg \sigma_q^f$), the uncertainty of the Flavor Composition (FC) is given by

$$\sigma_{\mathrm{FC}} = \sigma_g^f \frac{|\mathcal{R}_q - \mathcal{R}_g|}{f_g \mathcal{R}_g + (1 - f_g)\mathcal{R}_q}. \tag{4.13}$$

As one can see, this uncertainty is directly proportional to the residual in the response and the uncertainty on the gluon fraction [45].

The flavor response uncertainty is the second flavor-related uncertainty; it accounts for the fact that while most of the commonly used Monte Carlo generators exhibit a good level of agreement regarding the modeling of the quark response $\mathcal{R}_q$, they differ "considerably" concerning the response of gluon-initialized jets $\mathcal{R}_g$. This is of course due to the already described circumstance that most MC generators on the market incorporated LEP constraints into their tunes – for gluon jets, those constraints are statistically quite limited[4]. Consequently, this uncertainty is defined (due to the absence of alternatives) by the direct comparison of the gluon response $\mathcal{R}_g$ between two Monte Carlo generators

$$\sigma_{\mathrm{resp}} = f_g \left(\mathcal{R}_g^{\mathrm{Pythia}} - \mathcal{R}_g^{\mathrm{Herwig}} \right). \tag{4.14}$$

4.4.4 Charged Particle Jets

Jets build from topological clusters contain both: contributions from charged and neutral particles. This has the advantage of taking all constituents of the jets into account and hence allows to reconstruct the full four-momentum of the particle that initialized the jet. Furthermore, it permits analytic calculations of jet-related observables. The downside of using jets solely build from calorimeter information is the significantly worse spatial resolution compared to the inner part of the Detector. For this reason, jets exclusively build from topo-clusters are usually not suitable for jet substructure measurements (like this one) – or only to a very limited extent. To address this limitation, tracking information must be taken into account to allow for better resolution of the individual constituents inside a jet. There are two main approaches:

Track jets: Track jets are jets *directly* build from reconstructed tracks in the inner detector. This means, instead of topo-clusters, tracks are used in the jet finding algorithm to

[4]This can easily be understood by recalling that in a e^+e^--collider with a centre-of-mass energy close to the Z mass resonance (like LEP(II)), the cross section for $e^+e^- \to Z^* \to q\bar{q}$ is very large. To obtain a signature of a gluon in the final state, needs at least three jets in the final state $e^+e^- \to Z^* \to q\bar{q}g$. Since this is a NLO process, the rate of gluon jets is suppressed roughly by a factor of $\alpha_{\mathrm{S}}(M_Z)$, resulting in approximately two OOM less gluon jets compared to quarks.

reconstruct jets; the result might be, for instance, an anti-k_t $R = 0.4$ track jet. Since their isn't any calibration scheme for track jets available at ATLAS, those types of jets are usually not used in measurements since no energy-scale-related uncertainties exists. They are, however, commonly used in jet flavor tagging applications where the tagger's decision is based on detailed substructure information of the jet. Since taggers usually are calibrated to data directly, the aforementioned lack of a calibration scheme for track jets does not pose a problem.

"Jet-associated tracks": To avoid the above-mentioned problem involving track jets, one can instead use the set of tracks that are associated to a jet that has been reconstructed using, e.g., calo-clusters or PFlow Objecs (PFOs) as an input. As described in the previous section, jets build from topo-clusters are fully calibrated for jets with $R = 0.4$ and $R = 1.0$, respectively. After the tracks are associated to the jets based on a criterion to be specified in the next paragraph, the jet is reclustered using the matched tracks only and discarding the PFOs for good[5]. This approach usually results in very different jets compared to those obtained directly build from tracks and therefore excluding all neutral particles in the first place. One should take note, however, that track jets (as described above) usually contain more information about the substructure of a jet compared to associated track jets. Consider, for instance, the following process $gg \to t\bar{t} \to bW^+(\to q\bar{q})\,\bar{b}W^-(\to q'\bar{q}'')$ with the decay products of the W bosons being highly boosted. In case of topo-cluster jets, it's likely for the remnants of the W to be reconstructed in one single jet; for track jets, however, the higher angular resolution of the ID still allows to distinguish between the two jets. The track-to-jet association is usually done using *ghost- or ΔR-association*, which will be explained in the following paragraph.

The concept of a *ghost* particle was first introduced in an attempt to provide a meaningful definition of the "area" or (constituent) "catchment area" of a jet by *"[...] adding a dense coverage of infinitely soft 'ghosts' and counting how many are clustered inside a given jet[.]"* [136]. This sounds somewhat cumbersome, but it results in a well-defined notion of a jet area contrary to just using point-like constituents. Ghost-track association adopts this idea and utilizes it for track-to-jet association. In this association scheme, the momenta of all tracks in the event are temporarily set "infinitesimally" small – they can not be simply set to zero otherwise they are not clustered – prior to the application of the jet finding algorithm. After this step, jets are reconstructed, using topo-clusters and the infinitesimally small-p_{T} tracks, applying the desired jet finder. This may result in pure ghost-jets artifacts with infinitesimally small p_{T} that are discarded. After this step, all tracks inside the catchment area are taken to be associated to the respective jet.

An alternative approach to the ghost-based method would be a simple ΔR matching scheme with respect to the jet axis. However, the ΔR-cone-based track association fails to provide meaningful jets for recombination algorithms, which result in highly irregular jet areas, e.g., for C/A or k_t (see Figure 4.2).

4.5 Missing Transverse Energy $\boldsymbol{E}_{\mathbf{T}}^{\mathbf{miss}}$

At hadron colliders like the LHC, the full momentum p of the initial partons that participate in the hard interaction is fundamentally inaccessible on an event-on-event basis as it is subject to statistical fluctuations according to the parton density functions. This implies that the initial state of the interaction is unknown. The three-momentum of the hadron, however, can be decomposed into a component transverse, i.e., perpendicular p_{T} and longitudinal $p_{\parallel}$ to the direction of motion along the beam line $p = p_{\mathrm{T}} + p_{\parallel}$. Since the colliding protons move at almost the speed of light, the proton can (to good approximation) be described in the so-called "infinite momentum" or "Breit" frame, in which the particle mass can be neglected. Furthermore, the transverse component of the proton *and* its parton constituents is completely negligible compared to the longitudinal part. If the transverse momentum p_{T} vanishes in the

[5] This is not exactly what is done in this measurement. In this measurement, PFlow-based jets as well as tracks matched to the former ones enter the calculation of the triple-differential multijet cross section. The set of matched tracks is used to compute the jet substructure variable, while the PFlow-based jets are used to reconstruct the transverse momentum as well as the relative pseudorapidity.

initial state, the same must – of course – apply to the final state as well. This implies that the *total* sum of all final state particles' transverse momentum in the event must balance in order for the initial condition $\sum_i p_{\mathrm{T},i} = 0$ to be fulfilled. In practice, however, not all final state particles are reconstructed in the detector either because, e.g., their energy is too low to be detected, or, their interaction probability is very low as it is the case for neutrinos or for hypothetical particles extending the Standard Model. Therefore, the total three-momentum can be written as $\sum_i p_{\mathrm{T},i} = \sum_k p_{\mathrm{T},k}^{\mathrm{inv}} + \sum_j p_{\mathrm{T},j}^{\mathrm{vis}} = 0$, whereby $p_{\mathrm{T},j}^{\mathrm{vis}}$ denotes the total (visible) momentum of the reconstructed particles and $p_{\mathrm{T},k}^{\mathrm{inv}}$ being the contribution not reconstructed (non-visible) in the detector. Summarizing the non-reconstructed part of the total transverse momentum $E_{\mathrm{T}}^{\mathrm{miss}} := \left|\sum_k p_{\mathrm{T},k}^{\mathrm{inv}}\right|$, the *missing transverse energy* (or equivalently the *missing transverse momentum*) is given by

$$E_{\mathrm{T}}^{\mathrm{miss}} = -\left|\sum\nolimits_j p_{\mathrm{T},j}^{\mathrm{vis}}\right|. \tag{4.15}$$

The actual reconstruction of the missing transverse momentum is more complicated as one might expect from Equation 4.15. At ATLAS, $E_{\mathrm{T}}^{\mathrm{miss}}$ characterized by two main contributions: *firstly*, a so-called *hard term* that is calculated based on fully reconstructed and calibrated objects (leptons, photons and jets); *secondly*, a corresponding *soft term* summarizing signals not associated with any of the aforementioned hard objects – $E_{\mathrm{T}}^{\mathrm{miss}} = E_{\mathrm{T}}^{\mathrm{miss,hard}} + E_{\mathrm{T}}^{\mathrm{miss,soft}}$. The soft term is a significant contribution toä the total $E_{\mathrm{T}}^{\mathrm{miss}}$ and it needs to be reconstructed separately. Usually, the soft part is reconstructed from Inner Detector tracks that are associated to the primary vertex in the event, but are *not* connected with any of the hard objects.

5. Detector Unfolding[1]

Up to this point, the selection of chapters in this thesis represented a common workflow in particle physics starting with a (mathematical/quantitative) theory (Chapter 1) that is used to derive predictions by means of analytic calculations or MC simulation (Chapter 2) that can be compared to measurements in an experiment (Chapter 3) from which physical observables are derived (Chapter 4).

There is one step missing in this line of thought (besides the actual analysis idea of course): the observables reconstructed from measurements usually depend on the conditions the experiment has been performed. For instance, the cross section of a physical process is a *detector-independent* quantity that is (among other things) related to the total number of observed events in an particular experiment. In general, the measurement with the detector will distort particle properties which is why one usually distinguish between *Particle-Level* (PL) objects and *Detector-Level* (DL) objects. In order to make consistent comparisons to theory, the measured observables *must* be corrected for detector effects; this procedure is referred to as (detector) *unfolding*. After unfolding, the measured quantities are at PL, which not only allows for a direct comparison to theory (provided that the given observable is IRC-safe), but also to other results obtained from other experiments that have been corrected for detector effects.

5.1 Basic Concepts

Our picture of the *underlying* material reality surrounding us is invariably clouded due to the limiting precision of imperfect measuring apparatuses that lead to a distorted perception of the hidden "truth" similar to a view through gloomy glasses (this also includes our organs of perception of course). In physics – like in empirical sciences in general –, this discrepancy between the actual *form* of an *entity to be examined* and the representation of the *object that manifest in the experiment* must be taken into account when deriving general statements concerning reality based on observations or measurements. The reasons for this blurred image of nature are manifold ... ranging from limited technical precision of our experimental equipment, to known (and unknown) phenomena that affect the measurement, to absolutely fundamental limitations regarding our insight into the properties of most elementary constituents of material being (not to mention mental or holistic characteristics of, e.g., consciousness, love, friendship, beauty etc., which *a priory* withstand *any* gain of insight based on natural scientific approaches).

A simple example would be distortion of the light of the distant star on its way through the earth's atmosphere thence the light gets scattered which reduces the resolution on ground-based telescopes.

In case of elementary particle physics, experimental measurements are typically based on counting the occurrence of different categories of processes (commonly referred to as *events*), which can be used to determine, for instance, the cross section of a certain mechanism of interest given the luminosity of the accelerator. Depending on the researcher's interest and research goals, the measurement is further confined to specific phase space regions (usually referred to as *bins*); for each measurement bin, the number of events is counted after selection cuts have been applied to increase the purity of the specific process of interest. The measurand

[1] This chapter may safely be skipped provided that the reader is already familiar with the basic concepts of detector unfolding.

in this case is the event count, and there will be a disagreement between the expected number of events and (if there is a theoretical model available) the expected ones. The origin of this discrepancy can be reduced to the following three categories of effects (adapted from Ref. [140]): *detector effects*, *statistical fluctuations* and *contamination due to background processes*. Detector effects summarize differences between expected and observed distributions due to our incomplete knowledge of the measuring devices as a whole. A concrete example (which is quite relevant for the analysis presented in this thesis) is the imperfect reconstruction of the energy of a jet due to inactive material in the detector that causes the jet to be reconstructed in a wrong phase space bin. Those *migrations* between different phase space regions give rise to *fake* events that bias the cross section.

Since cross section measurements are based on statistically independent counting experiments, the observed number of events is expected to follow a Poisson distribution and hence accompanied by a *statistical* uncertainty that is given by the square root of the number of counted events. This uncertainty can be reduced by increasing the Luminosity (increase probability for interaction of the process of interest), by collecting more data over time or by combining several measurements – in summary: increase the number of events.

Last but not least, it is impossible to define perfectly pure regions of a specific process in phase space, fundamentally, due quantum mechanical indistinguishability and, experimentally, simply because of limited resolution and the therein resulting ambiguity of kinematic configurations. Furthermore, the estimation of background processes contribution to the total event count in one phase space bin may be inaccurate due to limited precision in the calculation of the event rate of the respective process or unknown phenomena contributing without being accounted for.

All of the aforementioned effects will result in the measured spectrum g of an observable $\mathcal{O}^{\mathrm{DL}}$ (whereby "DL" is an abbreviation for Detector Level) to differ from the Particle Level (PL) distribution f with $\mathcal{O}^{\mathrm{PL}}$ one would expect in case of an ideal detector with infinite event statistics and perfectly pure process samples. This makes the comparison and interpretation of measurements complicated since one assumed truth is associated to several experimental outcomes each one supposed to be a representative of the same underlying object. For this reason, one usually applies a correction to the measured observable in order to account for effects that cause the DL and the PL distribution to disagree. This procedure, which aims to map the observed measurement (which depends on the experimental conditions) to the truth one, is generally referred to as (detector) *unfolding*.

Mathematically, unfolding is formulated based on a Fredholm integral equation that relates the probability distribution over a PL observable $f(\mathcal{O}^{\mathrm{PL}})$ to the corresponding DL distribution $g(\mathcal{O}^{\mathrm{DL}})$ through the convolution with a *kernel* function $K(\mathcal{O}^{\mathrm{DL}}, \mathcal{O}^{\mathrm{PL}})$ according to

$$g(\mathcal{O}^{\mathrm{DL}}) = \int_{\mathrm{supp}(f)} \mathrm{d}\mathcal{O}^{\mathrm{PL}}\, f(\mathcal{O}^{\mathrm{PL}}) K(\mathcal{O}^{\mathrm{DL}}, \mathcal{O}^{\mathrm{PL}}), \tag{5.1}$$

whereby the integration volume denotes the support, i.e., $\mathrm{supp}(f) := \{\mathcal{O}^{\mathrm{PL}} \in \mathbb{X} \mid f(\mathcal{O}^{\mathrm{PL}}) \neq 0\}$ of the density function f.

In particle physics, the probability densities f, g are usually one-dimensional distributions over real numbers $\mathcal{O}^{\mathrm{PL/DL}} \in \mathbb{R}$ that are approximated by discrete histograms with an expected number of events ν_i (notation adopted from Ref. [41]) in the compact subset $U_i^g \subset \mathrm{supp}(g)$

$$\nu_i = \int_{U_i^g} \mathrm{d}\mathcal{O}^{\mathrm{DL}}\, g(\mathcal{O}^{\mathrm{DL}}), \tag{5.2}$$

with $\bigcup_{i=1}^{N} U_i^g \subseteq \mathrm{supp}(g)$ whereby $N \in \mathbb{N}$ denotes the number of *bins*. Similarly, the truth distribution is discretized over the set $\{U_i^f \subset \mathrm{supp}(g) \mid \bigcup_{i=1}^{M} U_i^f \subseteq \mathrm{supp}(f) \wedge U_i^f \cap U_j^f = \emptyset \text{ with } i \neq j\}$ with $M \in \mathbb{N}$ denoting the number of bins of the truth-related histogram(s) with an expected number of events

$$\mu_j = \int_{U_j^f} \mathrm{d}\mathcal{O}^{\mathrm{PL}}\, f(\mathcal{O}^{\mathrm{PL}}). \tag{5.3}$$

Consequently, the kernel function in Equation 5.1 is discretized too and approximated by a matrix R with $R \in \mathbb{R}^N \times \mathbb{R}^M$ henceforth referred to as *response matrix*. Properly normalized,

the response matrix $R(i,j) =: R_{ij}$ gives the probability to reconstruct an event with the truth value μ_j in phase space bin j as ν_i in bin i, i.e., for $i \neq j$ there are migrations between regions of phase space, for instance, due to incorrect energy measurements.

After approximating the probability densities f, g by means of histograms and describing the originally continuous kernel function K by a finite matrix with appropriate dimensionality, the Fredholm integral Equation 5.1 can be expressed in terms of a simple system of linear equations

$$\nu_i = \sum_{j=1}^{M} R_{ij}\mu_j + \beta_i, \tag{5.4}$$

which can be compactly written as a matrix equation $\boldsymbol{\nu} = \boldsymbol{R\mu} + \boldsymbol{\beta}$. The term β_i was added to account for the contributions of background events in the ith bin of the measurement.

In the (very) rare cases in which the PDF of the true distribution f is known analytically, the response matrix can be directly calculated provided that the integral function of f is well defined. Most of the time, however, the response matrix is derived from detailed simulations based on Monte Carlo methods that aim to mimic the experimental conditions as close as possible to account for effects not directly covered by uncertainties such as, e.g., dead material.

5.1.1 Construction of the Response Matrix

As outlined in Section 5.1, the response matrix encodes correlations between objects at PL and DL. Properly normalized, it corresponds to the probability to reconstruct the truth-value μ_j in phase space bin j as ν_i in bin i. Due to the limited resolution of all measuring apparatuses, missing information of the event may result in (potentially significant) migrations ($i \neq j$) between different regions of phase space, which in turn give rise to large off-diagonal elements in the response matrix. Since the size of systematic uncertainties is (roughly) proportional to the magnitude of the non-diagonal entries of the response matrix, the measurement should be constructed with the aspiration to obtain response matrices being *as diagonal as possible*; usually, the decisive factors that control the fraction of elements on the diagonal are the objects being used to reconstruct the observable of interest (for instance topoclusters versus tracks) as well as the binning of the phase space.

To construct the response matrix, the object of interest, e.g., a jet, must be reconstructed at PL *as well as* at DL. If this is not the case, the event does *not* contribute to the response matrix (because there is no jet to correlate with), but it affects the so-called efficiency or fake factor depending on whether the PL or the DL object fails the event selection. The reconstruction efficiency as well as the fake rate are the topic of the following section.

Subsequently, both object must be matched to each other in order to get a one-to-one mapping between particle and detector level. This matching scheme (in case of jets, the ΔR or ghost-matching scheme are common choices) should be as efficient as possible to avoid biases and a dependence on the respective matching scheme. Once the assignment has been established, the observables (with reference to the (differential) cross section) is measured can be calculated at PL and at DL, which is the information that enters the response matrix.

5.1.2 Efficiency and Fake Factors

The response matrix corresponds to the correlation between the observable(s) measured at PL and DL. In order to construct the response matrix, there must be a well-defined one-to-one correspondence between an object reconstructed at particle level and a corresponding one at detector level. This matching scheme, however, is not unique, and different assignment criteria may result in different response matrices. However, for each mapping scheme, it may happen that no suitable candidate at PL or DL is found. For instance: the limited resolution in some regions of the detector may result in a significant underestimation of the actual jet energy, giving rise to migrations into different regions of phase space[2], which may fall outside of the

[2]For the sake of completeness, it should be mentioned that particle- and detector-level objects are elements of two different spaces that are at most isomorphic to each other.

phase space volume defined by the selection criteria of the respective analysis. In this case, the event/jet can *not* be used to populate the response matrix as there is no correlative at PL and/or DL to correlate to. This situation is schematically illustrated in Figure 5.1. In Figure

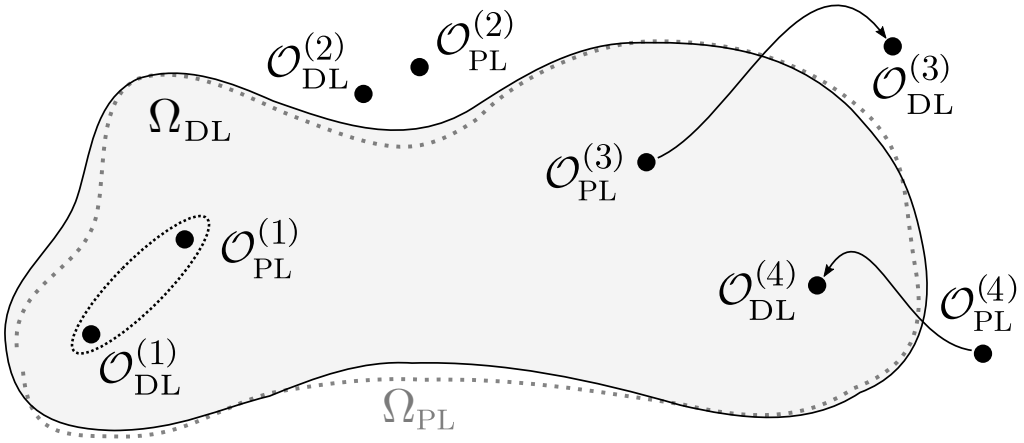


Fig. 5.1: Schematic illustration of the phase space volume of the measurement and some elements of truth- and reco-level objects. For $\mathcal{O}^{(1)}$, PD and DL are close to each other and within the PS such that they can uniquely be matched; for $\mathcal{O}^{(2)}$, both are inside of the volume and the event is not filled in the response matrix; $\mathcal{O}^{(3)}$, causes a loss in efficiency, and $\mathcal{O}^{(4)}$ causing fakes.

5.1, the reconstructed observable $\mathcal{O}^{(i)}_{\mathrm{PL/DL}}$ for event number i at PL and DL, respectively, both fall into the active phase space volume defined by the selection and hence can be matched to each other to populate the response matrix. In case of event j and k, however, the measured value of the observable at DL $\mathcal{O}^{(j)}_{\mathrm{DL}}$ (or at PL $\mathcal{O}^{(k)}_{\mathrm{PL}}$) fail the selection criteria. Of course, it also happens that neither PL nor DL object are outside of the phase space boundary; those events, however, are not considered to be of interest for the analysis.

Events for which no corresponding matching partner (either at PL or DL) exists can not simply be ignored as those effects are subject to modeling differences between different Monte Carlo generators and hence play an important role for the model-related uncertainties of the measured cross section. Furthermore, the cross section would get under- or overestimated.

In the following, a distinction has to be made between, *firstly*, events for which the observable at PL lies within the phase space volume while the corresponding object at DL does not pass the event selection, and, *secondly*, *vice versa*.

The former scenario results in a loss in *efficiency* because the event should enter the measurement as the truth-level object passes the selection. Mathematically, this effect is described by the *efficiency factor* according to

$$\epsilon_{\mathrm{eff}} = \frac{N^{\mathrm{DL}\wedge\mathrm{PL}}}{N^{\mathrm{PL}}}, \tag{5.5}$$

whereby $N^{\mathrm{DL}\wedge\mathrm{PL}}$ denotes the number of events that pass the selection at DL *and* PL and N^{PL} being the total number of PL events that pass the truth-level selection.

Similarly, the latter one corresponds to the situation where the event is not supposed to enter the active measurement volume, but due to migrations, the event is incorrectly accounted for at DL. Those events are referred to as *fakes* because detector inefficiencies "fake" an event which actually is supposed to be rejected by the selection. Thus, the corresponding variable quantifying the amount of falsely counted events is referred to as *fake factor* defined as

$$\epsilon_{\mathrm{fake}} = \frac{N^{\mathrm{PL}\wedge\mathrm{DL}}}{N^{\mathrm{DL}}}, \tag{5.6}$$

with N^{DL} being the total number of events passing the selection at DL.

Since both, efficiency (see Equation 5.5) and fake (see Equation 5.6) factor, affect the

number of counted events, the final (triple-differential) cross section must be corrected for those effects. Furthermore, because the reconstruction efficiency as well as the fake rate are derived from simulation, they are sensitive to modeling effects of different Monte Carlo generators and therefore must be taking into account in the model-related uncertainties (this topic will be discussed in great detail later in Chapter 10 when all sources of statistical and systematic uncertainties are discussed).

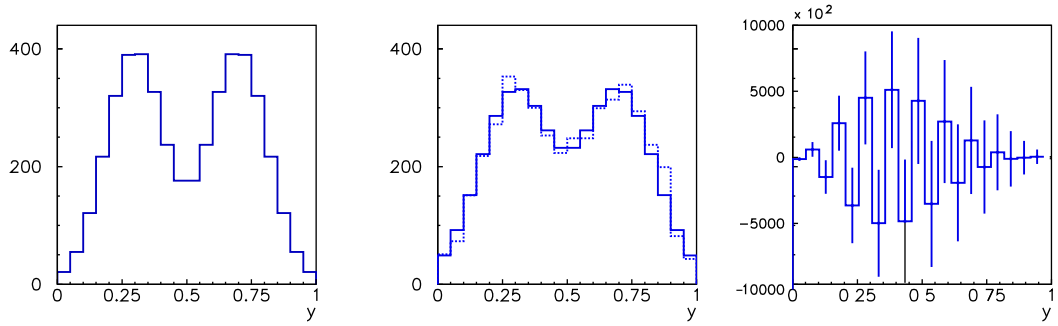
Finally, given the response matrix R, the efficiency factor ϵ_{eff} and fake factor ϵ_{fake}, the corrected number of events at DL is given by

$$N_i^{\text{PL}} = \sum_j N_{\text{DL}}^j \left(\left[R^{-1} \right] \right)_{\text{DL},j}^{\text{PL},i} \cdot \frac{\epsilon_{\text{fake},j}}{\epsilon_{\text{eff},i}}, \tag{5.7}$$

whereby the sum runs over all phase space bins. In Equation 5.7, $\left(\left[R^{-1} \right] \right)_{\text{DL},j}^{\text{PL},i}$ corresponds to the (i,j)th element of the inverted response matrix R. Regarding the inversion of the response matrix, different methods are available, which are subject of the upcoming section of this chapter.

5.2 Unfolding Methods

Equation 5.4 suggests a "simple" inversion of the response matrix, which is then applied to the measurement at DL to obtain the corresponding spectrum at PL. However, the "naive" inversion of the response matrix may lead to undesired behavior as shown by the often cited (and just as often misunderstood) example in Figure 5.1. The approach to the unfolding problem can also be



(a) True *(b)* Expected/measured *(c)* Unfolded

Plot 5.1: Illustration of possible "problems" of the unfolding procedure. The figures show the underlying truth distributions (*left*), the expected (solid line) and the observed/measured (dashed line) distribution (*middle*), as well as the unfolded result (*right*) (the figures have been adopted and modified from Ref. 41, Fig. 4 p. 5).

thought of as the maximum likelihood solution, which therefore provides an *unbiased* estimator with the smallest variance possible. However, it may cause some unwanted side effects. This is shown in Figure 5.1. The leftmost figure shows a double-peaked *true* distribution; the solid line in Figure 5.1b corresponds to the folded truth-level distribution with the dashed line being the data that was sampled according to a Poisson distribution. Finally, the rightmost figure shows the unfolded result. Obviously, the unfolded result in Figure 5.1a looks quite different than the one expected from Figure 5.1a, which was supposed to be recovered. The reason for this apparent "failure" of the method is due to the presence of considerable fraction of non-diagonal elements in the response matrix, which may occur if, e.g., the chosen binning of distributions does not account for the limited resolution of the detector. In this case, significant migrations are the consequence. The large fluctuations that can be observed in Figure 5.1c follow from the inverted response matrix that is applied to an instance vector of the data instead of the expectation value it was derived from. In the later case, the unfolding procedure recovers the correct truth value; fluctuations in the data, however, are magnified due to large correlations

present between bins. It is very important to note that the unfolded spectrum is not wrong – it just isn't desirable because the resulting distribution is not smooth and does not look like the one expected at truth level. Furthermore, the simple inversion approach does not allow for different binning choices at PL and DL, but limits the family response matrices to square matrices. The objective of alternative unfolding methods is therefore to find a reasonable compromise between a bias (which is not present in the simple matrix-inversion method as the estimator is unbiased) and the smoothness of the final unfolded distributions. A quick summary for common unfolding methods is shown in this section [41].

The smoothing of the unfolded distribution is archived by applying a regularization scheme. In the likelihood formulation of the unfolding problem (the estimator according to Equation 5.4 can be considered as the solution to a likelihood function $\mathcal{L}(\boldsymbol{\mu})$ with $\partial_{\mu_i} \log \mathcal{L}(\mu_i) = 0$), regularization can be thought of as an additional term $S(\boldsymbol{\mu})$ to the (log-)likelihood $\log \mathcal{L}(\boldsymbol{\mu}) + \lambda S(\boldsymbol{\mu})$, whereby $\lambda \in \mathrm{R}$ denotes to regularization parameter that controls the strength of the additional constrain and therefore can be interpreted in terms of a Lagrange multiplier. The concrete form of the regularization function $S(\boldsymbol{\mu})$ is subject of the respective regularization schemes. The following list contains a brief summary of regularization schemes commonly used in particle physics (quoted almost verbatim from the following, excellent Reference [41] by Francesco Spanó): Tikhonov schemes, Maximum-entropy schemes, the family of Non-iterative Bayes-inspired schemes, Iterative unbinned schemes, using event-by-event weights etc.

The family of *iterative* schemes was not mentioned above since it is used to unfolding the triple-differential cross section measurement presented in this thesis and therefore deserves a little more detailed treatment.

The unfolding scheme used for this analysis (following the official recommendations of the ATLAS collaboration) is based on a method that was introduced in 1994 by G. D'Agostini and is commonly referred to as (iterative) Bayesian unfolding [141]. As the name already insinuates, D'Agostini's unfolding method is based on Bayes' theorem. In this picture, the smearing over a reconstructed observable at detector level $\mathcal{O}^{\mathrm{DL}}$ is interpreted as a conditioned probability distribution $P(\mathcal{O}^{\mathrm{DL}}|\mathcal{O}^{\mathrm{PL}})$ given the truth value of the observable $\mathcal{O}^{\mathrm{PL}}$. In an ideal detector, this probability distribution would correspond to the Dirac delta distribution $\delta(\mathcal{O}^{\mathrm{DL}} - \mathcal{O}^{\mathrm{PL}})$. Correspondingly, there is a conditioned distribution $P(\mathcal{O}^{\mathrm{PL}}|\mathcal{O}^{\mathrm{DL}})$. The two conditioned probability distributions, $P(\mathcal{O}^{\mathrm{PL}}|\mathcal{O}^{\mathrm{DL}})$ and $P(\mathcal{O}^{\mathrm{DL}}|\mathcal{O}^{\mathrm{PL}})$, are related by Bayes' theorem according to

$$P(\mathcal{O}_i^{\mathrm{PL}}|\mathcal{O}_j^{\mathrm{DL}}) = \frac{P(\mathcal{O}_j^{\mathrm{DL}}|\mathcal{O}_i^{\mathrm{PL}}) P_0\left(\mathcal{O}_i^{\mathrm{PL}}\right)}{\sum_k P(\mathcal{O}_j^{\mathrm{DL}}|\mathcal{O}_k^{\mathrm{PL}}) P_0(\mathcal{O}_k^{\mathrm{PL}})}, \tag{5.8}$$

with $P_0(\mathcal{O}_k^{\mathrm{PL}})$ denoting the prior over the truth values. The information of the conditioned distribution $P(\mathcal{O}^{\mathrm{DL}}|\mathcal{O}^{\mathrm{PL}})$ is basically given by the response matrix that is derived in simulation using Monte Carlo. The objective of the iterative Bayesian approach is to (approximately) solve Equation 5.8 to obtain $P(\mathcal{O}_i^{\mathrm{PL}}|\mathcal{O}_j^{\mathrm{DL}})$, i.e., the distribution over the observable at PL given a measurement at DL. Given the solution to Equation 5.8, the unfolding problem according to Equation 5.7 can be written as

$$N_i^{\mathrm{PL}} = \sum_j P(\mathcal{O}_j^{\mathrm{DL}}) P(\mathcal{O}_i^{\mathrm{PL}}|\mathcal{O}_j^{\mathrm{DL}}) \cdot \frac{\epsilon_{\mathrm{fake},j}}{\epsilon_{\mathrm{eff},i}}. \tag{5.9}$$

Equation 5.8 can be applied *iteratively* whereby the *prior* distribution is updated after each iteration. Each iteration corresponds to a regularization of the resulting response matrix; therefore, the number of iteration is a crucial parameter of the method that allows to balance variance and the bias of the resulting estimator. In practice, the number of iterations is determined based on the configuration that results in the *smallest* total uncertainty of the measurement.

6. Machine Learning[1,2]

Leaving particle physics aside for a moment, this chapter is supposed to provide a brief yet substantial introduction into the broad area of machine learning with an emphasis on deep neural networks. In the context of the measurement(s) presented in this report, machine learning is used to train a *binary* classifier with the objective to discriminate between quark- and gluon-initialized jets. The score predicted by the neural networks z_{NN} is then used as an (one-dimensional) jet substructure observable, with regard to which the triple-differential multijet cross section is measured and unfolded to correct for detector-related effects (see Chapter 5). This approach differs from the most common application of machine learning within the context of high-energy particle physics, i.e., the utilization as a classifier model that is used to discriminate between signal and background processes for, e.g., particle searches. Nonetheless, all models used in this analysis are trained in a similar fashion to ordinary classifier networks, that is, they are optimized regarding their discrimination power between showers originating from (light) quarks and gluons, respectively.

The neural networks that are used to assign a score to each jet are based on the "Deep Sets" architecture, which allows to constrain the function learned/approximated by the neural network to be permutation invariant, i.e., the prediction of the model does not depend on the arbitrary ordering scheme imposed on the input arguments. Limiting the set of models to *functions over sets* that are permutation invariant *and* allow for variable-length inputs is attractive because it naturally respects symmetries commonly associated with jets. For instance, the number of emissions inside of a jet is a stochastic process that (within statistical physical laws) varying from jet to jet; hence, the model should not be limited to a certain number of constituents. Furthermore, the jet is a phenomenological object of underlying QCD-related processes that is determined by its final state configuration. To "account" for quantum-mechanical indistinguishability, the probability assigned to each jet should be agnostic with respect to the arbitrary ordering imposed on the constituents. Besides the physics' motivation, the permutation, variable-length parameterization of the model is interesting as it improves training stability and opens a complete new space of functions that can be approximated by means of a neural network.

Regarding the structure of this chapter, the first section (6.1) serves as a description as well as a recapitulation of the historical development of neural networks and machine learning – from its modest beginnings in the mid-twentieth of the previous century to the highly complex, state-of-the-art models that are in the focus of today's research. By doing so, the section is limited to "milestones" (according to the author's subjective opinion) that caused a significant step forward in the area of machine learning or even – to cite the words of the physicist and philosopher Thomas S. Kuhn [142] –, a change of paradigm. Following the historical overview, the fundamental principle behind artificial neural networks and the feed-forward (6.2) topology are introduced. Now, equipped with a basic understanding of the fundamental concepts behind neural networks, the next section (6.3) introduces the aforementioned "Deep Sets" architecture that allows to learn permutation agnostic functions. This section is limited to the key essential

[1] Essential parts of the succeeding text passages have been adopted (almost) literally from my Master's thesis "*Towards a Data-Driven Simulation of QCD Radiation with Generative Models utilizing Machine Learning Methods*", 2019, Universität Heidelberg, Christof Sauer [49].

[2] This chapter may safely be skipped provided that the reader is already familiar with the basic concepts of machine learining.

aspects of the "Deep Sets" architecture/parameterization, with the concrete application in the context of this measurement being subject to an subsequent chapter. Following the introduction of the model, the next section (6.4) covers more technical aspects that are important when working with machine-learning approaches like, e.g., the definition of the loss function or the update of the model's parameters by means of gradient descent and the backpropagation algorithm.

6.1 A Brief History of Neural Networks

The highly advanced and complex computer models, which more and more find their way into the daily life of modern societies, come with a long and rich history that goes back even further than the appearance of the very first computers – at least conceptually. The principle idea is/was to model the functionality of the (human) brain with its enormous amount of interconnections between individual neurons and the activation potential within these cells. The growing understanding of neurobiology and the processes that govern the human brain in the last three centuries therefore greatly influenced the early concepts of artificial neural networks .

At the beginning of the eighteenth century, biologists and physicists alike studied the effect of electricity on the nervous system. The Italian naturalist Luigi Galvani [143] first demonstrated the presence of electrical currents in animal tissues. Later in 1871, Julius Bernstein, with the help of Emil du Bois-Reymond, deepened the knowledge of the processes in the nervous system with his *"Membrane Theory of Electrical Potentials"* [144], which is considered to be the first description of the action potential. Shortly thereafter, the Spanish neuroscientist and pathologist Santiago Ramón y Cajal [145] reported the first detailed anatomy of the nervous system, which later inspired scientists to adopt and apply those novel insights to mathematical and computer models.

The idea to model the functionality of the brain with its interconnections of neurons – a term that was first introduced by Heinrich W. G. von Waldeyer-Hartz [146] – and hence the beginning of artificial neural networks is often taken to be the research article *"A Logical Calculus of Ideas Immanent in Nervous Activity"* of Warren McCulloch and Walter Pitts in 1943 [147]. With the starting gun being fired, it did not take long until the concepts introduced by McCulloch and Pits received the attention of the scientific community. The '40s of the previous century also had a profound influence on brain theory, whereby special emphasis should be placed on Donald Hebb's *"Organization of Behavior: A Neuropsychological Theory"* in particular the therein introduced *"[.] Neurophysiological Postulate"* (also known as *Hebb's rule*) that describes the pre-post synaptic enhancement that is essential for the process of "learning" [148]. This concept was then translated to the artificial McCulloch-Pitts neuron by weighting each input. The gathered knowledge in neurobiology and the (at this time) hypothetical neural network models can be considered as important preparation work for the fast development of the field in the 1950s that was driven by the advancement of computers with increased computational power as well as significantly increased memory capacities.

The discoveries and scientific research by Warren McCulloch and Walter Pitts, Hebbs *et al.* culminated in Frank Rosenblatt's *perceptron* in 1962, which can be considered as the first artificial neural network in a modern sense. Frank Rosenblatt introduced the perceptron in his famous book *"Principles of Neurodynamics"* [149]. The perceptron is a very simple model that consists of a fixed number of inputs $\{x_1, x_2, \dots x_n\}$ (Rosenblatt only used five at the time) that each one being weighted with a real number $\{w_1, w_2, \dots w_n\}$ and resulting a weighted input $\{w_1 x_1, w_2 x_2, \dots w_n x_n\}$ with $x_k, w_k \in \mathbb{R}$ to the node (artificial neuron) where the inputs are finally added together $\sum_{k=1}^{n} w_k x_k$. Inspired by McCulloch and Pits' work, the artificial neuron had a binary activation according to the threshold value $\theta \in \mathbb{R}$ with $\sigma(x) = \Theta(x - \theta)$, where Θ is the Heaviside step function.

This simple model introduced by Rosenblatt was already quite successful; though, it was limited to a restricted family of functions it was able to approximate. For example, Rosenblatt perceptron is incapable representing a simple logical XOR (eXclusive-OR) or XNOR (eXclusive-NOR) gate due to its limitation to only learn a single, linear decision boundary. Rosenblatt was

not aware of these limitations; he – overly enthusiastically – proclaimed: *"Given an elementary [..]perceptron, a stimulus world W, and any classification $C(W)$ for which a solution exists; let all stimuli in W occur in any sequence, provided that each stimulus must reoccur in a finite time; then beginning from an arbitrary initial state, an error correction procedure will always yield a solution to $C(W)$ in a finite time [...]"* [149]. This initial elation caused the machine learning community to completely exaggerate the potential and applicability of neural networks at that time – disappointment and disillusionment was the consequence. In 1969, seven years after Rosenblatt's publication, however, Marvin Minsky and Seymour Papert published the essay *"An Introduction to Computational Geometry"* [150], which drew the attention to the inherent limitations of the perceptron and its restrictions to only learn linearly separable functions (the aforementioned XOR and XNOR gate is a simple example for a non-linearly separable function). The paper by Minsky and Papert had put the breaks on the initial euphoria of the novel method introduced by Rosenblatt and almost caused a complete standstill of research in the area until the beginning of the '80s.

After the initial hype in the '60s, the discipline of neural networks fell into a deep sleep. It was not until the year 1982, that the area awoke from the dark ages; rose out of the ashes of its eventful past in manifold dazzling colors, like the phoenix or the Egyptian Bennu. The age of machine learning is yet to come – for the better or the worse remains to be seen[3]. The person triggering the renaissance of neural networks was the American physicist John Hopfield of Caltech. In his directive essay *"Neural networks and physical systems with emergent collective computational abilities"* [151], he provided a structured and systematic analysis of the potential as well as the limitations of a neural network at that time. In 1985, he, Hopfield, in cooperation with D. W. Tank published another important paper: *"'Neural' computation of decisions in optimization problems"* [152].

The '80s mark a turning point and a rapid development of the field of machine learning boosted by the essay of Hopfield and the political circumstances of the cold war and the fear of the USA to be outperformed by other countries in this promising technology, the development in machine learning accelerated rapidly. In 1986 another method experienced its revival: a process called *backpropagation*. The backpropagation (automatic differentiation) algorithm was first introduced by Seppo Linnainmaa in 1970 [153]; however, the method was not appreciated until the year 1986 when David Rumelhart, Geoffrey Hinton, and Ronald Williams published a paper that used *"[l]earning representations by back-propagating errors"* [154]. This algorithm was the prerequisite to train the large complex networks emerging at that time.

Also, the theoretical understanding of the mechanisms and the mathematical fundamentals grew, enabling great steps forward. The paper by George Cybenko *"Approximation by Superpositions of a Sigmoidal Function"* from 1989 [155] is considered to be the first proof of the so-called *Universal Approximation Theorem* (UAT) for the special case on sigmoid activation functions solely. The theorem states as follows:

Theorem (**UAT**; Ref. 155, Thm. 1, p. 306)**.** *Let σ be any continuous discriminatory function. Then finite sums of the form*

$$G(x) = \sum_{j=1}^{N} \alpha_j \sigma(y_j^T + \theta_j)$$

are dense in $C(I_n)$ [space of real-valued continuous functions]. In other words, given any $f \in C(I_n)$ and $\epsilon > 0$, there is a sum, $G(x)$, of the above form, for which

$$|G(x) - f(x)| < \epsilon \quad \text{for all } x \in I_n.$$

(Proof: see Ref. 155, p. 306)

Two years later, Kurth Hornik provided another, more general proof of the universal approximation theorem [156]. The UAT can be considered as the theoretical basis of neural networks that tells us that – in principle – there are no limitations in potential applications –

[3] Probably the latter one ...

except the bounds of reality of course. Under the assumption of an arbitrary complex model, a neural network can approximate, respectively, learn *any* continuous function.

The continuous growth in complexity of today's networks come with an enormous amount of trainable parameters (weights) goes hand in hand with an increased demand on computer performance regarding the requirements of CPU power and RAM storage. For this reason, the complexity and the performance of neural network models scale with the enhancement of computing power, similar to parallelization of mathematical operations on modern GPUs.

With the universal approximation theorem at the beginning of the '90s, neural networks have established themselves. In the following decades, the field of machine learning and artificial intelligence have made considerable progress and numerous discoveries. Today, artificial intelligence is an integral part of our daily life – but not without controversy. It is hardly possible to foresee the implications of the increasing presence of machine learning methods and big-data analysis within the context of the progressive digitization and its repercussions on our daily life. However, it is very likely that this age marks the beginning of the era of "artificial intelligence".

The increased interest at the beginning of the '90s caused a cascade of numerous discoveries. One milestone was followed by another one so that it becomes difficult to distill the major steps that caused significant progress of the field as a whole. The most significant of these were – according to the author's opinion –: the "discovery" of recurrent artificial neural networks by John Hopfield [151] and recurrent neural networks derived therefrom, the Long Short-Term Memory (LSTM) (recurrent neural) networks introduced by Sepp Hochreiter and Jürgen Schmidhuber in 1997 [157] (the development of LSTM, which are much better at capturing long-term dependencies, were the main driver for the great success of RNNs in the applications of speech recognition); the invention of support vector machines as an algorithm for optimal margin classifiers [158]; reinforcement learning based on Markov decision processes [159] (for an overview, see Ref. [160]); the realization of generative models by means of generative *adversarial* neural networks [161] and autoencoders (conceptually introduced by Ballard [162]).

6.2 Artificial Neural Networks

The basic building blocks of an *artificial neural network* have already been introduced in Theorem 6.1. More Formally, a neural network can be described in terms of a triple $(\mathcal{N}, \mathcal{V}, w)$, whereby $\mathcal{N}$ corresponds to the set of *neurons*, whereat the *connection* between the pairs of neurons $(i, j) \in \mathcal{V}$ is given by the set $\mathcal{V} := \left\{(i, j)|i, j \in \mathbb{N}^{\geq|\mathcal{N}|}\right\}$. The *weights* are defined by the function $w : \mathcal{V} \to \mathbb{R}$ that gives the strength $w_{ij} := w(i, j)$ of the connection $(i, j) \in \mathcal{V}$. Hence, a weight $w_{i'j'} = 0$ cuts of the connection between neuron i' and j' (no stimulus). Furthermore, neural networks need a *propagation/transfer function* T_w that receives the output of several neurons $\{i_1, i_2, \cdots i_n\}$ with the output $(o_{i_1}, o_{i_2}, \cdots o_{i_n})$ connected via weights $\{w_{i_1,j}, w_{i_2,j} \cdots w_{i_n,j}\}$ to one neutron j and transforms it to $T_{w_j} := T_w(o_{i_1}, \cdots o_{i_n}, w_{i_1,j}, \cdots w_{i_n,j}) \in \mathbb{R}$. In principle, there is no restriction on T_w besides differentiability (which is required by the backpropagation algorithm). In practice, however, in almost *all* practical applications, the so-called *weighted sum* is used. This weighted sum is an affine transformation between outputs of the neurons and the weights

$$T_{w_j} = \sum_{i \in \{1, \cdots, n\}} o_i w_{i,j} + b_j, \tag{6.1}$$

where b_j is the so-called learnable *bias* $b_j \in \mathcal{B}$ term [163]. The actual neural network f, i.e., the function it learns/approximates defined above, is still rather limited in its applications since it is restricted to affine functions solely due to the chosen form of the propagation function T. Hence, the network is not able to approximate, e.g., non-linear functions since the concatenation of linear ($b = 0$) functions is linear again (homomorphism) according to

$$f = \bigcirc_{j \in \mathcal{N}} T_{w_j}^{b=0} = \bigcirc_{j \in \mathcal{N}} \left(\sum_{i \in \{1, \cdots n\}} o_i w_{i,j} \right) = \sum_{i \in \{1, \cdots n\}} \left(o_i \bigcirc_{j \in \mathcal{N}} w_{i,j} \right). \tag{6.2}$$

A neural network defined like Equation 6.2 would be nothing but a simple linear regression model. To improve the approximation power of general *feed-forward* neural networks, the output of the propagation function T_{w_j} is further processed by what is called the *activation function* (also known as *transfer function*) σ

$$\sigma_j = \sigma(T_{w_j}; \alpha_1, \cdots, \alpha_n), \tag{6.3}$$

with $\alpha_1, \cdots, \alpha_n \in \mathbb{R}$ being some adjustable *hyperparameters* (i.e., fixed parameters not learned by the network). The output of the activation function may also be "time dependent" $\sigma_j(t)$ and hence depend on previous activation. There exists a huge variety of different activation functions and choosing the "most appropriate" one is part of the model's adjustment. For example, Rosenblatt's perceptron introduced in Section 6.1 had a threshold function Θ. The choice of the activation function also depend on the network's task whether it is, e.g. a regression or a classification task. Thus, the latter one expects probabilities which requires an appropriate activation function that maps the network's output to the interval $[0, 1]$. The most popular activation functions used today are the *sigmoid*, *Fermi logistic function* $\sigma(z) = 1/(1 + e^{-z})$ (as it used for the proof of the UAT 6.1), the *hyperbolic tangent* $\cosh(z) = (e^z + e^{-z})/2$, and the so-called *Rectified Linear Units* (ReLU) $\mathrm{ReLU}(z) = \max(0, z)$ [164, 165] (for a comprehensive list of activation functions see, e.g., 166).

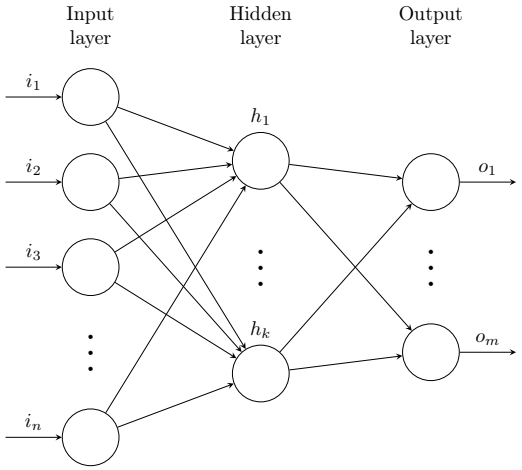


Fig. 6.1: "Shallow" feed-forward neural network with n inputs, one hidden layer with k neurons, and m outputs. The trainable weights of the networks are represented by the connections between the individual neurons.

ReLUs are the most frequently encountered activation functions in today's neural networks although it only has a non-linearly in one point. Compared to other saturating activation functions, however, ReLUs do not suffer from the vanishing or exploding gradient problem. Furthermore, the derivations needed for the backpropagation algorithm are simple to compute and hence significantly improve performance. On the other hand, ReLUs tends to result in sparse gradients. This effect can be reduced by using so-called Leaky ReLUs $\ell\mathrm{ReLU}(z) = \max(-\epsilon z, z)$ that adds a small slope $\epsilon \in \mathbb{R}^{>0}$ for negative arguments [167].

Summarizing the statements above, it can be said that neural networks are a composition of simple linear or affine transformations T and non-linear (activation) functions σ with trainable weights w, which can be thought as the interconnections between the network's neurons. These are the main components that *all* neural networks have in common. However, there exists a large number of different topologies or design concepts, i.e., the "geometrical" arrangement of the weights and its interconnections in the network. The following section introduces some of the most known and widespread network topologies.

The (by far) most common form of a network topology is the so-called *feed-forward* design. In this "conservative" topology, the network's neurons are grouped into different layers. A distinction is made between the *input* layers, *hidden* layers. and the *output* layer. The input

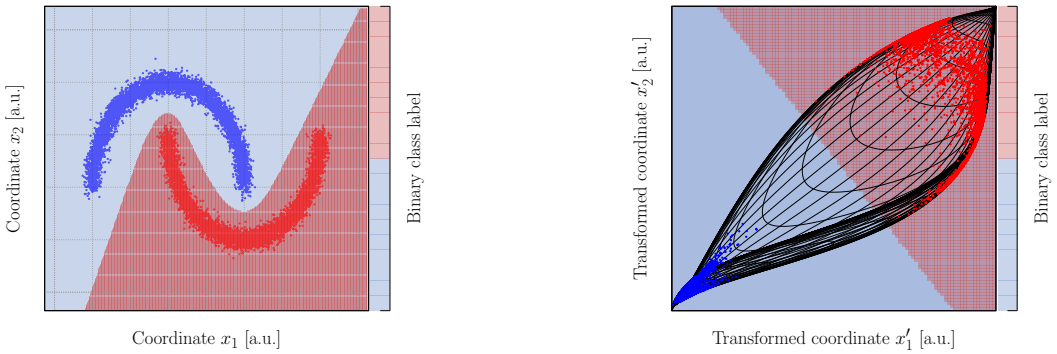
layer is the outermost level of a neural network that does not perform any computations on the data but serves as the "interface" between the outside world and the network. The hidden layers are the enclosed level of the neural network that is not directly connected to the outer world. All layers besides the input and the output layers are referred to as *hidden*. Within the hidden layers, the actual mathematical operations are performed that transform the data from layer to layer into different representations. Last but not least, the final layer is the output of the neural network that performs some final computations and returns the processed input data. The arrangement of weights and nodes (respectively neurons) can be summarized in what is known as a *directed graph*. Figure 6.1 shows a directed graph for a very simple feed-forward neural network with only one hidden layer (also referred to as "shallow" neural network.). As explained above, a feed-forward neural network is a composition of simple linear or affine transformations $T_w^{(l)} : \mathbb{R}^n \to \mathbb{R}^m$ and non-linear (activation) functions $\sigma^{(l)} : \mathbb{R}^n \to \mathbb{R}^n$ (with l denoting the lth layer in the network) that are arranged in layers without any feedback (loops), i.e., the output of any layer does not affect that very same layer. With all those elements and ingredients, a feed-forward neural network can be written as:

$$f(x) = \underbrace{\left[\sigma^{(L)} \circ T_w^{(L)}\right]}_{\text{output layer}} \circ \underbrace{\left[\sigma^{(L-1)} \circ T_w^{(L-1)}\right] \circ \cdots \circ \left[\sigma^{2} \circ T_w^{(2)}\right]}_{\text{hidden layers}} \circ \underbrace{\left[\sigma^{(1)} \circ T_w^{(1)}\right](x)}_{\text{input layer}}, \tag{6.4}$$

with x denoting the input to the input layer. So the output $f^{(l)}(x)$ of the lth layer is given by

$$f_w^{(l)}(x) = \begin{cases} x & : l = 0 \\ \left(\left[\sigma^{(l)} \circ T_w^{(l)}\right] \circ f^{(l-1)}\right)(x) & : l > 0 \end{cases}. \tag{6.5}$$

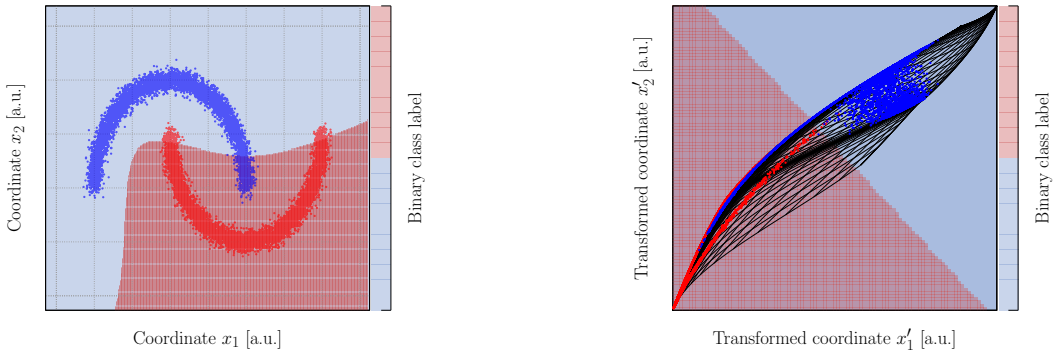
With the definition of a feed-forward neural network according to Equation 6.4, the system is mathematically well defined by means of its trained/adjusted weights (see Section 6.4).



(a) Raw data *(b)* Transformed data

Plot 6.1: Discrimination boundary, which was learned by the neural network, for the raw data (*left*) and its new representation in the transformed coordinate system of the second hidden layer (*right*).

Due to the choice of the propagation function to be linear or affine, the classification network tries to learn a new representation of the data (by means of a (homotopic) coordinate transformation) in each layer that allows for a separation of the transformed input by a hyperplane (linear separation).To better understand the internal working mechanism of neural networks, consider the "simple" classification problem in Figure 6.1. In order to visualize the new representation of the data in the second hidden layer of the network, the input data is forward passed form the first to the second layer $x^{(2)} = \sigma^{(2)}(T_w^{(2)}(\sigma^{(1)}(T_w^{(1)}(\boldsymbol{x}))))$. (The architecture of the network is very primitive; hence, its complexity and capability is highly limited. It was chosen such that the transformed coordinate system can be visualized easily)



(a) Raw data

(b) Transformed data

Plot 6.2: Failed classification due to inopportune set of hyperparameters.

Figure 6.1 shows the learned classification boundary of the neural network (the training has been terminated after the classification accuracy reached 100 %). In configuration space $\boldsymbol{x}$, the data points are separated by a complex, non-linear decision boundary. In the transformed representation $\boldsymbol{x}'$, however, the data points are transformed in a complicated way such that they can be separated by a simple hyperplane. So, the network learns a new representation of the data (through a complicated coordinate transformation) in which the classification is simple.

The network may very well fail to learn an appropriate transformation that allows for a linear separation (e.g., due to the architecture that limits the complexity of the network, the amount of training data available, training steps or (most likely) an inopportune configuration of hyperparameters). Figure 6.2 gives an example of a failed attempt to learn a decision boundary due to a *purposely* ill-chosen setup. As shown in the this section, discriminative models (or feed-forward neural networks with softmax output) try to learn a complex transformation that allows for a linear separation of the transformed data in the representation of the penultimate layer. This property does not come as a surprise since it is imposed by the requirement alternating sequence of a linear (or affine) transformation T_w followed by a non-linear activation σ. One may think of a more general transformation than T that allows for a better transformation of the underlying manifold.

6.3 "Deep Sets" [3]

The kind of "simple" neural networks introduced in the previous section are limited to data with fixed dimensionality, i.e., the length of the input vector $x \in \mathcal{X}$, with $\mathcal{X}$ being the domain of the input data, to the machine learning model does *not* change between elements. This allows to approximate a vast space of function $f : \mathcal{X} \to \mathcal{Y}$ that transform the domain of input variables $\mathcal{X}$ into the image set $\mathcal{Y}$. The properties of the set of possible target values $\mathcal{Y}$ usually determines whether the machine learning algorithm is a classification or a regression task. In case of classification, the image set is discrete $\mathcal{Y} = \mathbb{Z}_{N_c}$ whereby the number of elements N_c corresponds to the number of classes or categories the model is supposed to discriminate ($N_c = 2$ for binary classification). A common example is binary classification, that is, the discrimination between just two classes or categories of data; in this case, the image set usually is set to be $\{0, 1\}$. In case of regression, on the other hand, the neural network approximates a function that maps the input variables to a continuous space of output values. A simple example of a regression task would be a function $f : \mathcal{X} \subset \mathbb{R}^d \to \mathcal{Y} \subset \mathbb{R}$ that maps a d-dimensional vector space onto the continuous space of real numbers.

This concept, however, can not be easily translated to set-valued data in the form of $\mathbb{X} = \{x_1, x_2, ..., x_M\}$ with $x_i \in \mathfrak{X}$ (notation has been adopted from Ref. [3]) where the input domain $\mathcal{X} = 2^{\mathfrak{X}} = \wp(\mathfrak{X})$ is the power set, i.e., the set of all subsets of $\mathfrak{X}$ including the empty set

with cardinality of zero. The cardinality of each set $x_i \in \mathbb{X}$ may vary from element to element. While feed-forward neural networks assume time-*independent* data, many of today's learning tasks, actually have to deal with sequential data that changes over time. A natural example would be predicting stock prices, analyzing DNA sequences or predicting protein folding, composing a piece of music or, to mention a more prominent and relevant example, (natural) language processing as it is needed for translation, speech recognition or text generation. In general terms, it can be said that feed-forward neural networks reach their limits in cases where the input data is varying in its length and highly "time"-correlated as they learn relationships between input features. To this end *Recurrent Neural Networks* (RNNs) have been invented in the '80s of the previous century (see Section 6.1). This kind of architecture not only allows to deal with variable-length inputs, but it also comes with in internal memory unit that allows to learn correlations between different states and inputs.

In case of discrimination or a multi-class classification task, the network needs to assign a probability to each label that maps it to a class with a certain probability. This adds the constrain $\sigma^{(L)} : \mathbb{R}^n \to [0,1]^{n_c}$ on the image set of the last layer, whereby n_c denotes the number of classes (in practice this restriction is enforced, e.g., by a *softmax* layer $\sigma(x)_j = e^{x_j} / \sum_{k=1}^{n_c} e^{x_k}$).

In principle RNNs can (and have) been used to analyze set-valued data. However, it was shown in Ref. [168] that the output of the function learned by an RNN highly depends on the ordering of the input data. For this reason, when working with RNNs, an arbitrary order must be imposed on the set and strictly hold during training and inference in order to obtain consistent results. This circumstance disregards the fundamental property of sets to be generally unordered, i.e., a randomly imposed permutation of the set's elements still corresponds to *exactly* the same object. Similarly, a proper function f defined on sets has to respect this essential characteristic feature and should therefore be unresponsive to the ordering of the elements of the input set. In mathematical terms: "a function $f : 2^{\mathfrak{X}} \to \mathcal{Y}$ acting on sets must be permutation invariant to the order of objects in the set, i.e, for any permutation $\pi : f(\{x_1, ..., x_M\}) = f(\{x_{\pi(1)}, ..., x_{\pi(M)}\})$" (Ref. 3, Property 1, p. 2).

The structure and properties of functions on sets have been extensively studied by Zaheer et al. in 2017 [3]; all information provided in this section is based their excellent work if not explicitly mentioned otherwise.

Since functions defined over a domain of sets are highly non-trivial, the authors start their considerations following a strategy of increasing generality and therefore constrain the class of permutation invariant functions to the case where $\mathfrak{X}$ is countable set (i.e., there exists an injective function that maps $\mathfrak{X}$ to the set of natural numbers $\mathbb{N}$) with the image being given by the set of real numbers $\mathrm{Y} = \mathbb{R}$. Based on those simplifications, the following Theorem characterizes the structure of the function f:

Theorem (Ref. 3, Thm. 2, p. 2)**.** *A function $f(X)$ operating on a set X having elements from a countable universe, is a valid set function, i.e., invariant to the permutation of instances in X, iff it can be decomposed in the form*

$$\rho\left(\sum_{x \in X} \phi(x)\right),$$

for suitable transformations ϕ and ρ.
(Proof: see Ref. 3, Appendix A, p. 12)

For a full proof of Theorem 6.3, the interested reader is referred to the quoted publication. Nonetheless, since the decomposition of the function f into ϕ and ρ is the basic motivation for the neural network architecture derived from it and used in this measurement, it is worth pausing for a moment and reflect on the statements made in the proposition above.

The quintessence of the Theorem is the observation that a permutation invariant function $f : \mathfrak{X} \to \mathcal{Y}$ defined over a set $\mathcal{X}$ with *countable* ($|\mathfrak{X}| < \aleph_0$) elements can be constructed through two continuous functions $\rho : \mathfrak{X} \to \mathbb{R}$ and $\phi : \mathbb{R} \to \mathbb{R}$ by the following rule $f(X) = \rho\left(\sum_{x \in X} \phi(x)\right)$. The important point is that in the decomposition of f, the functions ρ and ϕ do *not* operate on sets directly but on fixed-size objects for which the sum over all embeddings $\phi(x)$ corresponds to all elements of the set. In practice, the two functions ρ and ϕ are approximated by neural networks. The requirement of permutation invariance follows

trivially form the form of decomposition by noting that all finite sums are commutative, i.e., $\sum_{x \in X} \phi(x) = \sum_{x \in \pi(X)} \phi(x)$.

6.4 Training of Neural Networks

Section 6.2 introduced the basic concepts of neural networks regarding their mathematical formulation in terms of weights, activation and propagation functions, as well as different topologies in particular feed-forward (6.2) and "Deep Sets" (6.3). This section gives a short introduction to the actual training procedure, in particular, on how to update the model's parameters through gradient descent (6.4.2) and how to efficiently compute gradients (6.4.3). Furthermore, the elementary concept of a loss function in supervised and unsupervised learning is covered in this section.

6.4.1 The Loss Function

In order to quantitatively measure the inconsistency between the predicted values by the neural network $\hat{\boldsymbol{y}} = f_\theta(\boldsymbol{x})$ (now denoting the network's weights by $\boldsymbol{\theta}$) and the actual *label* $\boldsymbol{y}$ (available only in case of (semi-)supervised learning) for the training data $(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{D}$ one needs to define a *loss function* $\mathcal{L}(\hat{\boldsymbol{y}}, \boldsymbol{y})$. This loss function provides an estimate of the network's approximation error; hence, it should be defined to be a non-negative quantity ($\mathcal{L} \geq 0$), with a decreasing loss corresponding to reduced approximation error. Furthermore, the loss must be a *differentiable* function to get meaningful derivatives needed in the backpropagation and gradient descent algorithm to update the network's weights (the conceptually very different paradigm of reinforcement learning also allows for non-differentiable loss functions to be used).

There are plenty of different loss functions available, most of them serving a particular purpose, e.g., whether the objective is a regression or a discrimination task. A prominent example for a loss function in a simple regression task would be the *Mean Squared Error* (MSE), $\mathcal{L} = \frac{1}{N} \sum_{i=1}^{N} \left(\hat{y}^{(i)} - y^{(i)}\right)^2$, while in a classification task the usual choice would be the *binary cross-entropy* (multi-class cross-entropy for multi-class classification), $\mathcal{L} = -\frac{1}{n} \sum_{i=1}^{n} \left[y^{(i)} \log(\hat{y}^{(i)}) + (1 - y^{(i)}) \log(1 - \hat{y}^{(i)})\right]$.

At this point, there is nothing to be gained from providing a long enumeration of different loss functions and their properties. But one should keep the loss function and its role in the training process of neural networks in mind: the loss function is a measure of similarity between predictions of the neural network and the training data and hence is utterly important for the training process. It is the fundamental quantity to compute the corrections to the network's weights using gradient descent and backpropagation, which is subject to the next sections.

6.4.2 Gradient Descent and Optimizers

With the loss function at hand, the learning objective of neural networks is to find a configuration of weights/parameters in a very (very!) high-dimensional space, which defines the model, that minimizes $\mathcal{L}$. If $\mathcal{L}$ is non-convex, there is no guarantee that the optimization procedure will yield a *global* minimum of the loss function. Besides that, it is not expected that the minimization will result in a perfect configuration of weights since exact optimization is NP-hard.

Most optimization algorithms used in machine learning today are based on a *first-order* iterative optimization algorithm called *gradient descent*. The objective is to update the models parameters $\boldsymbol{\theta} \in \mathbb{R}^d$ (i.e., the weight) in the opposite direction of the gradient – which points in the direction of the "steepest slope" – of the cost function[4] $\nabla_\theta \mathcal{L}(\theta)$. The most simple form of such an algorithm requires one additional, adjustable parameter, i.e., the so-called learning rate $\alpha \in \mathbb{R}^{>0}$ that corresponds to the step size taken towards the (local) minimum. It is important to internalize that all optimization algorithms based on gradient descent are at *first-order*. While the gradient of the loss function only measures the change of the slope, the second-order derivative provides information about the direction based on the curvature of

[4]Through this thesis the terms *loss*, *cost* and *objective* function are used interchangeably.

the cost's error surfaces. So, the restriction to first-order gradients represents a significant limitation. An example of an optimization algorithm that also uses second-order derivatives would be Newton's method. However, computing second derivatives is "expensive" and often computationally intractable as for evaluating the Hessian matrix one needs to compute N_θ^2 derivatives, whereby N_θ is the number of weights in the network. The networks presented in this thesis have a total number of parameters in the order of approximately $N_\theta \sim \mathcal{O}(10^5)$. In other words, the second-order optimization algorithm would have to compute, roughly, one *billion* ($\mathcal{O}(10^{10})$) derivatives for only *one* iteration step! This is unfeasible for high-dimensional data sets; hence, the algorithms rely on first-order approximations.

Furthermore, a distinction is made between three different versions of gradient descent based on the information that is used to compute the gradients. *Batch gradient descent* computes the gradients for the entire data section and updates the parameters weighted by the constant and global learning rate

$$\boldsymbol{\theta} \rightarrow \boldsymbol{\theta} - \alpha \nabla_\theta \mathcal{L}(\theta). \tag{6.6}$$

This optimization method might be intractable for large data sets that does not fit into memory at once. The second variant is known by the name *stochastic gradient descent*. It is the other extreme and performs an update of the training parameters for each training example

$$\boldsymbol{\theta} \rightarrow \boldsymbol{\theta} - \alpha \nabla_\theta \mathcal{L}(\theta; x^{(i)}, y^{(i)}), \tag{6.7}$$

which might cause redundant computations and usually results in very volatile losses. Finally, *mini-batch gradient descent* reaches a compromise among the two previous approaches by performing an update of the model's weights for each *mini-batch* consisting of n_{mb} samples

$$\boldsymbol{\theta} \rightarrow \boldsymbol{\theta} - \alpha \nabla_\theta \mathcal{L}\left(\theta; x^{(i:i+n_{\text{mb}})}, y^{(i:i+n_{\text{mb}})}\right). \tag{6.8}$$

This one usually provides more stable convergence and hence is the method of choice – also in this thesis.

As has been described above, gradient descent only uses the information of first-order derivatives. However, some methods allow improving the estimation of the direction in the loss' error surface by taking into account the information of previous iterations. This is important since it has a profound impact on the performance of the optimization. There exists an unmanageable and somewhat confusing variety of extensions to the vanilla gradient descent that takes this approach; however, they can be broadly subdivided into methods that use *momentum* respectively *moving averages* and *adaptive learning rates*.

6.4.3 The Backpropagation Algorithm

Gradient descent provides rules to update the weights based on the gradient of a predefined error function. Backpropagation referrers to the numerical implementation of gradient descent that provides a method for performing *automatic differentiation* of complex, nested functions (such as multi-layer neural networks) by successive application of the chain and power rule.

The backpropagation algorithm consists of a *forward-pass* and a *backward-pass* (in this order). In the forward-pass step, the training data is propagated forward through the entire network that produces an output/prediction $\hat{\boldsymbol{y}}$. Now, the loss function $\mathcal{L}(\hat{\boldsymbol{y}}, \boldsymbol{y})$ is evaluated for the network's prediction (or a batch of predictions) and the actual label $\boldsymbol{y}$ of the data. (This procedure, of course, is only reasonable in the context of supervised learning where the *truth* information is available during training). The starting point of backpropagation is the very intuitive and natural definition of the "error" according to

$$\delta_k^{(l)} = \frac{\partial \mathcal{L}}{\partial z_k^{(l)}} = \frac{\partial \mathcal{L}}{\partial \sigma_k^{(l)}} \frac{\partial \sigma_k^{(l)}}{\partial z_k^{(l)}} = \sum_{i=1} \delta_i^{(l+1)} w_{ik}^{(l+1)} \frac{\partial \sigma}{\partial z_k^{(l)}}, \tag{6.9}$$

with $z_k^{(l)}$ referring to the output of the propagation function of the kth node/neuron in layer $l \in \{1, \ldots, L\}$. In Equation 6.9, the error has been (re-)express using the chain rule. Since

backpropagation starts from the output layer, the first step is to compute $\delta_k^{(L)}$ for neurons in this layer. In this case $\delta_k^{(L)}$, is simply given by $\delta_k^{(L)} = \frac{\partial \mathcal{L}}{\partial a_k^{(L)}} \sigma' \left(z_k^{(L)} \right)$, whereby $\sigma^{(L)}$ denotes the activation function in the output layer. Now, one needs to compute the gradients of $\mathcal{L}$ with respect to the weights $w_{kj}^{(l)}$. Using the chain rule again, this is given by

$$\frac{\partial \mathcal{L}}{\partial w_{kj}^{(l)}} = \frac{\partial \mathcal{L}}{\partial z_k^{(l)}} \frac{\partial z_k^{(l)}}{\partial w_{kj}^{(l)}} = \delta_k^{(l)} \frac{\partial z_k^{(l)}}{\partial w_{kj}^{(l)}} \tag{6.10}$$

The second multiplicand can be written as $\frac{\partial z_k^{(l)}}{\partial w_{kj}^{(l)}} = \sigma_j^{(l-1)}$ so that $\frac{\partial \mathcal{L}}{\partial w_{kj}^{(l)}} = \delta_k^{(l)} \sigma_j^{(l-1)}$. (Analogously, the very same can be done for the biases $\frac{\partial \mathcal{L}}{\partial b_k^{(l)}} = \delta_k^{(l)}$.) By subsequently applying the chain rule to Equation 6.10, which yields

$$\frac{\partial \mathcal{L}}{\partial w_{kl}^{(l)}} = \sum_{\alpha\beta\gamma\ldots\omega} \frac{\partial \mathcal{L}}{\partial z_\alpha^{(L)}} \frac{\partial z_\alpha^{(L)}}{\partial z_\beta^{(L-1)}} \frac{\partial z_\beta^{(L-1)}}{\partial z_\gamma^{(L-2)}} \cdot \ldots \cdot \frac{\partial z_\gamma^{(l+1)}}{\partial z_\omega^{(l)}} \sigma_j^{(l-1)}; \tag{6.11}$$

it now becomes clear why the algorithm is called backpropagation: with the chain rule, the "error" from the very first layer L is gradually backtracked to the respective layer l. The sum in Equation 6.11 is complicated and goes over all *connected* paths in the neural network.

With Equation 6.11 and gradient descent (see Section 6.4.2) at one's disposal, the weights of the neural network can be iteratively updated until the cost function is justifiable minimized and the model provides a reasonable approximation of the training data.

7. Data-Driven Feature Extraction

Hadronic jets are the phenomenological manifestations of the fundamental, *color-charged* particles of QCD measured an experiment; hence, they are a signature of the strong interaction. This association between a parton and a particle jet is by no means unproblematic as the jet is constructed from a spray of *color-neutral* hadrons and therefore, by construction, neutral in color by itself. The jet (see Chapter 1.3), however, is directly connected to a *color-charged* parton participating in the hard scattering process.

However, not only the association of a hadronic shower reconstructed in a detector and the originating particle that initialized it is conceptually problematic; a careful reflection on the everyday definition and notion of quark- and gluon-initialized jets quickly reveals that those objects are indeed "ill"-defined in most circumstances. As described in the introduction of this thesis, there is no "precise" definition quark and gluon jets to the present day. This problem arises because higher-order contributions to fundamentally different processes involving quarks and gluons in its final state may result in identical, i.e., indistinguishable configurations in the detector, which results in a superposition of states due to quantum mechanical indistinguishability. The reconstructed jets therefore can not uniquely be assigned to a quark nor gluon anymore. Furthermore, most definitions of quark- and gluon-initialized jets are theoretically ill-defined, i.e., they are not IRC-safe.

For instance, a very common definition of the flavor of a jet is based on the (partially arbitrary) event record (i.e., the set of particles produced by a Monte Carlo generator) by assigning the *Particle Data Group IDentification* (PDGID) number of the highest-p_{T} constituent clustered in the jet to the jet as a whole. This may seems intuitive because it is in harmony with the simple picture supported by the parton shower that models the jets as a high-energetic parton accompanied by consecutive emissions (see Chapter 2.4). Despite being commonly used, this definition of quark and gluon jets is *not* collinear safe because, for instance, the additional splitting of a gluon into two quarks $g \to q\bar{q}$ must result in the jet being identified as initialized by a quark, provided that the gluon was the highest-momentum particle in the jet. There are more examples to highlight inconsistencies in the definition of quark and gluon jets, however, the examples given so far should be sufficient to illustrate the general problem, before going into more detail. Generally, not having an IRC-safe definition of quark- and gluon-initialized jets, result in considerable complications regarding the comparison of analytically calculated features associated with quark and gluons and experimental measurements.

The lack of a well-defined notion of quarks and gluons at hadron level by means of their phenomenological signatures in the detector makes it very complicated to consistently study related properties thereof. The problem can be partially addressed by studying QCD-related properties in regions in phase space that are know to be highly enriched (or even pure) in quarks or gluons without the need of the precisely defining them. This is possible because the quark and the gluon as an element of the fundamental and adjoint representation of the SU(3) symmetry group gives rise to certain statistically distinctive features between both categories. This was, for instance, done at the Large Electron-Positron Collider (LEP), which resulted in highly enriched (almost pure) data of quark-initialized jets and therefore was perfectly suited to study quark-related properties. Gluon-enriched samples, on the other hand, are very limited (this would be a perfect argument for the construction of a Higgs-factory, which would result in very pure samples of gluons). At hadron colliders like the LHC, there are plenty of gluons produced, however, one always deals with *mixed* samples of quarks and gluons. This *degeneracy*

can not easily be broken as the fractions of quarks and gluons in the sample is unknown. The fractions (or, alternatively, templates) can be extracted from simulation; however, the obtained results are highly model dependent as using different Monte Carlo generators and different definitions of jets originating from quarks and gluons will result in different results. Furthermore, it is difficult to provide a well-defined uncertainty on the calculated results besides the simple comparison of various generators – which (admit it ...) does not makes any sense at all. Explaining a solely data-driven, model-independent method to break the aforementioned degeneracy of mixed samples of quark- and gluon-initialized jets and extracting their underlying fractions in different regions in phase space is subject of this chapter.

This chapter is structured as follows: The first section (7.1) provides a detailed introduction into the problem of consistently defining quark- and gluon-initialized jets at hadron level. Besides common pitfalls in their definition, this section also provides an overview of the most common definitions of quark and gluon jets. A possible way out of the misery is provided in the subsequent section (7.2) that introduces an operational definition of quark- and gluon-*like* (the adjective use is important as will become clear in a moment) at distribution level, avoiding aforementioned difficulties with jet-by-jet-based labeling schemes. This *operational* definition is independent of model-related information; however, in order to break the degeneracy of samples from hadron-hadron collisions, the underlying JSS observable that is used to measure the cross section must result in phase space regions that are almost pure in either quark- or gluon-initialized jets. Mathematically, this means that the underlying probability distributions of the observable must be *mutual irreducible*. The rather "abstract concept" of mutual irreducibility is studied in light of QCD in the next section (7.3). After that (A), the method for extracting the underlying quark/gluon probability distributions as well as their fractions is connected to a statistical demixing procedure called "*topic modeling*", which has originally been developed as a tool for extracting emergent themes in an extensive corpus of text-based documents. Afterwards (7.4), the method is extended to non-mutual irreducible observables by means of a simultaneous measurement with a mutual irreducible one, or, by making use of the *universality* of the extracted fractions. Finally (7.5), a step-by-step introduction on how to extract the underlying quark/gluon features and fractions from an unfolded multi-differential cross section is provided.

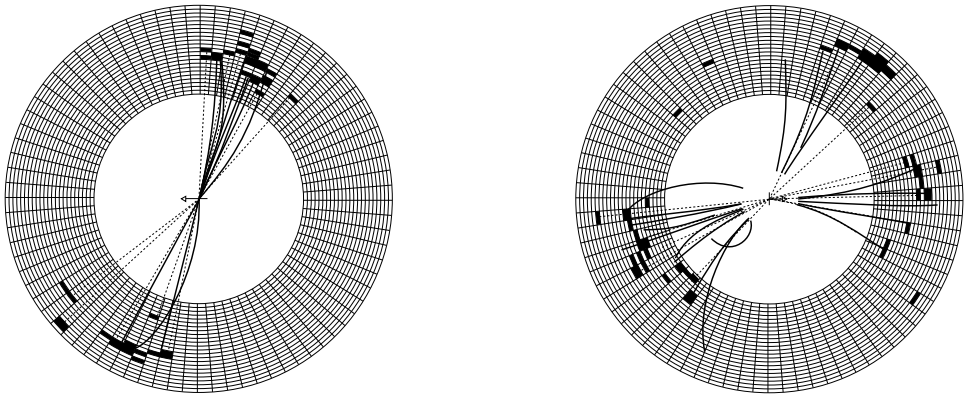
7.1 Definition of Quark and Gluon Jets

"Quarks and gluons are the fundamental quanta (quantum fields) of QCD (see Chapter 1.1.3) that are associated with two different representations of the underlying non-abelian SU(3) symmetry group. In the Lagrangian of the theory (see Equation 1.6), the gluon field is given by a tensor field of rank 2 defined over spacetime with its values being elements of the adjoint (vector) bundle of the aforementioned symmetry group. The quark on the other hand is given by the fundamental representation of SU(3)." – this (or something similar) might be the answer to the question: "what are quarks and gluons?" if asked a theorist[1]. Certainly a good and valid answer ..., but maybe not the one needed by an experimentalist or instrumentalist for his daily working routine. From an experimentalist's point of view, the more important question is how the fundamental particles associated with the strong force manifest itself in an experiment (i.e., their "phenomenology") and how consistently define them. The answer to this questions is harder than it may seems at first glance.

7.1.1 Problems Defining Quark and Gluon Jets

As a first example of quark- and gluon-initialized jets, take the two event displays recorded with the JADE experiment at LEP in Figure 7.1 that shows a two- and a three jet event. At LEP(II), electrons and positrons are collided at a centre-of-mass energy close to the resonance mass of the Z boson. The two-jet event in Figure 7.1a therefore most likely resembles the

[1] This statement is supported by the author's very own experience during lunch at R1 at CERN.



(a) (b)

Fig. 7.1: Event display of a two-jet (*left*) and a three-jet (*right*) event, recorded with the JADE detector at the Positron-Electron Tandem Ring Accelerator (PETRA) at DESY in Hamburg, Germany. (Taken from Ref. 22, Fig. 9.6 p. 132).

process $e^+e^- \to Z \to q\bar{q}$ with two quark-initialized jets in the final state. Since the two-jet cross section for e^+e^--collisions at this energy is very large, LEP data is highly enriched with quarks and therefore perfectly suited to study quark-related properties. Basically, the requirement of two jets (exclusively) in the final state defines a selection that results in *almost* pure samples of quark-initialized jets.

In case of the gluon, the situation is different though. For a gluon-initialized jet to be present in the final state, the process must be (in the language of perturbation theory) of higher order (at least NLO). The lowest process that (probably) includes a gluon in the final state is a three-jet event as the one shown in Figure 7.1b. The additional gluon usually is a result of the accelerated color-charged quark emitting radiation similar to an electron emitting electromagnetic radiation if accelerated. This higher-order process introduces an additional factor of α_S. In consequence, the number of gluon-initialized jets is suppressed by at least an order of magnitude ($\alpha_S(M_Z) \approx 1/10$) compared to jets due to quarks; as a result, the former ones are much less constrained from data, which results in larger uncertainties associated to gluon-related features. Furthermore, for a three-jet event, the selection of the participating gluon-initialized jet is not fully efficient because in some cases the radiated gluon can not be distinguished from the other quark jets with high efficiency. Most of the time, however, it is sufficient to associate the gluon with the energetically softest jet participating in the event, eventually combined with a mass-window cut close to the mass of the Z boson. In the example above, comparatively straightforward to consistently define quark and gluon jets at hadron level based on the final state. In case of hadron colliders, like the LHC (see Chapter 3.1), however, the situation changes dramatically as it (almost) becomes impossible to isolate pure regions of quark- and gluon-initialized jets in phase space, while at the same time maintaining sufficient statistics. Generally, at hadron colliders, one *always* deals with mixed samples of quarks and gluons. As a further complication, the notion of quark and gluon jets becomes theoretically ill-defined at parton level beyond Born level.

As an illustrative example, consider the two different second-order processes shown in Figure 7.2 that both share the same initial and final state, i.e., they are experimentally indistinguishable and irreducible. In the first processes (see Figure 7.2a), a quark-antiquark pair couples to the gluon mediator with one of the quarks emitting a gluon because it is an accelerated (color) charge. The three partons in the final state would give rise to three jets (provided that the gluon is resolved) in the detector one of which being gluon initialized, while the other two are caused by quarks. The situation, however, is more involved as other processes with the *same* order in α_S and the *same* final state configuration contribute too. An example is shown in Figure 7.2 that represents the same initial state but with a tree-gluon vertex instead. If one of the two gluons splits into a quark-antiquark pair (see Chapter 2.4

Fig. 7.2: Example of two processes evolve from the same initial state involving the emission of a gluon (*left*) and gluon splitting into two quarks (*right*) resulting in the same final-state configuration.

concerning parton showers), the particle content of the final state will be identical and therefore quantum mechanically indistinguishable with respect to process depicted in Figure 7.2b. Since the two processes are indistinguishable regarding their final states, a super position of the two contributions must be considered according to the laws of quantum mechanics, which may give rise to highly non-trivial interference effects. Now, however, a "peculiarity" is revealed: both processes in Figure 7.2 result in two quarks and one gluon being present in the final state; but, on the underlying LO process (the emission of a gluon or the splitting of a gluon into a quark-antiquark pair are considered higher-order corrections to the Born contribution) involve two quarks in one case but two gluons in the other one. Since the two processes interfere with each other, Figure 7.2 represents and example for the notion of quark and gluon jets becomes "ill"-defined already at parton level.

This illustrative example only serves as a quick motivation for potential problems when speaking about quarks and gluons, which are usually not considered in general linguistic usage. The upcoming section provides a detailed overview of different approaches to define quark and gluon jets based on the Les Houches workshop on "Physics at TeV Colliders" that took place in 2015 [23].

7.1.2 "What is a Quark/Gluon Jet?"

As it was shown in the previous section, the seemingly simple question of how to define quark and gluon jets is much more subtle than expected at first glance. First of all, it should be noted that *there is no unambiguous definition of quark and gluon jets at hadron level.* This is due to the aforementioned interference effects in quantum mechanics that does not allow for a one-to-one correspondence between the partons involved in the internal process and their phenomenological manifestations, i.e., the jets, in the final state[2]. Furthermore, the showering as well as the non-perturbative formation of stable hadrons add another layer of complexity to the labeling of jets in terms of the two categories (light) quark or gluon. This circumstance of not having a well-defined description of quark and gluon jets represents a problem as many analyses rely on this information and therefore depend on the (partially arbitrary) information in how the flavor of a jet was defined in this particular case, which also makes the comparison of different studies with non-identical jet-labeling schemes somewhat inconsistent.

For instance, as shown in Chapter 4.4.3, the Global Sequential Calibration component of the ATLAS calibration scheme of the jet energy scale aims to account for the different calorimeter responses of quark- and gluon-initialized jets due to differences in their respective shower shapes and response of the calorimeter. This correction is solely based on Monte Carlo; hence, the concrete result, i.e., the associated correction factor (see Equation 4.9), depends on the definition used to label jets being either quark- or gluon-initialized. Consequently, the

[2]In fact, this statement is perfectly consistent with QM in that sense that is not possible to make precise statements concerning the outcome/result of a particular measurement. Instead, one can only make statements about an *ensemble* of measurements.

flavor-related uncertainty in the calorimeter response (which is given by the difference between two Monte Carlo event generators[3] (see Equation 4.14)) depends on the particular convention of how quark and gluon jets are defined.

The following paragraphs are almost a verbatim representation of Section 2 of the excellent publication *"Systematics of quark/gluon tagging"* [169]. Here, the different approaches to define quark and gluon jets are grouped into three categories based on the underlying common "theme" or "principles". The order of suggested options for defining the flavor of a jet follow a path of increasing "well-definiteness" of the respective definition, starting from most ill-defined interpretation to the most well-defined statement of meaning.

The first group of definitions *"[...] try to identify [.] quark[s] [and gluons] as [..] object[s] unto itself, without reference to the specific final state of interest."* This approach assigns a flavor *jet-on-jet based*, i.e., each jet in an event is (binary) labelled as a quark (of some flavor) or a gluon based on some identification or classification scheme completely independent of the actual underlying process. Using those methods hence allow to create completely labelled data sets, which are, for instance, the prerequisite to train machine learning models in a supervised fashion. This approach, however, suffers from the aforementioned problem of the superposition of different processes involving different partons that yield an identical final state. Imposing an unambiguous label beyond leading-order calculations for each jet consequently does not take this intrinsic *in*definiteness existing on jet-on-jet level into account. Besides the effect of interference of final state, methods in this class of labeling schemes do not account for the circumstance of jets being fundamentally process dependent in general due to, e.g., non-trivial color connections etc. This phenomenon is refereed to as *sample dependence*, which simply defines the well-established fact that a quark or a gluon may differ between different processes. For instance, a gluon jet originating from a (pure) QCD/multijet event (i.e., only the strong force being involved) differs from a gluon-associated jet created in an, e.g., Z + jet event. This aspect will be of relevance again later on for the motivation of this measurement being done in dijet events solely to not encounter (or rather reduce) the problem of *sample-dependent* jets, which spoils one of the basic assumptions of the algorithm to extract the underlying quark/gluon distributions introduced in the next section.

Following now, a quick summary of labeling schemes that fall into this category according to the conclusions of the already mentioned Les Houches workshop on *"Physics at TeV Colliders"* in 2015.

- *"A quark parton. This definition (incorrectly) assumes that there is a one-to-one map between a jet and its initiating parton. Because it neglects the important role of additional radiation in determining the structure of a jet, we immediately dismiss this definition."*

- *"A Born-level quark parton. This definition at least acknowledges the importance of radiative corrections to jet production, but it leaves open the question of how exactly to define the underlying Born-level process from an observed final state. (For one answer valid at the parton level, see flavored jet algorithms below.)"*

- *"An initiating quark parton in a final state parton shower. We suspect that this is the definition most LHC experimentalists have in mind. This definition assumes that the parton-shower history is meaningful, though, which may not be the case beyond the strongly-ordered or LL approximations. Because the parton shower is semi-classical, this definition neglects the impact of genuinely quantum radiative corrections as well as non-perturbative hadronization."*

- *"A maximum-p_{T} quark parton within a jet in a final state parton shower. This definition uses the hardest parton within the active jet area encountered at any stage of the*

[3] It is not obvious by any means that some "arbitrary" definition of quark and gluon jets are equally well suited for any event generator. On the contrary, many jet-labeling schemes that are used in practice have been optimized based on a particular model (e.g., Pythia), which may introduce inconsistencies when applying the very same method to other generators.

shower evolution, including the initial hard scattering process. This 'max-p_{T}' the shower evolution, including the initial hard scattering process. This 'max-p_{T}' prescription is a variant on the initiating parton prescription above (see further discussion in ref. [55][4]). It differs from the initiating parton by a calculable amount in a LL shower [36] and is based on the same (naive) assumption that the parton-shower history is meaningful."[5]

- *"An eikonal [(from Greek εἰκών meaning "image")] line with baryon number 1/3 and carrying triplet color charge. This is another semi-classical definition that attempts to use a well-defined limit of QCD to define quarks in terms of light-like Wilson lines. Philosophically, this is similar to the parton-shower picture, with a similar concern about how to extrapolate this definition away from the strict eikonal limit."*

- *"A parton-level jet object that has been quark-tagged using an IRC-safe flavored jet algorithm. This is the strategy adopted in Ref. [56]. While this definition neglects the impact of hadronization, it does allow for the calculation of quark jet cross sections at all perturbative orders, including quantum corrections."*

The upcoming definition of quark and gluon jets changes the perspective and tries to provide a description based on the respective process. In a nutshell, the method utilizes a feature of certain processes that allow (within certain limits in phase space) a factorization of short-distance from the long-distance, non-perturbative fragmentation, which allows for a definition of quark and gluons in terms of observables based on jet grooming.

- *"A quark operator appearing in a hard matrix element in the context of a factorization theorem.This is similar to the attitude taken in ref. [54]. In the context of a well-defined cross section measurement, one can (sometimes) go to a limit in phase space where the hard production of short-distance quarks and gluons factorizes from the subsequent long-distance fragmentation. This yields a nice (gauge-covariant) operator definition of a quark jet, which can be made precise for observables based on jet grooming [57, 58]. That said, even if a factorization theorem does exist for the measurement of interest, this definition is potentially ambiguous beyond leading power."*

Last but not least, the next definition (which is relevant for the studies presented in this thesis) aims to identify quarks and gluons in terms of the phase space given by the hadronic cross section of some process, i.e., there is no aspiration to attribute the flavor of the jet to some object at parton level. It is important to appreciate that this approach does treat quarks and gluons as an *adjective* rather than a substantive – as it was the case for most of the other definitions – because it assigns a "likeness" ("quark-likeness" and "gluon-likeness") to each jet according to its location in phase space instead of forcing a "hard"(=binary) label on a jet-on-jet basis. Of course, in order for this approach to provide reasonable results, phase space regions that are sufficiently (a mathematically exact criterion will be provided in the

[4] The following references correspond to the actual bibliography entries present in Ref. [169].

[36] M. Dasgupta, F. Dreyer, G.P. Salam and G. Soyez, *Small-radius jets to all orders in QCD*, JHEP 04 (2015) 039 [arXiv:1411.5182]

[54] J. Gallicchio and M.D. Schwartz, *Pure Samples of Quark and Gluon Jets at the LHC*, JHEP 10 (2011) 103 [arXiv:1104.1175]

[55] A. Buckley and C. Pollard, *QCD-aware partonic jet clustering for truth-jet flavor labeling, Eur. Phys. J. C 76 (2016) 71 [arXiv:1507.00508]*

[56] A. Banfi, G.P. Salam and G. Zanderighi, *Infrared safe definition of jet flavor*, Eur. Phys. J. C 47 (2006) 113 [hep-ph/0601139]

[57] C. Frye, A.J. Larkoski, M.DṠchwartz and K. Yan, *Precision physics with pile-up substructure beyond the next-to-leading logarithm*, JHEP 07 (2016) 064 [arXiv:1603.09338]

[5] The above-described method that determines the flavor of the jet based on the PDG particle number of the highest-p_{T} constituent inside of a jet is commonly used by many analysis at the ATLAS experiment. It provides the advantage of being infrared safe (by construction) because the additional emission of a soft parton does not change the flavor of the jet abruptly. However, it is *not* collinear safe as, for instance, the additional splitting of a gluon into two quarks $g \to q\bar{q}$ would result in the jet being identified as initialized by a quark.

upcoming sections) enriched by either quark or gluon jets must exist in order to extract enough information of the respective jet flavor. The last part of the statement appears like a circular argument: to get a well-defined definition of quark and gluon jets in terms of a hadronic cross section measurement, regions in phase space that are "adequately" pure in either of those two categories must exist; however, the definition of those enriched regions in phase space are usually determined based on an intuitive and (possibly) once again "ill"-defined understanding of the nature of quarks and gluons. This problem is unavoidable to a certain degree. But, given a minimal set of assumptions regarding our understanding of quarks and gluons as the fundamental quanta of QCD as well as an appropriate class of observables, the method at hand still yields a well-defined, meaningful definition of quark- and gluon-like jets (at distribution level) while avoiding most of the issues attributed to the other schemes mentioned above.

- *"A phase space region (as defined by an unambiguous hadronic fiducial cross section measurement) that yields an enriched sample of quarks (as interpreted by some suitable, though fundamentally ambiguous, criterion). Here, the goal is to tag a phase space region as being quark-like, rather than try to determine a truth definition of a quark. This definition has the advantage of being explicitly tied to hadronic final states and to the discriminant variables of interest. The main challenge with this definition is how to determine the criterion that corresponds to successful quark enrichment. For that, we have to rely to some degree on the other less well-defined notions of what a quark jet is."*

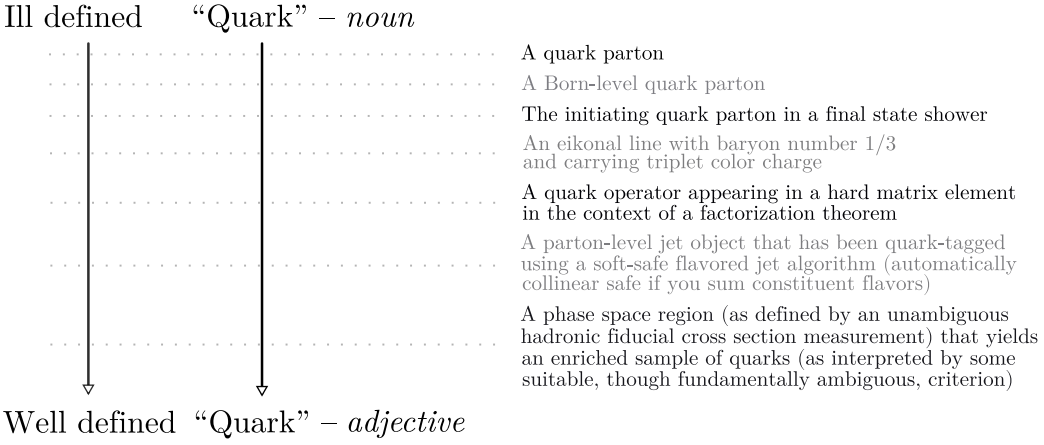


Fig. 7.3: Summary of the different approaches to define quark and gluon jets from ill- to well-defined interpretations based on the original slide from the June 10, 2015 summary report of the quark/gluon Les Houches subgroup Ref. [23] shown in Figure (Fig.) IV.46.

The last definition deserves a little more attention because it forms the basis for the interpretation of quark and gluon jets studied this analysis. As it was the case before, the major part of the content of this paragraph was adopted from Ref. [23].

The idea is to *unambiguously* measure the hadronic cross section of some process of interest (which results in a mixed sample of quark- and gluon-initialized jets) with respect to some jet substructure observable $\mathcal{O}_{\text{JSS}}$ that acts as a quark-gluon discriminant with a certain efficiency. In general, the resulting differential cross section $\mathrm{d}^{n+1}\sigma/\mathrm{d}\mathcal{O}_{\text{JSS}}\mathrm{d}^n\mathcal{V}$, whereby $\mathrm{d}\mathcal{O}_{\text{JSS}}\mathrm{d}^n\mathcal{V}$ denotes the total phase space volume, does *not* correspond to the probability of the jet being a quark or a gluon because the samples are mixtures of both as mentioned already. The probability (distribution) of a jet J with an observed JSS value $\mathcal{O}^J_{\text{JSS}}$ being initialized by a quark or a gluon $P_{q/g}(\mathcal{O}^J_{\text{JSS}})$ can *not* be determined based on the cross section measurement of a single sample of quark and gluon jets alone. To provide enough information to extract the underlying target distributions, the aforementioned (differential) cross section measurement must be supplemented by a second measurement of the *same object* but based on a sample consisting

of a *different mixture* of quarks and gluons.

Regarding the substructure observable $\mathcal{O}^{J}_{\mathrm{JSS}}$, it should be possible to define regions in phase space that are enriched (not necessarily pure!) by either quarks or gluons. Depending on the exact form of $\mathcal{O}_{\mathrm{JSS}}$, but without loss of generality, one can construct the quark-gluon discriminant such that in the limit of its support, $\mathrm{supp}\,(\mathcal{O}_{\mathrm{JSS}})$, the distribution is dominated (ideally pure) in one of the two categories, e.g.,

$$\mathcal{O}_{\mathrm{JSS}} \xrightarrow{\text{“quark}-\text{like”}} \min \mathrm{supp}\,(\mathcal{O}_{\mathrm{JSS}})\,, \tag{7.1}$$

$$\mathcal{O}_{\mathrm{JSS}} \xrightarrow{\text{“gluon}-\text{like”}} \max \mathrm{supp}\,(\mathcal{O}_{\mathrm{JSS}})\,. \tag{7.2}$$

Of course, this interpretation assumes that there is a rough understanding of how quarks and gluons (as the fundamental particles of QCD) manifest itself in an experiment, which leads back to the (supposedly) circular argument mentioned above. However, the statements according to Equation 7.1 do not require a precise definition of the “true” nature of quarks and gluons, but just a conjecture regarding their signature in an experiment. For instance, quarks and gluons correspond to the fundamental and adjoint representation, respectively, of the SU(3) symmetry group, which gives rise to different eigenvalues of the Casimir operator that can be considered as different color charges assigned to quarks and gluons. This simple fact already implies different experimental signatures expected for quark- and gluon-initialized jets. The number of constituents n_{const} inside a jet, for instance, is proportional to the number of emissions of an accelerated color charge. Because the color factor of gluons is greater than the color factor of quarks, gluon jets (independently of the precise definition of this object!) come with more emissions *on average* and hence more constituents, i.e., $\mathbb{E}_{n_{\mathrm{const}}\sim\mathbb{P}_g}\left[n_{\mathrm{const}}\right] > \mathbb{E}_{n_{\mathrm{const}}\sim\mathbb{P}_q}\left[n_{\mathrm{const}}\right]$. Consequently, due to energy-momentum conservation, one also expects jets to be broader on average if initialized by a gluon because the same energy budget (which is on average identical in case of quarks and gluons) is distributed over a larger number of particles. Those are just two examples of a large amount of possible quark-gluon discriminators available that do not require a precise definition of those objects. The quintessence of this paragraph is that there is no need for a precise definition of quarks and gluons in order to define regions in phase space enriched by jets initialized by the fundamental quanta of QCD.

Finally – to emphasize again –, the essence of this definition is to define quarks and gluons in terms of the hadron-level cross section with respect to a discriminant variable that is sensitive to the different representations of the QCD symmetry group. This means in particular that there is no effort to assign a “hard“ (i.e., “must be either a (light) quark or a gluon”) label to each jet in an event. *The actual interpretation of the measurement in terms of quark-like and gluon-like features is done on distribution level rather than on individual jets that are present in an event.*

7.2 An Operational Definition of Quark and Gluon Jets

The previous section introduced several possibilities of defining quark and gluon jets from ill- to well-defined concepts. One insight was that in order to obtain a well-defined definition, the notion of classifying jets on a jet-on-jet basis must be abandoned in favor of an approach that interprets the result in terms of quark- and gluon-likes features with regard to the hadronic cross section, i.e., at the level of individual distributions, making it consistent with quantum mechanics.

This section directly connects to the foregoing one by introducing the actual definition of quark and gluon jets that thenceforth will be used in this analysis; it also provides the theoretical foundation for the measurement of quark and gluon fractions.

The path has already been – in broad terms – mapped out; the starting point of the discussion leading towards the *operational definition* of quark and gluon jets is given by the *conceptional* one, which – for the sake of completeness and lucidity – is repeated once more.

Definition 1 (**Conceptual Definition of Quarks/Gluon Jets [23, 42]**). *A phase space region (as defined by an unambiguous hadronic fiducial cross section measurement) that yields an*

enriched sample of quarks (as interpreted by some suitable, though fundamentally ambiguous criterion)

The advantages and disadvantages of Definition 2 was already discussed at length in Section 7.1.2. The aim now is to convert this conceptional definition ("conceptional" since it does not provide clear instructions on how to apply this picture of quarks and gluons in practice) into an *operational*[6] recipe, which can be used in an actual analysis.

To arrive at an operational definition, this paragraph starts by reflecting on the basic questions on how a final cross section measurement consists of contributions from (light) quarks and gluons.

Consider two samples of hadronic jets (i.e., a proper *set* of jets that provide the basis for the (multi-)differential cross section measurement) given by

$$M_i^k := \left\{ J \in \mathbb{R}^{(1,3)} \mid \Phi_i(J) \wedge p_{\mathrm{T}}^J \in [p_{\mathrm{T}}^{\mathrm{min,k}}, p_{\mathrm{T}}^{\mathrm{max,k}}] \right\} \subset \mathbb{R}^{(1,3)}, \tag{7.3}$$

whereby Φ_i with $i \in \{1, 2\}$ denotes a (logical) *predicate* that may correspond to two processes or two different selections with the same process and $k \in \mathbb{N}$ denoting the kth (histogram) bin of transverse momentum such that $M_i = \bigcup_{k\in\mathbb{N}} M_i^k$ gives the full ith mixture of the measurement[7]. Furthermore, the mixtures M_i^k are mutually disjoint, meaning, $M_i^k \cap M_j^k = \varnothing$ if $i \neq j$. The two samples M_1 and M_2 correspond to two *mixtures* of quark and gluon jets[8], meaning that the fractions of quark- and gluon-initialized jets in the two samples $f_{q/g}^{M_{1/2},k}$ ($f_{q/g}^{M_1,k} \neq f_{q/g}^{M_2,k}$) are different.

The basic assumption of the operational definition is that the cross section $P_{M_i}(\mathcal{O}_{\mathrm{JSS}}) := \int \mathrm{d}\mathcal{V}_i \mathrm{d}^{n+1}\sigma / \mathrm{d}^n\mathcal{V}_i \mathrm{d}\mathcal{O}_{\mathrm{JSS}}$ that has been reconstructed based on the ith sample of jets is a mixture of contributions from quark and gluon jets according to

$$P_{M_i}^k(\mathcal{O}_{\mathrm{JSS}}) = f_q^{M_i^k} P_q^k(\mathcal{O}_{\mathrm{JSS}}) + f_g^{M_i^k} P_g^k(\mathcal{O}_{\mathrm{JSS}}), \tag{7.4}$$

with $P_{q/g}^k(\mathcal{O}_{\mathrm{JSS}})$ being the distributions related to quarks and gluons in the kth transverse momentum bin over the JSS variable $\mathcal{O}_{\mathrm{JSS}}$ that are supposed to be extracted. By assumption, there are only two categories: (light) quarks and gluons; hence, $f_q^{M_i} + f_g^{M_i} = 1$.

This decomposition seems very plausible and reasonable, but it is not unproblematic not to mention unreserved valid. One problem is that this assumption ignores the effect of the already mentioned sample dependence which states that quark and gluon jets are *not* universal, i.e., quark/gluon jets from different processes may differ fundamentally. In the context of this study, this problem is mostly avoided by solely dealing with jets of the same process (multijet event production).

The objective of the operational definition is to extract the quark/gluon fractions $f_{q/g}^{M_i^k}$ as well as the underlying distributions $P_{q/g}^k(\mathcal{O}_{\mathrm{JSS}})$. According to Equation 7.4, this problems appears to be ill defined as the resulting linear system of (two) equations is underconstrained (four unknowns). Usually, the additional information that is needed to solve Equation 7.4 is inferred from theory or simulation by estimating the quark/gluon fractions or by extracting templates from Monte Carlo (e.g., Pythia). In the operational definition of quark and gluon jets, however, no information from simulation is supposed to be used. This problem is addressed by an additional constrain that must be fulfilled by the observable $\mathcal{O}_{\mathrm{JSS}}$, which limits the space of possible JSS variables that can be used in combination with the operational definition of

[6] According to W. Edwards Deming's book "The New Economics: For Industry, Government, Education" *"An operational definition is a procedure agreed upon for translation of a concept into measurement of some kind[.]"* [170].

[7] Don't be confused; this is just a smart ass' way of saying that there are two different processes/mixtures of jets (like, for instance, M_{QCD} and $M_{\mathrm{Z+jet}}$) or, like it is done in this analysis, two sets defined by a selection with different efficiencies in quark and gluon jets (like, for instace, M_{central} and M_{forward}). We just need two mixtures of jets that differ wrt. their composition of quarks and gluons.

[8] In the scope of this analysis, the two mixtures *always* refer to different mixtures of quark and gluon jets. However, the concept can easily be adopted to other processes that fulfill the basic assumptions of the method outlined in this section.

quark and gluon jets.

To gain an understanding of the additional property of $\mathcal{O}_{\text{JSS}}$, consider the classification power between quarks and gluons provided by the two distributions $P^k_{q/g}(\mathcal{O}_{\text{JSS}})$. According to the *Neyman-Pearson lemma*, the *optimal* classifier statistics for discriminating two categories is given by their (log-)likelihood ratio $L_{q/g}(\mathcal{O}_{\text{JSS}}) := \log\left(P^k_q(\mathcal{O}_{\text{JSS}})/P^k_g(\mathcal{O}_{\text{JSS}})\right)$ (the logarithm just represents a monotonic transformation and will be ignored in the following). A corresponding likelihood ratio can be constructed for the two different mixtures $L_{M_1/M_2}(\mathcal{O}_{\text{JSS}}) := \log\left(P^k_{M_1}(\mathcal{O}_{\text{JSS}})/P^k_{M_2}(\mathcal{O}_{\text{JSS}})\right)$ that gives the optimal classifier statistic between M_1 and M_2 (those objects can actually be measured!). Equation 7.4 (the ratio of $P^k_{M_1}$ and $P^k_{M_2}$) can now be expressed in terms of $L_{q/g}$ and L_{M_1/M_2}, which yields the following expression as easily verified

$$L^k_{M_1/M_2}(\mathcal{O}_{\text{JSS}}) = \frac{P^k_{M_1}(\mathcal{O}_{\text{JSS}})}{P^k_{M_2}(\mathcal{O}_{\text{JSS}})} = \frac{f_q^{M_1^k} L^k_{q/g}(\mathcal{O}_{\text{JSS}}) + f_g^{M_1^k}}{f_q^{M_2^k} L^k_{q/g}(\mathcal{O}_{\text{JSS}}) + f_g^{M_2^k}}, \tag{7.5}$$

whereby, once again, the likelihood ratios are determined for each p_{T}-bin for which the quark/gluon fractions and the underlying distributions are supposed to be extracted. The derivative of Equation 7.5 with respect to $L^k_{q/g}$ is given by

$$\left|\partial_{L^k_{q/g}} L^k_{M_1/M_2}\right| = \frac{f_g^{M_1^k} f_q^{M_2^k} - f_g^{M_1^k} f_g^{M_2^k}}{\left(f_q^{M_2^k} L^k_{q/g} + f_g^{M_2^k}\right)^2}, \tag{7.6}$$

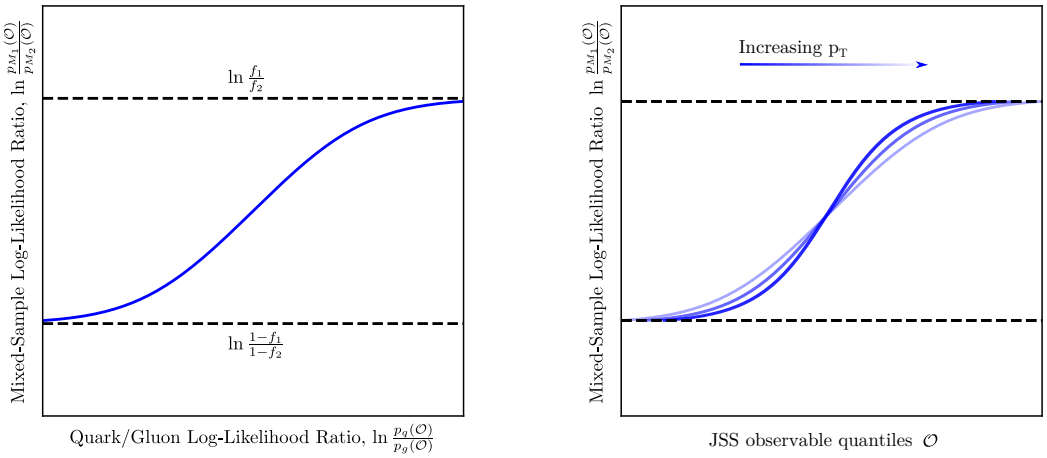
which implies that the $L^k_{M_1/M_2}$ is a monotonic function of $L^k_{q/g}$ provided that the fractions are different $f_g^{M_1^k} \neq f_g^{M_2^k}$. Here it can already been seen that it is very important for the two samples of jets being different mixtures (with different fractions) of quark and gluon jets. The monotony plays a crucial role which once again highlights the importance of the two samples M_1 and M_2 to differ regarding their fraction of quark- and gluon-like jets. The conceptual graph of $(L^k_{M_1/M_2}, L^k_{q/g})$ is shown in Figure 7.1, which nicely illustrates the monotonous course. Figure 7.1 also shows the limits of Equation 7.6 that only depend on the fractions in the different mixtures, which opens a window to break the degeneracy present in Equation 7.4

$$\lim_{L^k_{q/g}\to\max} L^k_{M_1/M_2}\left(L^k_{q/g}\right) = \frac{f_q^{M_1^k}}{f_q^{M_2^k}}, \tag{7.7}$$

$$\lim_{L^k_{q/g}\to\min} L^k_{M_1/M_2}\left(L^k_{q/g}\right) = \frac{f_g^{M_1^k}}{f_g^{M_2^k}}. \tag{7.8}$$

This is an interesting first result. Provided that the likelihood ratio of the two mixtures is known (or at least approximated sufficiently well), the entire substructure feature space was mapped to a single number that can be used to determine an optimal (according to the Neyman-Pearson lemma) boundary in phase space to identify regions enriched in quark- and gluon-initialized jets. So far, however, this criterion only exists hypothetically because there is now procedure yet that describes a method to identify the enriched regions besides the fact that they are given by some selection cut on $L^k_{q/g}$ *or* equivalently(!) $L^k_{M_1/M_2}$. The connection to the targeted operational definition of quark and gluon jets lies in the insight of $L^k_{q/g}$ being an optimal classifier that can be used to determine regions in phase space that are (almost) pure in either quarks or gluons given by the extreme points of $L^k_{q/g}$. Provided that the jet substructure observable $\mathcal{O}_{\text{JSS}}$ results in regions in phase space $U^k_q(\mathcal{O}_{\text{JSS}}), U^k_g(\mathcal{O}_{\text{JSS}}) \subset \text{supp}(P^k(\mathcal{O}_{\text{JSS}}))$ with $U^k_q \cap U^k_g = \varnothing$ that are pure in quarks and gluons, which implies $\text{supp}(P^k_{q/g}(U^k_{g/g})) = \varnothing$ and $\text{supp}(P^k_{q/g}(U^k_{g/q})) \neq \varnothing$. In those pure regions of phase space, the likelihood *must* vanish $L^k_{q/g}(U^k_g) = 0$ and $L^k_{q/g}(U^k_q) = 0$, which is the equivalent of the statement that the jet substructure observable $\mathcal{O}_{\text{JSS}}$ is sufficiently powerful to result in probability distributions $P^k_{q/g}(\mathcal{O}_{\text{JSS}})$ that are *mutually irreducible*.

Because the concept of mutually irreducibility is absolutely crucial, the following theorem



Plot 7.1: "The monotonic relationship between the mixed-sample log-likelihood ratio $[L^k_{M_1/M_2}]$ and the quark-gluon log-likelihood ratio $[L^k_{q/g}]$ [...] for illustrative fraction values [(left) and the mixed-sample log-likelihood ratio as a function of the observable quantiles (right)]. The relationship between the maximum and minimum values of the mixed-sample and quark/gluon log- likelihoods [...] is visually evident in that the red curve horizontally asymptotes to the two black dashed curves. [...]" (adapted and modified from Ref. 42, Fig. 1, p. 6).

provides a righteous definition of this statement in mathematical terms before continuing the path to the operational definition of quark and gluon jets.

Theorem (Mutual irreducibility)**.** *Two probability density functions $P^k_q(\mathcal{O}_{\text{JSS}})$ and $P^k_g(\mathcal{O}_{\text{JSS}})$ defined on the support $\mathcal{O}_{\text{JSS}}$ are mutual irreducible iff $\exists!\lambda \in [0,1)$ with*

$$P^k_g(\mathcal{O}_{\text{JSS}}) = \lambda P^k_q(\mathcal{O}_{\text{JSS}}) + (1-\lambda)P^{k,\prime}(\mathcal{O}_{\text{JSS}}), \tag{7.9}$$

with $P^{k,\prime}(\mathcal{O}_{\text{JSS}})$ being some probability function.

As it will be discussed later, there is a certain class of observables in QCD that result in mutual irreducible distributions according to Theorem 7.2.

To summarize: the regions in phase space enriched by quark/gluon jets are given by the extreme points of the likelihood ratio $L^k_{q/g}$ of the underlying quark/gluon probability distributions. According to Equation 7.4, the likelihood ratio of the quark/gluon distributions and the corresponding likelihood ratio of probability distributions (i.e., the measured differential cross section) of the two mixtures $L^k_{M_1/M_2}$ are monotonically related to each other, which especially implies that their extreme points are related as well. At this stage of the derivation, it is beneficial to introduce a new quantity – the so-called *reducibility factor* of two distributions

$$\kappa_{\text{AB}} = \inf_{\mathcal{O}} \frac{P_{\text{A}}(\mathcal{O})}{P_{\text{B}}(\mathcal{O})}, \tag{7.10}$$

which corresponds to the *infimum* (in practice, the continuous probability distributions are approximated by binned histograms, in which case the *infimum* turns into a *minimum*) of the likelihood ratio of two probability distributions P_A and P_B. The value of the observable $\mathcal{O}_{\text{JSS}}$ that satisfies the infimum(minimum) operation is referred to as the *anchor (bin)*

$$\mathcal{O}^{\downarrow} := \arg \inf_{\mathcal{O}}(\min_{\mathcal{O}}) \frac{P_{\text{A}}(\mathcal{O})}{P_{\text{B}}(\mathcal{O})}, \tag{7.11}$$

which plays a crucial role in the extraction of the underlying quark/gluon distributions as well as the fractions as will be discussed later. The anchor bins can be considered as the region

(bin) in phase space that is assumed to be (almost) pure in either quarks or gluons.

The requirement of mutual irreducibility can now be expressed in terms of the reducibility factors according to Equation 7.10.

Theorem (Mutual irreducibility)**.** *Two probability density functions $P_q^k(\mathcal{O}_{\text{JSS}})$ and $P_g^k(\mathcal{O}_{\text{JSS}})$ defined on the support $\mathcal{O}_{\text{JSS}}$ are mutual irreducible iff the reducibility factors κ_{qg} and κ_{gq} simultaneously vanish*

$$\kappa_{qg} = \inf_{\mathcal{O}_{\text{JSS}}} \frac{P_q^k(\mathcal{O}_{\text{JSS}})}{P_g^k(\mathcal{O}_{\text{JSS}})} = 0, \tag{7.12}$$

$$\kappa_{gq} = \inf_{\mathcal{O}_{\text{JSS}}} \frac{P_g^k(\mathcal{O}_{\text{JSS}})}{P_q^k(\mathcal{O}_{\text{JSS}})} = 0. \tag{7.13}$$

The equivalence of Theorem 7.2 and 7.2 is directly apparent by substituting $P_g^k(\mathcal{O}_{\text{JSS}}) = \lambda P_q^k(\mathcal{O}_{\text{JSS}}) + (1-\lambda)P^{k,\prime}(\mathcal{O}_{\text{JSS}})$ into the definition of the reducibility factor and requiring mutual irreducibility, i.e., $\exists!\lambda \in [0,1)$.

The motivation for introducing the reducibility factors κ is that the reducibility factors of the two mixtures (which can directly be measured at distribution level!) are directly related to the *quark/gluon fractions* in Equation 7.4. This means, that the degeneracy of the system of linear equations is resolved by the additional requirement of the jet substructure variable $\mathcal{O}_{\text{JSS}}$ resulting in distributions that are mutual irreducible. This can be seen by applying the infimum function to both sides of Equation 7.5 leading to

$$\kappa_{M_iM_j}^k = \inf_{\mathcal{O}_{\text{JSS}}} L_{M_1/M_2}^k(\mathcal{O}_{\text{JSS}}) = \inf_{\mathcal{O}_{\text{JSS}}} \frac{f_q^{M_i^k} L_{q/g}^k(\mathcal{O}_{\text{JSS}}) + f_g^{M_i^k}}{f_q^{M_j^k} L_{q/g}^k(\mathcal{O}_{\text{JSS}}) + f_g^{M_j^k}}, \tag{7.14}$$

with $\kappa_{M_iM_j}^k$ being the reducibility factors of the kth transverse momentum bin. Assumed that the first mixture M_1 is quark-enriched and M_2 is gluon-enriched, Equation 7.14 can be entirely expressed in terms of reducibility factors at the level of distributions by

$$\kappa_{M_1M_2}^k = \frac{f_q^{M_1^k}\kappa_{qg}^k + f_g^{M_1^k}}{f_q^{M_2^k}\kappa_{qg}^k + f_g^{M_2^k}}, \qquad \kappa_{M_2M_1}^k = \frac{f_q^{M_2^k} + f_g^{M_2^k}\kappa_{gq}^k}{f_q^{M_1^k} + f_g^{M_1^k}\kappa_{gq}^k}, \tag{7.15}$$

whereby the infimum operation was applied to the nominator and denominator using the monotony of L_{M_1/M_2}^k with respect to $L_{q/g}^k$.

Now, putting things together. If, *and only if*, the underlying quark gluon distributions are mutual irreducible, i.e., $\kappa_{qg}^k = 0$ and $\kappa_{gq}^k = 0$ (see Theorem 7.2), the fraction of quark $f_q^{M_i^k}$ and gluon $f_g^{M_i^k}$ jets in the different samples are directly given by the *measured* reducibility factors of the two mixtures $\kappa_{M_iM_j}^k$ according to

$$\kappa_{M_1M_2}^k = \frac{f_g^{M_1^k}}{f_g^{M_2^k}}, \qquad \kappa_{M_2M_1}^k = \frac{f_q^{M_2^k}}{f_q^{M_1^k}}, \tag{7.16}$$

which can finally be solved for the unknown fractions given the measured values of the reducibility factors of the two different mixtures and taking into account that only two categories (quarks and gluons) are assumed, which implies the additional constrain $f_g^{M_i^k} + f_q^{M_i^k} = 1$. Solving the Equations 7.22 for the quark fractions gives

$$f_q^{M_1^k} = \frac{1-\kappa_{M_1M_2}^k}{1-\kappa_{M_1M_2}^k\kappa_{M_2M_1}^k}, \qquad f_q^{M_2^k} = f_q^{M_1^k}\kappa_{M_2M_1}^k, \tag{7.17}$$

which is the desired result. Applying those quark/gluon fractions to Equation 7.4 also allows

to extract the underlying distributions of quarks and gluons in terms of the reducibility factors

$$P_q^k(\mathcal{O}_{\mathrm{JSS}}) = \frac{P_{M_1}^k(\mathcal{O}_{\mathrm{JSS}}) - \kappa_{M_1M_2}^k P_{M_2}^k(\mathcal{O}_{\mathrm{JSS}})}{1 - \kappa_{M_1M_2}^k}, \tag{7.18}$$

$$P_g^k(\mathcal{O}_{\mathrm{JSS}}) = \frac{P_{M_2}^k(\mathcal{O}_{\mathrm{JSS}}) - \kappa_{M_2M_1}^k P_{M_1}^k(\mathcal{O}_{\mathrm{JSS}})}{1 - \kappa_{M_2M_1}^k}, \tag{7.19}$$

which finally solves the problem according to Equation 7.3.

At this point – before closing this section by quoting the final, operational definition of quark- and gluon-initialized jets –, it is appropriate to pause for a moment and reflect briefly on what has just been said. Apparently – almost miraculously –, the initially underconstrained problem 7.4 becomes solvable according to Equations 7.17 and 7.19 in a solely data-driven manner. This may be very surprising at first; but, closely following the individual steps of the derivation above illustrates that the additional information needed in order to obtain a solvable system of equation comes from the further constrain on the quark/gluon discriminant to yield probability distributions that are mutually irreducible. Without this requirement, the quark/gluon fractions (and consequently the underlying distributions as well) can not be extracted because the quark/gluon reducibility factors (which are experimentally inaccessible) do not vanish as can be seen in Equation 7.15. Those reducibility factors can be interpreted in the context of perturbative QCD, which is the content of the subsequent section. However, before the attention will be directed at the interpretation of the reducibility factors in terms of some concrete jet substructure observables (relevant for this analysis), this section closes by condensing all the insights just acquired into the final operational definition of quark and gluon jets, which provides the foundation of the measurement of quark/gluon fractions presented in this thesis.

Definition 2 (Operational Definition of Quarks/Gluon Jets [42]). *Given two samples M_1 and M_2 of QCD jets at a fixed p_{T} [bin $k \in \mathbb{N}$] obtained by a suitable jet-finding procedure, taking M_1 to be "quark-enriched" compared to M_2 [(which is "gluon-enriched")], and a jet substructure feature space $\mathcal{O}$, the quark and gluon jet distributions are defined to be:*

$$P_q^k(\mathcal{O}_{\mathrm{JSS}}) = \frac{P_{M_1}^k(\mathcal{O}_{\mathrm{JSS}}) - \kappa_{M_1M_2}^k P_{M_2}^k(\mathcal{O}_{\mathrm{JSS}})}{1 - \kappa_{M_1M_2}^k}, \tag{7.20}$$

$$P_g^k(\mathcal{O}_{\mathrm{JSS}}) = \frac{P_{M_2}^k(\mathcal{O}_{\mathrm{JSS}}) - \kappa_{M_2M_1}^k P_{M_1}^k(\mathcal{O}_{\mathrm{JSS}})}{1 - \kappa_{M_2M_1}^k}; \tag{7.21}$$

accordingly, the fractions of quarks(gluons) in the two mixtures are given by

$$\kappa_{M_1M_2}^k = \frac{f_g^{M_1^k}}{f_g^{M_2^k}}, \qquad \kappa_{M_2M_1}^k = \frac{f_q^{M_2^k}}{f_q^{M_1^k}}, \tag{7.22}$$

where $\kappa_{M_1M_2}^k$, $\kappa_{M_2M_1}^k$, $P_{M_i}^k(\mathcal{O}_{\mathrm{JSS}})$ are directly obtainable from M_1 and M_2.

As immediately apparent, the operational definition only makes a statement of quarks and gluons at the level of the measured hadronic cross section, i.e., it does *not* attempt to define quarks and gluons at the level of individual jets in the event, but rather it terms of well-defined probability distributions. That is in harmony with the conceptional Definition 2, which defines quark and gluon jets by enriched regions of phase space. Furthermore, as stated in Ref. [42], the interpretation of quark- and gluon-initialized jets also is in accordance with QFT, which only allows to derive theoretical predictions by probability distributions like cross sections rather than making predictions for single events. By abandoning the aspiration to assign/force a label to each jet in an event, one obtains a well-defined interpretation of quarks and gluons at distribution level; however, there are situations where a jet-on-jet label is indispensable like, for instance, the already mentioned case where the jet sample is used to train a neural network in a supervised fashion. In those circumstances, the operational Definition 2 can also be used to assign a per-jet label provided that the underlying distributions of quark- and gluon-initialize

jets are already known. For this, the JSS observable $\mathcal{O}_{\mathrm{JSS}}$) (for which the underlying quark and gluon distributions have been extracted) must be reconstructed for each jet J

$$\omega_g^J = \frac{f_g^{M^k} P_g^k(\mathcal{O}_{\mathrm{JSS}})}{f_q^{M^k} P_q^k(\mathcal{O}_{\mathrm{JSS}}) + f_g^{M^k} P_g^k(\mathcal{O}_{\mathrm{JSS}})} \in [0,1]. \tag{7.23}$$

Using Equation 7.23 result in an additional event weight (which must be taken into account when populating the histograms) that indicates the probability of the current jet being initialized by a gluon (or quark $\omega_q^J = 1 - \omega_g^J$). *Consequently*, using the operation definition to assign a per-jet label does *not* correspond in a binary decision, but rather accounts for the fact that in some circumstances such a choice is neither possible nor reasonable and therefore assigns a "soft label" in terms of a weight.

7.3 Mutual Irreducibility in Light of QCD

The previous section emphasized the significance of the quark/gluon discriminants to result in mutual irreducible distributions. In case of the observable fulfilling this property, the extracted underlying probability distributions of the underlying classes as well as the fractions, respectively, are guaranteed (in the limit of the quality at which the underlying assumptions are fulfilled) to correspond to quarks and gluons (see Equation 7.15). In this section, the concept of mutual irreducibility is further investigated and explained through concrete examples of some jet substructure observables in QCD. The goal is to construct categories of observables that are mutual irreducible and hence suited for the purpose of this analysis.

As a reminder, the definition of mutual irreducibility is given in Theorem 7.2. To put the definition into words, mutual irreducibility implies that the underlying probability distributions of quarks and gluons $P_{q/g}^k(\mathcal{O}_{\mathrm{JSS}})$ over a jet substructure variable $\mathcal{O}_{\mathrm{JSS}}$ are not a mixtures of each other given another (arbitrary) probability distribution.

As it was shown in the previous section, a necessary (and sufficient condition) for an observable to result in mutual irreducible distributions is that the associated reducibility factors $\kappa_{qg}^k = \kappa_{qg}^k = 0$ vanish. This rather abstract condition can be translated into statements about jet substructure observables in QCD by recalling that vanishing reducibility factors imply that the current feature space (in which the jet is represented) results in phase space regions that are (sufficiently) pure either in quark or gluon jets. Hence, mutual irreducibility is a property of jet-related observables that encodes information on whether the quark-gluon discriminator is suitably powerful to result in sufficiently enriched regions in phase space for the extraction of the underling quark/gluon distributions and fractions to succeed or not. This is a qualitative statement; however, mutual irreducibility can be checked for jet substructure observable explicitly provided that analytic calculations exists. In the limit of Leading-Logarithmic (LL) calculations, mutual irreducibility is guaranteed for all jet observables that count the number of emissions inside of a jet [9]. The reason for all substructure observables that count the number of emissions inside of a jet being mutual irreducible at LL is due to the identical emission profile which is why the probability density only differs by the color factors ($C_{\mathrm{F}} = 4/3$ and $C_{\mathrm{A}} = 3$) between quark- and gluon initialized jets [29]. This property results in counting-observables being (more or less) approximately Poission distributed (because the mean and the variance of the distribution scales with the number of the emissions inside the jet) for which mutual irreducibility follows immediately.

The classes of mutual and non-mutual irreducible substructure observables (at LL) can be generalized further by studying their properties at LL (the following content is based on the discussion in Section 3 ("Leading-Logarithmic Analysis") of the paper *"Casimir Meets Poisson: Improved Quark/Gluon Discrimination with Counting Observables"* by Christopher Frye *et al.*

[9] That is one of the main reasons for measuring the inclusive multi-differential cross section of dijet events, among other variables, with respect to the *Soft Drop multiplicity* n_{SD} for different configurations of the Soft Drop algorithm because it results in mutual irreducible distributions of the underlying quark/gluon distributions. This will be proven explicitly in Chapter 11.1.4 by means of a LL calculation of this observable. The conclusion, i.e., counting-based observables being mutual irreducible at LL, is anticipated unproven at this point.

$$\lambda_\beta^\kappa = \sum_{j\in J} z_i^\kappa \theta_i^\beta \qquad (7.27)$$

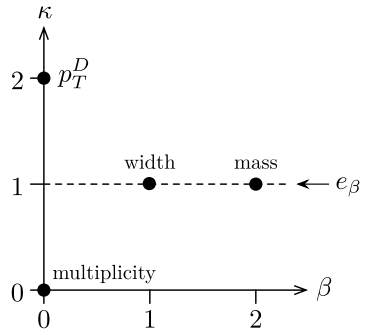


Fig. 7.4: "Visualization of the space of observables λ_β^κ, which includes several well-known jet observables used in quark/gluon discrimination[.]" (adapted and modified from Ref. 24, Fig. 1, p. 3).

[26]).

As heuristically motivated – a rigorous derivation and prove is provided later in Chapter 11.1.4 –, counting-based observables are approximately Poisson distributed $P_{q/g}(n_{\text{count}}) = \mu(C_{\text{F/A}})^{n_{\text{count}}} e^{-\mu(C_{\text{F/A}})}/n_{\text{count}}!$, with the mean of the distribution $\mu(C_{\text{F/A}})$ depending on the color factor. Applying the definition of the reducibility factors according to Equation 7.10 yields

$$\kappa_{qg} = \inf_{n_{\text{count}}} \frac{P_q(n_{\text{count}})}{P_g(n_{\text{count}})} = \inf_{n_{\text{count}}} \left[\frac{\mu(C_{\text{F}})}{\mu(C_{\text{A}})}\right]^{n_{\text{count}}} e^{\mu(C_{\text{A}})-\mu(C_{\text{F}})} = 0, \qquad (7.24)$$

$$\kappa_{gq} = \inf_{n_{\text{count}}} \frac{P_g(n_{\text{count}})}{P_q(n_{\text{count}})} = \inf_{n_{\text{count}}} \left[\frac{\mu(C_{\text{A}})}{\mu(C_{\text{F}})}\right]^{n_{\text{count}}} e^{\mu(C_{\text{F}})-\mu(C_{\text{A}})} = e^{\mu(C_{\text{F}})-\mu(C_{\text{A}})} \approx 0, \qquad (7.25)$$

whereby the first expression vanishes because $C_{\text{F}}/C_{\text{A}} < 1$ and $n_{\text{count}} \in \mathbb{N}$. The second reducibility factor does not vanish exactly, but it is exponentially suppressed by the area, Π, of the emission plane $\mu(C_{\text{F}}) - \mu(C_{\text{A}}) \propto (C_{\text{F}} - C_{\text{A}})\Pi$.

The next class of substructure observables are those defined by the sum over the jet's constituent four-momenta (or derived quantities) according to

$$f(\{p_j\}_{j\in J}) = \sum_{j\in J} f(p_j), \qquad (7.26)$$

with j being one arbitrary constituent in the reconstructed jet J. There is a large variety of quark/gluon discriminants that are of the form of Equation 7.26, like, for instance, the family of generalized angularities $\lambda_\beta^\kappa = \sum_{j\in J} z_i^\kappa \theta_i^\beta$ [169, 171], whereby κ and β control the momentum and angle weighting. Depending of the configuration of (κ, β), Equation 7.26 corresponds to different observables (see Figure 7.4): $(0,0) \cong$ constituent multiplicity (see counting-based observables), $(2,0) \cong$ "transverse momentum", $(1,1) \cong$ jet width or broadening, $(1,2) \cong$ mass etc. Going back to Equation 7.26, at LL, emissions are uniformly distributed in the plane spanned by $(\log 1/\theta, \log 1/z)$ (the bulk commonly observed in the Lund jet plane is due to hadronization effects), with a constant density. In the strongly-ordered limit and assuming f to be linear in energy ($\kappa = 1$ in Equation 7.26), the lowest analytic prediction at LL is given by the *dominant* emission in the jet

$$\sum_{j\in J} f(p_j) \xrightarrow{\text{LL}} \max_{j\in J} f(p_j), \qquad (7.28)$$

in which case the probability for the observable satisfying $f(p_j) < f_{\max}$ being equal to the probability of *not* having an emission(=Sudakov form factor, see Chapter 2.4) with $f(p_j) > f_{\max}$, i.e.,

$$\Sigma_i(f_{\max}) := \int_{[0,f_{\max}]} \mathrm{d}f\, p(f) = e^{-\frac{2\alpha_{\text{S}} C_i}{\pi} A(f_{\max})}, \qquad (7.29)$$

with $A(f_{\max}) = \int_{f(t,\theta)>f_{\max}} \frac{\mathrm{d}\theta}{\theta}\frac{\mathrm{d}z}{z}$ denoting the excluded (no emissions) area in the Lund plane and $2\alpha_{\mathrm{S}}C_i/\pi$ being the constant density in phase space due to the aforementioned uniformly distributed emissions. Calculating the cumulative distributions according to Equation 7.29 for quarks and gluons, one find the following "famous" relation

$$\Sigma_g = [\Sigma_q]^{\frac{C_{\mathrm{A}}}{C_{\mathrm{F}}}} = [\Sigma_q]^{\frac{9}{4}}, \tag{7.30}$$

which is generally known as *Casimir scaling*. This simple calculations reveals a property of Sudakov-distributed observables (observable with a Sudakov form factor at LL accuracy), i.e., their discrimination power with respect to quarks and gluons as measured by the ROC curve (see Equation 7.30) is generally limited by Casimir scaling, which states that the efficiency of gluon identification is limited by $\epsilon_q^{C_{\mathrm{A}}/\mathrm{F}}$. Based on Equation 7.30, the upper limit on the discrimination power of an observable "suffering" from Casimir scaling can be computed based on the AUC according to $\mathrm{AUC}_{\mathrm{Casimir}} = 1 - \int_{[0,1]} \mathrm{d}\Sigma_q\, \Sigma_q^{\frac{9}{4}} \approx 0.70$, which is an upper limit as higher-order corrections reduce classification performance.

With the quark-gluon distributions related by Casimir scaling, the reducibility factors for Sudakov-distributed observables can be simple calculated

$$\kappa_{qg} = \inf_f \frac{P_q(f)}{P_g(f)} = \frac{C_{\mathrm{A}}}{C_{\mathrm{F}}} \inf_f \Sigma_q(f)^{1-\frac{C_{\mathrm{A}}}{C_{\mathrm{F}}}} = \frac{C_{\mathrm{A}}}{C_{\mathrm{F}}}, \tag{7.31}$$

$$\kappa_{gq} = \inf_f \frac{P_g(f)}{P_q(f)} = \frac{C_{\mathrm{A}}}{C_{\mathrm{F}}} \inf_f \Sigma_q(f)^{\frac{C_{\mathrm{A}}}{C_{\mathrm{F}}}-1} = 0, \tag{7.32}$$

whereby the factor $C_{\mathrm{A}}/C_{\mathrm{F}}$ is due to the normalization of the probability distributions as the cumulative distribution and the probability density are related by $P_{q/g}(f) = \partial_f \Sigma_{q/g}(f)$.

As obvious from Equation 7.31, the irreducibility factors do not nearly vanish simultaneously – not even exponentially as it was the case for the counting-based observables that are not dominated by the leading emission. Therefore, all observables in accordance with Equation 7.26 do *not* result in mutual irreducible probability distributions and therefore can not be used to extract the underlying fractions of quarks and gluons based on the operational Definition 2. The problem of non-mutual irreducible substructure variables can be "solved" or "bypassed" by making use of the fact that the extracted fractions (based on mutual irreducible variables of course) are universal, i.e., once the quark/gluon fractions are known, they can be used to resolve the degeneracy in Equation 7.4 and the underlying distributions with respect to a non-mutual irreducible observable can be calculated according to Equation 7.19. An alternative approach (which will be further explained in one of the upcoming sections) is very similar, but relies on the simultaneous measurement (the joint distribution) of a variable that is mutual irreducible and another one that is non-mutual irreducible. In this case, obtaining the underlying quark/gluon distributions of the non-mutual irreducible observable is simply given by a reweighting procedure, which is explained in great detail in Appendix 7.4.

7.4 Extract Topics for Non-Mutual Irreducible Observables

Since mutual irreducibility for $P_{q/g}(x)$ with $P_{M_a}(x) = f_q^{(a)} P_q(x) + f_g^{(a)} P_g(x)$ is guaranteed (x could be, for instance, the constituent multiplicity n_{const}, the SD multiplicity n_{SD} or the NN-score z_{NN}), the jet topics $P_{T1}(x)$, $P_{T2}(x)$ (to adopt the language of the previous section) and the corresponding fractions can be extracted. To get the quark/gluon distributions for an observable y that does *not* lead to mutual irreducibility (like, e.g., the jet mass or the transverse momentum of the jet), the idea is to measure the joint distributions (for instance a *multi*-differential cross section)

$$P_{M_a}(y, x), \tag{7.33}$$

whereby M_a with $a \in \{1, 2\}$ denotes different statistical *mixtures*, to extract the topics $P_{T1}(y)$, $P_{T2}(y)$ in a simultaneous measurement of x and y, which is associated with the underlying quark/gluon probability distributions over the respective jet substructure observable.

The first step is to express the probability distribution $P(y)$ as a marginal distribution over a "latent" variable x

$$P(y) = \int \mathrm{d}x\, P(y,x), \tag{7.34}$$

with $P(y,x)$ being the joint distribution marginalized over the two different statistical mixtures $P(y,x) = \sum_{M_a} P_{M_a}(y,x)$ necessary for topic extraction. Factorizing the joint distribution according to Equation 7.34 yields

$$P(y) = \int \mathrm{d}x\, P(y|x)P(x), \tag{7.35}$$

with $P(x) = \sum_{M_a} P_{M_a}(x)$ and $P_{M_a}(x) = f_q^{(a)} P_q(x) + f_g^{(a)} P_g(x)$ whereby the fractions $f_{q/g}^{(a)}$ and topics $P_{q/g}(x)$ are known because x is *by definition* mutually irreducible. The latter point is crucial because the quark/gluon fractions – once derived based on an observable that is mutual irreducible – are independent regardless of the respective observable – they are *universal*. The distribution $P(x)$ of the observable x, leading to mutual irreducibility, can be reexpressed in terms of the topics and the quark/gluon fractions

$$\begin{aligned}
P(x) &= \sum_a P_{M_a}(x) && (7.36)\\
&= \sum_a \left\{ f_q^{(a)} P_q(x) + f_g^{(a)} P_g(x) \right\} && (7.37)\\
&= f_q^{(1)} P_q(x) + f_g^{(1)} P_g(x) + f_q^{(2)} P_q(x) + f_g^{(2)} P_g(x) && (7.38)\\
&= P_q(x) \left(f_q^{(1)} + f_q^{(2)} \right) + P_g(x) \left(f_g^{(1)} + f_g^{(2)} \right) && (7.39)\\
&= 2P_q(x) f_q + 2P_g(x) f_g, && (7.40)
\end{aligned}$$

whereby the total fractions $f_q := (f_q^{(1)} + f_q^{(2)})/2$ and $f_g := (f_g^{(1)} + f_g^{(2)})/2$ have been introduced. The very same can be done for $P(y)$ according to

$$P(y) = 2P_q(y) f_q + 2P_g(y) f_g. \tag{7.41}$$

Using those results, Equation 7.41 can be written as

$$P_q(y) f_q + P_g(y) f_g = \int \mathrm{d}x\, P(y|x) \left\{ f_q P_q(x) + f_g P_g(x) \right\}, \tag{7.42}$$

whereby the additional factor of '2' cancels. This equation can be factorized into two linear independent terms

$$P_q(y) f_q + P_g(y) f_g = \int \mathrm{d}x\, P(y|x) \left\{ P_q(x) f_q + P_g(x) \right\}, \tag{7.43}$$

$$P_q(y) f_q + P_g(y) f_g = \int \mathrm{d}x\, P(y|x) P_q(x) f_q + \int \mathrm{d}x\, P(y|x) P_g(x), \tag{7.44}$$

which yields

$$f_q \left\{ P_q(y) - \int \mathrm{d}x\, P(y|x) P_q(x) \right\} + f_g \left\{ P_g(y) - \int \mathrm{d}x\, P(y|x) P_g(x) \right\} = 0, \tag{7.45}$$

whereby the clutched expressions have to vanish independently due to $f_{q/g} \neq 0$ and $f_q \neq f_g$ (two different mixtures). This finally gives the two expressions for the quark/gluon distributions for the non-mutual irreducible observable y

$$P_q(y) = \int \mathrm{d}x\, P(y|x) P_q(x), \tag{7.46}$$

$$P_g(y) = \int \mathrm{d}x\, P(y|x) P_g(x). \tag{7.47}$$

The derivation of the quark/gluon distributions of the observable y can be summarized by the following statement.

Corollary 0.1. *Given that the topics $P_g(x)$ and $P_q(x)$ for the observable x that, by definition, leads to mutual irreducible distributions, the two probability functions $P_g(y)$ and $P_q(y)$ for 'any' observable y are mutual irreducible too if*

$$P_{q/g}(y) = \int \mathrm{d}x\, P(y|x) P_{q/g}(x). \tag{7.48}$$

Proof. The proof is done by contradiction. Assumed the two distributions $P_g(y)$ and $P_q(y)$ with $P_{q/g}(y) = \int \mathrm{d}x\, P(y|x) P_{g/g}(x)$ are *not* mutual irreducible, i.e., $\exists \lambda \in [0,1]$ such that

$$P_g(y) = \lambda P_q(y) + (1-\lambda) P'(y), \tag{7.49}$$

with $P'(y)$ being some (arbitrary) probability function satisfying this condition. Using the definition of $P_g(y)$ and $P_q(y)$ in terms of the joint distribution, it follows that

$$\begin{aligned}
0 &= P_g(y) - \lambda P_q(y) - (1-\lambda) P'(y), && (7.50)\\
&= \int \mathrm{d}x\, P(y|x) P_g(x) - \lambda \int \mathrm{d}x\, P(y|x) P_q(x) - (1-\lambda) P'(y), && (7.51)\\
&= \int \mathrm{d}x\, P(y|x) \left\{ P_g(x) - P_q(x)\lambda \right\} - (1-\lambda) P'(y), && (7.52)\\
&= \int \mathrm{d}x\, P(y|x) \left\{ P_g(x) - P_q(x)\lambda \right\} - (1-\lambda) \int \mathrm{d}x\, P(y|x) P'(x), && (7.53)\\
&= \int \mathrm{d}x\, P(y|x) \underbrace{\left\{ P_g(x) - P_q(x)\lambda - (1-\lambda) P'(x) \right\}}_{\neq 0} \neq 0 && (7.54)
\end{aligned}$$

The term in the brackets is guaranteed to *not* vanish because $P_g(x)$ and $P_q(x)$ are mutual irreducible; hence, the mutual irreducibility of $P_g(y)$ and $P_q(x)$ follows immediately.

□

Extracting the "jet topics" (for non-mutual irreducible observables) based on a joint measurement offers the advantage of easily propagating uncertainties from the unmixed cross section (topics) onto the underlying quark/gluon distributions of the non-mutual irreducible observable. However, one is not limited to a simultaneous, i.e., multi-dimensional measurement in order to obtain "jet topics" for non-mutual irreducible observables. In fact, once the degeneracy in Equation 7.3 has been broken by extracting the quark/gluon fractions, quark/gluon distributions can be measured for *all* observables provided that the phase space selection is identical. The latter point is utterly important since even slight changes in the event selection may result in considerable differences in the phase space and hence different fractions of quark- and gluon-like jets. Therefore, if the event/jet selection changes, the quark/gluon fractions must *re*measured based on a mutual irreducible observable.

Given the "universal" quark/gluon fractions $f_i^{M_\alpha^k}$ have been measured in some phase space volume, the underlying quark and gluon distributions of *any* JSS observable (mutual as well as non-mutual irreducible) are given by

$$P_g^k(\mathcal{O}_{\mathrm{JSS}}) = \frac{f_q^{M_2^k} P_{M_1}^k(\mathcal{O}_{\mathrm{JSS}}) - f_q^{M_1^k} P_{M_2}^k(\mathcal{O}_{\mathrm{JSS}})}{f_q^{M_1^k} f_g^{M_2^k} - f_q^{M_1^k} f_g^{M_2^k}}, \tag{7.55}$$

$$P_q^k(\mathcal{O}_{\mathrm{JSS}}) = \frac{f_g^{M_2^k} P_{M_1}^k(\mathcal{O}_{\mathrm{JSS}}) - f_g^{M_1^k} P_{M_2}^k(\mathcal{O}_{\mathrm{JSS}})}{f_g^{M_1^k} f_q^{M_2^k} - f_g^{M_1^k} f_q^{M_2^k}}, \tag{7.56}$$

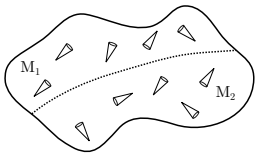
provided the substructure variable was measured based on the same mixtures of jets.

7.5 Extract "Jet Topics" – A Step-by-Step Instruction

This chapter introduced a data-driven method that is used to extract the underlying quark/gluon (probability) distributions as well as the (universal) quark/gluon fractions of two disjoint mixtures of jets. The motivation and theoretical background of the procedure was explained in great detail in Section 7.2 and a connection with probabilistic topic modeling from data science was established in Section A. So far, the emphasis was to provide a sound mathematical basis in terms of clear definitions and rigorous statements like, for instance, the importance of mutual irreducibility and its interpretation within the framework of QCD (see Section 7.3). This, however, may not be the ideal language for gaining intuition of how to apply this method in a real-world measurement. Those who encounter the presented method for the very first time might miss a clear presentation of the individual steps necessary to obtain quark/gluon fractions and topics from the measurement of a hadronic jet cross section.

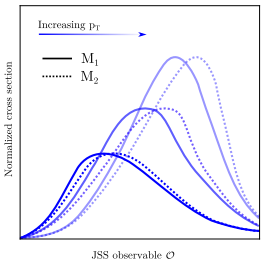
This section completes the chapter "Data-Driven Feature Extraction using Topic Modeling" by providing a clean *modus operandi* based on a step-by-step instruction on how to extract underlying quark/gluon features and fractions in a jet substructure measurement at hadron colliders.

Two mixtures of jets with different quark/gluon fractions The first step of the measurement is to define two mixtures of jets (see, e.g., Equation 7.3) that results in two *disjoint* samples of jets with different fractions of quark- and gluon-initialized jets. There are several ways to define different jet mixtures.



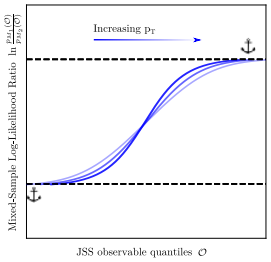
One option is to use two different processes that are known to result in quark- and gluon-enriched phase spaces. For instance, an obvious choice would be to select $Z+$jets (e.g., by triggering on the decay products of the Z) to obtain a sample enriched by quarks and multijet events to get a second mixture dominated by gluons. This works fine. However, using two processes is a significant complication from an experimentalist's point of view. Furthermore, mixing jets from different processes introduces the aforementioned problem of *sample dependence*.

An alternative approach of obtaining two mixtures of jets (which is used for the measurement presented in this thesis) is to use the very same process, but to introduce an additional selection criterion that provides some sensitivity to quark and gluon jets. In this analysis, this is done by measuring the relative dijet production rate for central(="gluon-enriched") and forward(="quark-enriched") jets, which will be further elaborated later.

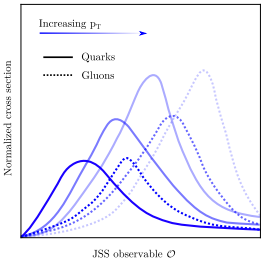


Measurement of the (multi-)differential cross section Once the two jet samples of quark/gluon mixtures have been defined, the (multi-)differential cross section with respect to some (mutual irreducible) quark/gluon discriminant is measured based on the two mixtures of jets defined the aforementioned first step. After the cross section has been reconstructed (in data and simulation), it is unfolded to particle level to correct the measured spectra in data for the finite detector efficiency as well as acceptance and resolution effects from the detector (see Section 5). In this measurement, the dijet cross section is simultaneously unfolded in central and forward region. This procedure is repeated for several systematic variations to propagate uncertainties through the detector that are translated into an uncertainty on the final "jet topics" as well as the extracted quark/gluon fractions. After unfolding, the (multi-)differential cross section is at hadron level (corrected for detector effects) and ready for the calculation of the reducibility factors. *Comment:* This step basically is an *entire* analysis on its own!

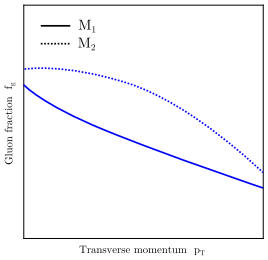
Extraction of the reducibility factors $\kappa_{M_i M_j}$ from the likelihood of the two mixtures Once the (multi-)differential cross section measurement has been unfolded to particle level, the reducibility factors $\kappa_{M_i M_j}$ (needed to extract the underlying quark/gluon distributions according to Equation 7.19 as well as the corresponding fractions in the mixtures based on Equation 7.17) must be calculated in accordance with Equation 7.10. This must be done for *each* bin k in transverse momentum individually, i.e., for each p_T bin, the ratio of the corresponding *single*-differential cross section $(1/\sigma)\,\mathrm{d}\sigma^k_{M_\alpha}/\mathrm{d}\mathcal{O}_\mathrm{JSS}$ with respect to the substructure observable for the two mixtures M_1 and M_2 is calculated and the anchor bins/regions (see Equation 7.11) thereof. The extraction of the anchors is much more involved than described here – further information on the concrete determination of the anchor regions will follow shortly.



Extraction of the underlying quark/gluon distributions ("jet topics") After the reducibility factors $\kappa^k_{M_i M_j}$ have been extracted for each bin in transverse momentum of the measurement based on the ratio of the (relative) jet rates in the two mixtures, the underlying quark/gluon distributions (the jet topics) (see Equation 7.19) can be calculated by maximally subtracting the cross sections measured in the two mixtures from each other. Provided that the substructure observable is mutual irreducible, the obtained probability distributions are guaranteed (to the degree to which the underlying assumption of mutual irreducibility is fulfilled) to correspond to quark- and gluon-like jet features according to the operational Definition 2.



Extraction of the underlying quark/gluon fractions This step is performed simultaneously to the extraction of the jet topics, but it is listed separately to mimic the order in which the results are presented in this thesis. Like the quark/gluon probability distributions, the fraction of quark- and gluon-like jets is directly calculated from the reducibility factors $\kappa^k_{M_i M_j}$ in each bin of transverse momentum according to Equation 7.17. The result are quark/gluon fractions $f^{M^k_\alpha}_{q/g}$ for each transverse momentum bin and mixture M_α. In the end, the total fraction of quark- and gluon-initialized jets is stated by combining the fractions in central $f^{M^k_1}_{q/g}$ and forward $f^{M^k_2}_{q/g}$ region according to $f^{M^k_\alpha}_{q/g} = (f^{M^k_1}_{q/g} + f^{M^k_1}_{q/g})/2$ to get the total fraction in the dijet samples in each p_T bin.



Part III

Analyses and Results

The third and last part of this thesis finally turns the focus of attention to the actual analysis, i.e., the measurement of the triple-differential multijet cross section as a function of several jet substructure observables (sensitive to characteristic features of hadronic showers initialized by quark and gluon jets, respectively), as well as the subsequent extraction of the underlying quark/gluon probability distributions and the corresponding fractions in the respective regions of phase space volume. However, before the results of the cross section measurement or the quark/gluon fractions are presented, some technical (but absolutely essential) aspects of the analysis must be discussed like, for instance, the selection of events or the definition of the systematic uncertainties etc.

This part is structured as follows: the "first" Chapter 8 introduces all samples in real pp-collision data, which have been recorded within the course of Run 2, and Monte Carlo simulation of the corresponding campaigns that are used in the analysis.

Thereafter, in Chapter 9, the selection of analysis-relevant collision events by means of the trigger as well as the phase space is discussed; furthermore, the reconstruction of the (physical) objects (vertices, tracks, jets) is explained.

The next Chapter 10 is rather technical in nature too; it introduces all the sources of statistical and systematic uncertainties and provides a precise prescription for their calculation.

Chapter 11 presents the results for the measurement of the triple-differential multijet cross section with respect to the counting-based JSS observables. This chapter includes a discussion of the detector-level spectra, the unfolding as well as the unfolded cross section including all sources of systematic and statistical uncertainties.

Chapter 12 is structurally identical to the previous one but for the triple-differential cross section as a function of the NN-based observables instead.

Chapter 13 uses the unfolded cross sections to extract the underlying quark/gluon probability distributions as well as the quark/gluon fractions based on the method that has been explained in the previous chapter.

The last Chapter 14 finishes this thesis by discussing the acquired insights and providing in outlook into possible further studies.

8. Data and Monte Carlo Samples

The triple-differential multijet cross section is measured in pp-collision over a phase space volume that is defined by the inclusive dijet selection criteria at particle *and* detector level. The actual event and object selection cuts (minimum-p_{T} threshold, acceptance cuts etc.) are subject of the upcomming Chapter 9. This chapter introduces all data sets (including real pp-collision data as well as data obtained from Monte Carlo simulation) from which the differential multijet cross section is reconstructed.

The first section (8.1) introduces the proton-proton collision data that was recorded with the ATLAS detector at the LHC (see Chapter 3.1) during the second Run of data-taking, which corresponds to the years from 2015 to 2018 (inclusive). The next section (8.2) introduces data obtained from simulation only. It introduces all state-of-the-art (general purpose) Monte Carlo generators that have been used to simulate multijet events according to the relevant conditions at the LHC such as, e.g., pile-up, beam energy etc. Those samples are available at particle and detector level, whereby the latter one includes a full GEANT4 simulation of the ATLAS detector. The various Monte Carlo generators differ with respect to the underlying PDFs, the order of the calculation of the hard subprocess (matrix element, see Chapter 2.3), the parton shower model (see Chapter 2.4) as well as the hadronization models (see Chapter 2.5). The effect of the different models on the reconstructed multijet spectrum is studied and used to derive systematic uncertainties on the final triple-differential multijet cross section measurement. Furthermore, the response matrix is constructed from the *nominal* Monte Carlo sample, which is used to fully correct the measured spectra for acceptance and detector effects (see unfolding in Chapter 5).

8.1 LHC Run 2 (2015–2018) pp-Collision Data

Going from Run 1 (2009–2013) to Run 2 (2015–2018), the LHC as well as the ATLAS experiment have undergone numerous changes that affect several detector components; changes in the LHC beam conditions; reconstruction and (re-)definition of physical objects; as well as changes in the simulation of the experiment and the underlying physical models.

Concerning the detector system, the IBL (see Chapter 3.3.3) was installed between the beam pipe and the current inner pixel layer and thenceforth is the detector module most close to the beam line and interaction point. The new layer significantly improves the b-tagging efficiency and helps to maintain the track reconstruction performance given the considerable higher delivered luminosity by the LHC during Run 2; it also improves the reconstruction and resolution of primary and secondary (see b-tagging) vertices. Further changes include the repair of plenty of gas leaks in the TRT (see Chapter 3.3.3) that have been observed in the course of Run 1, as well as new muon detector chambers at certain locations of the detector [172].

The most important change regarding the beam conditions of the LHC was the significant increase in the beam energy going from 8 TeV to 13 TeV centre-of-mass energy as well as the increase in the delivered record luminosity. The increase in the centre-of-mass energy and the luminosity has several effects on the event topology such as a change in the fraction of quark and gluons as well as an impact on the pile-up contributions to the event. Besides the centre-of-mass energy, the temporal spacing of the proton bunches was reduced from 50 ns to 25 ns, accompanied by a considerable increase in pile-up and an need for very fast read out of the detector information. Important for the reconstruction of jets, the topo-

cluster reconstruction algorithm was adjusted to match the new pile-up conditions in terms of increased noise contribution. Furthermore, the reconstruction of tracks was updated regarding performance and to account for the additional information provided by the IBL. Also, the calibration scheme of the jet energy scale/resolution was updated to match the new data-taking conditions of the ATLAS experiment and to account for Run 1 data to improve the *in situ* part of the calibration steps (see Chapter 4.4.3).

8.2 Monte Carlo Samples

The ATLAS Physics Modeling Group (PMG) provides a set of dedicated samples that have been generated using different general-purpose Monte Carlo event generators available. Those Monte Carlo generators make use of different underlying models for the distinct steps in the event simulation chain introduced in Chapter 2.2.2 like, for instance, using a matrix element calculation at higher (or lower) order; varying parton shower models and/or different ordering variables and matrix element matching schemes; different hadronization approaches and PDF sets etc.

This section introduces the list of Monte Carlo generated samples that have been used for the measurement of the inclusive triple-differential multijet production rate at a centre-of-mass energy of 13 TeV, which is described in this thesis. In doing so, special attention is paid to the distinguishing features of the different generators, their configuration, as well as the assigned domain of application of the individual samples with respect to this measurement, particularly focusing on the definition of systematic uncertainties.

8.2.1 "Multijet Samples for 13 TeV ATLAS Analyses" [4]

Hadronic jets in the final state are among to the most interesting objects at the ATLAS experiment as they provide an excellent test of our understanding of the strong interaction at short distances by means of (perturbative) quantum chromodynamics. At the LHC, *multijet events*, i.e., events with exclusively hadronic jets being present in the final state[1], are among the most dominant production channels; therefore, they are both a "blessing" and a "curse" at the same time – a "blessing", for those analyses aiming for detailed measurements of jet properties because plenty of those objects are available; a "curse", for all analysis (e.g., many searches) dealing with multijet events as a predominant background process.

In order to perform a measurement of a jet-related observable – regardless of whether it is about a global event shape like, e.g., the thrust axis, or a substructure variable such as the number of charged constituents inside the jet –, one needs a set of Monte Carlo samples for the purpose of comparing data with predictions from theory and to derive uncertainties on the measurement based on our best models currently available. The choice of the Monte Carlo samples and the way they are generated should reflect the range of competing models used to describe the same process or phenomenon where there is no *a priori* justification for one approach being superior to any other.

An important example is the choice of the parton shower algorithm to simulate QCD

Fig. 8.1: Generator labels and their color scheme used through the entire analysis.

[1] More precisely, a multijet event is characterized by multiple hadronic jets being present in the final state with the absence of prompt isolated electrons or muons. This definition ensures that multijet samples are *statistically dominated* by processes involving the strong force.

radiative corrections to the cross section of the hard subprocess. There is some ambiguity in the choice of the ordering/evolution of the parton shower emission history due to the concurrence of several ordering variables in the collinear limit of the approximation. In the collinear limit ($\theta \to 0$), all choices are equivalent and yield the very same result; however, since the parton shower algorithm is used for emission beyond this limit as well, different choices of the evolution scale will give (sometimes substantially!) different results.

Another example is the choice of the hadronization model to simulate the transition from partonic to hadronic final states. Most of the approaches to hadronization (relevant for this analysis) have been discussed in Chapter 2.5, but there are more available. Most of those models show a similar performance[2], some have been optimized for specific applications like hadronic tau decays etc. As for the parton shower model, the concrete choice of the respective model is specific to the concrete application. The answer, to which generator to, use may even depend on the region in phase space or the processes the analysis focuses on.

In general, there is no "Swiss army knife", i.e., one event generator that meets all requirements at once. For this reason, the measurement presented in this thesis makes use of several different event generators each one with a very specific application in the analysis, which will be further elaborated below.

A summary of all Monte Carlo event generators along with their different configurations in terms of the underlying calculation of the matrix element, the parton shower and hadronization model as well as the underlying PDF is shown in Table 8.1. Furthermore, Figure 8.1 shows the generator labels and their color scheme that is used through the entire analysis.

MC Generator	Matrix Element	Parton Shower	Hadronization	PDF(s)
Pythia 8.230 (Nominal)	$2 \to 2$ *@ LO*	Dipole-based p_{T}-ordered	Lund string	NNPDF 2.3
Powheg v2 +Pythia 8.230	$2 \to 2$ *@ NLO*	Dipole-based p_{T}-ordered	Lund string	NNPDF30NLO
Sherpa 2.2.5	$2 \to 2$ @ LO	CSS	*Cluster (AHADIC)*	CT14NNL
	$2 \to 2$ @ LO	CSS	Pythia*'s (v6.4) Lund String*	CT14NNL
Herwig 7.1.3	$2 \to 2$ @ NLO $2 \to 3$ @ LO	*Angular-ordered*	Cluster	MMHT2014NLO
	$2 \to 2$ @ NLO $2 \to 3$ @ LO	*Dipole*	Cluster	MMHT2014NLO

Table 8.1: List of simulated multijet samples based different general-purpose Monte Carlo event generators for the simulation of proton-proton collisions used for the measurements presented in this thesis.

Pythia8

*"*Pythia *is a program for the generation of high-energy physics collision events, i.e.[,] for the description of collisions at high energies between electrons, protons, photons and heavy nuclei."* – literal quote from the official Pythia website [173].

Following the ATLAS recommendations (the collaboration provides a list of recommendations that analyses are supposed to follow in order to ensure comparability as well as "minimum" level of quality and performance standards), the baseline Monte Carlo generator is Pythia version 8.230 [174, 175, 176] that is used to produce the nominal samples for this

[2]This is barely surprising because most models get tuned to data at some point and therefore agree in certain regions of phase space.

analysis (the corresponding cross section is henceforth referred to as the "nominal/baseline" cross section/spectrum). Also, the baseline generator is used to construct the response matrix of the measurement (see Chapter 5.1.1), which unfolds the measured spectra in real data to particle level.

As described in Chapter 2.2.2, the event generation is done in several consecutive steps. In Pythia, the calculation of the matrix element for dijet processes is done at *leading order* in perturbation theory. The generated events are then further interfaced to Pythia's parton shower algorithm, which uses the k_t of the emissions as an ordering variable. The modeling of fragmentation and hadronization step is based on the (Lund) string model that was comprehensively explained in Chapter 2.5.

As described in Ref. [4], renormalization μ_{R} and factorization scales μ_{F} of the shower are set to the geometric mean of the squared transverse masses of the two outgoing particles $\mu_{\mathrm{R}}^2 = \mu_{\mathrm{F}}^2 = (p_{\mathrm{T}}^2 - m_3^2)(p_{\mathrm{T}}^2 - m_4^2)$, whereby the indices '3' and '4' denote the two outgoing particles from the calculation of the matrix element.

The NNPDF23LO PDF set [177] is used during the calculation of the matrix element, the parton shower simulation (space-like showers), and in the simulation of the Multi-Parton Interactions (MPI). Furthermore, to improve the agreement between data and simulation, ATLAS' "A14" [178] set of tuned parameters for the Pythia event generator was used that tunes the generator's parton shower for initial and final state radiation and multiple parton interaction parameters to a number of observables based ATLAS Run 1 data (see previous section).

Powheg/MadGraph5_aMC@NLO+Pythia8

The nominal samples outlined above are generated in a standalone Pythia simulation, i.e., all event generation steps (except for the simulation of the detector of course) explained in Chapter 2.2.2 are done within Pythia *solely*. This results in a calculation of the matrix element at LO as Pythia's matrix element generator only supports calculations at that order in perturbation theory. In order to estimate the effect of higher-order terms in the perturbative expansion (see Chapter 1.2.2), the calculation of the matrix element can be "outsourced" to a dedicated event generator that allows for higher-order calculations of the matrix element.

In case of the *"multijet simulation for* 13 TeV *ATLAS analyses"* [4], there are two MC generators that implement NLO calculations of the hard (sub)process: MadGraph5_aMC@NLO (version 2.3.3.1) [179] and Powheg V2 [180, 181]. In both cases, the hard scattering matrix element of the multijet events are calculated at NLO accuracy and interfaced to Pythia8 (via the "Les Houches" interface [182]) to simulate the remaining components of the event topology.

In this analysis, only the combination of Powheg and Pythia8 (thenceforth referred to as Powheg+Pythia8 throughout the rest of this thesis) is used to estimate the uncertainty associated to the obmission of higher-order contributions/corrections to the matrix element. In the calculation of the matrix element, the transverse momentum of the underlying (LO) Born-level configuration is used as the renormalization and factorization scale. The PDF set, which is used to sample the momenta of the initial partons participating in the hard interaction, is the NNPDF30NLO.

The configuration of the subsequent Pythia8 generator basically corresponds to the setup of the nominal Monte Carlo generator (see previous section), i.e., the NNPDF23LO PDF is used for the shower simulation (ISR) as well as MPI and the "A14" tune is applied. This is very important in order to insure that the uncertainty derived based on the comparison of the nominal Pythia samples with the one obtained from Powheg+Pythia8 really captures the effect on the final cross section due to the different levels of accuracy of the matrix element calculation.

Sherpa

"Sherpa [183] [(not to be confused with eponymous Tibetan ethnic groups native to the mountainous regions of Nepal)] is a Monte Carlo event generator for the Simulation of High-Energy Reactions of PArticles [(Sherpa)] in lepton-lepton, lepton-photon, photon-photon, lepton-hadron and hadron-hadron collisions." – literal quote from the official Sherpa website [184].

It, SHERPA, is – like PYTHIA – a fully standalone Monte Carlo event generator that allows to simulate the entire event chain in high-energy particle collisions up to the complete hadronic final states (the detector simulation, of course, is experiment-specific and therefore not part of the simulation procedure).

The SHERPA samples used in this analysis are generated using SHERPA version 2.2.5 [183]. Like it was the case for the nominal PYTHIA8 samples, the calculation of the matrix element is done at LO for $2 \rightarrow 2$ processes (like, e.g.,multijet production). The subsequent showering is done by SHERPA's implementation of the CSS (Catani-Seymour *dipole* factorization) parton shower [185], with a k_t-based ordering of the emissions. SHERPA uses two different PDF sets: the CT14NNLO [186] PDF set is used for the calculation of the matrix element, while the CT10 set [187] is used for modeling the underlying multi-parton interactions.

The SHERPA-based samples are generally used to estimate the effect of different hadronization models (see Chapter 2.5) on the final hadronic cross section. For this reason, two SHERPA samples have been generated that differ with respect to their hadronization model. As summarized in Table 8.1, there is one SHERPA sample based on the *(Lund) string* hadronization model and another one that generates color-singlet states according to the *cluster-based* approach to hadronization (see Chapter 2.5). More precisely, the concrete implementation of the cluster-based fragmentation is done by SHERPA's native "AHADIC" model [188]; the string-based hadronization follows PYTHIA's (version 6.4 [189]) implementation of the Lund string model. In the latter case, the output generated by SHERPA's parton shower simulation is directly interfaced to PYTHIA6. Furthermore, for all SHERPA samples used in this measurement, the decay to stable final state particles (see Chapter 2.5.1) was simulated based on the decay tables implemented by PYTHIA.

HERWIG(@NLO)

"HERWIG *[(short and sweet)] is a multi-purpose particle physics event generator."* – literal quote from the official HERWIG website [190].

Besides the POWHEG+PYTHIA8 and the MADGRAPH5_AMC@NLO+PYTHIA8 samples, respectively, there are two samples of dijet production generated using HERWIG 7.1.3 [191, 192, 193] that include a calculation of the underlying hard event at NLO. The renormalisation and factorization scales are set to the transverse momentum of the leading jet in the event. For the calculation of the matrix element, the MMHT2014NLO [194] PDF set is used. The hard cross section is calculated using Matchbox [195], which is necessary for the subsequent Matching and Merging step (based MC@NLO-like subtraction algorithm [196]) to match the higher-order calculation to the parton shower.

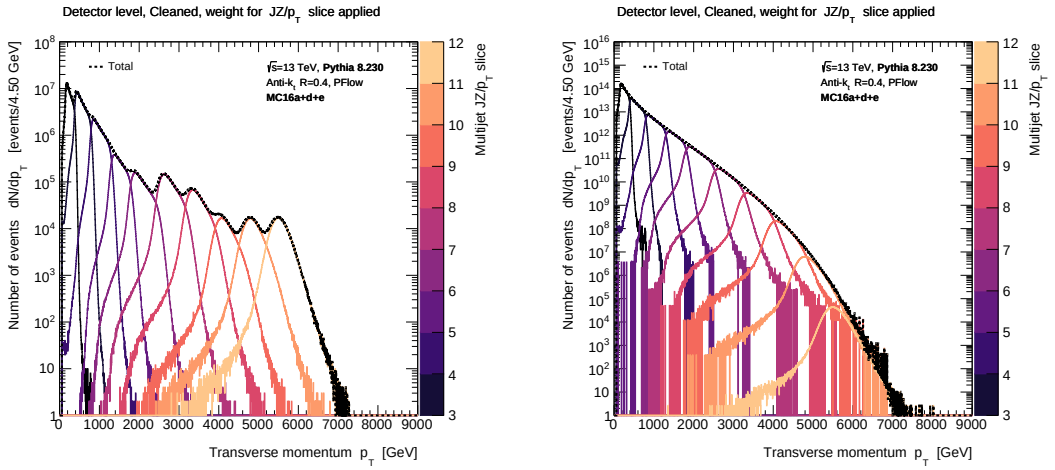
There are two NLO samples based on HERWIG that differ regarding their parton shower model. The first sample makes use of HERWIG's default showering algorithm that orders the emissions based on the radiation angle (angular-ordered). The other sample uses an alternative parton shower model based on a *dipole* model similar to the one used by SHERPA. Both showering algorithms use the same PDF set as for the calculation of the matrix element and are tuned to improve the modeling of quark- and gluon-initialized jets [196].

As can be seen from Table 8.1, both HERWIG samples use a cluster-based model for the final hadronization step.

8.2.2 Event (Re-)Weighting

It is important to ensure that the reconstructed spectrum of the transverse momentum is smooth. This is usually *not* the case because the events are independently generated in different exclusive ranges of transverse momentum with different integrated luminosities to generate sufficient statistics even in the tails of the distribution associated with low production rates. Those ranges are commonly referred to as "p_{T}-" or "JZ-slices", which is a sub-sample of the final multijet dataset. To recover the original (continuous) p_{T} spectrum, the events must be reweighted by a additional weight, which has to be calculated on an event-on-event basis, depending on the transverse momentum of the leading-p_{T} jet, the Data Set IDentification (number) DSID (which is a six-digit number) as well as the run number corresponding to the

respective Monte Carlo campaign (e.g., MC16a, MC16d, MC16e etc.). Great caution needs to be exercised when computing the event weights because discontinuities in the final p_{T} spectra might bias the entire measurement. The respective event weights must be calculated such that



(a) No p_{T} reweighting applied

(b) Correct p_{T} reweighting applied

Plot 8.1: Transverse momentum distribution for PYTHIA8 for each JZ slice and the total spectrum (solid black line) before (*left*) and after (*right*) applying the correction factor for the respective p_{T}/JZ-multijet slice.

the reweighted spectra corresponds to the integrated luminosity of Run 2. To calculate the event weight w_i^{JZ} associated with a certain p_{T}/JZ-slice and to recover the correct multijet cross section, the MC weight w_i^{MC} (which is supplied by the respective Monte Carlo event generator and accounts for, e.g., matching, merging, showering etc., i.e., MC-specific components) must be multiplied by the event filter efficiency ϵ, the (total) cross section and the Luminosity $\mathcal{L}$; finally, the resulting value is normalized by the total number of events *or* the sum of weights depending on the respective weighting scheme (which may differ between Monte Carlo generators) according to

$$w^{\mathrm{JZ}} := \frac{\epsilon\sigma\mathcal{L}}{\sum_{i=1}^{N_{\mathrm{events}}} w_i} \in \mathbb{R}, \tag{8.1}$$

whereby the denominator is given by the sum of all MC weights ($w_i = w_i^{\mathrm{MC}}$) in case of the so-called JZW slicing scheme or the total number of events ($w_i = 1$ for all $i \in \mathbb{N}^{<N_{\mathrm{events}}}$) for JZ-based reweighting approach.

The total event weight (not including any weight-based systematic variations) is then given by

$$w_i^{\mathrm{event}} = w_i^{\mathrm{JZ}} \cdot w_i^{\mathrm{MC}}, \tag{8.2}$$

i.e., applying w_i^{event} takes into account the reweighting due to the Monte Carlo generator as well as the reweighting of the p_{T}-slices to obtain a smooth transverse momentum spectrum. The final result of the reweighting procedure is shown in Figure 8.1. Corresponding distributions exist for *all* generators listed in Table 8.1 and all (substructure) observables studied in the context of this analysis.

An additional comment regarding the total event weight according to Equation 8.2. Applying the total event weight w_i^{event} results in the so-called "nominal" distribution(s)[3]; however, the event weight might include several other factors according to $w_i^{\mathrm{event}} \rightarrow w_i^{\mathrm{event}} \cdot \prod_k w_{k,i}^{\mathrm{var}}$ that

[3]Within the context of this report, the term "nominal" is used in a slightly different context, explicitly referring to distributions that habe been obtained based on the *nominal* Monte Carlo sample/generator. The nominal generator/sample (PYTHIA 8.230) *always* refers to the sample that was used to construct the response matrix to unfold the measured spectrum in data to particle/hadron level.

result in "variations" with respect to the nominal spectra. Usually, those weight-based variations are applied with the intention to change some specific conditions of the Monte Carlo sample like, for instance, the pile-up conditions, the underlying PDF, tunes of the parton shower, without the need to re-run the entire Monte Carlo simulation for all events – which is virtually impossible given the amount of time and resources needed to generate those samples. The weight-based variations will play an important role later in the context of the systematic uncertainties of the triple-differential multijet cross section measurement.

9. Event Selection and Event Reconstruction

This chapter finally turns the focus of attention of this report to the *actual* analysis by introducing the higher-level physical objects that are reconstructed on an event-on-event basis, which are then used in the measurements of the multi-differential multijet cross section(s). Since the relative production rate of multi*jet* events is measured with respect to a jet substructure observable that is sensitive to the quark- or gluon-like nature of the hadronic shower, the discussion is mostly centered around the reconstruction of jets as well as charged particles' trajectories. The tracks "associated"[1]/matched to calorimeter-based jets are of major importance within the context of this measurement as they maintain strong correlations between particle and detector level, allowing to unfold the jet substructure measurement in the first place. Furthermore, the phase space *volume* is defined on particle and detector level over which the cross section is measured.

This chapter proceeds as follows: in the first section (9.1), the central object of this measurement is introduced, i.e., the definition of the (normalized) triple-differential multijet cross section. Following, the next section (9.2) is dedicated to the object and event selection at *detector* and *particle/hadron* level. The reconstruction of jets as well as the matching of tracks and charged particles, respectively, is described along with the definition of quark- and gluon-initialized jets according to the PDGID code of the maximum-p_{T} constituent inside of the jet. For real collision events, the good-run certified data and the trigger selection strategy is introduced. Finally, the last section of this chapter (9.4) is supposed to provide a comparison between data and simulation (at particle and detector level) by means of a selected list of observables (jet substructure variables as well as global event shapes) and their distributions given the phase space volume and objects defined in the previous sections to estimate the data-to-MC agreement and to ensure that the reweighting scheme described in Chapter 8 works properly.

9.1 Definition of the Normalized Triple-Differential Multijet Cross Section

The analysis presented in this thesis includes several "independent"[2] measurements of the triple-differential multijet cross section in bins of a *jet substructure variable* $\mathcal{O}_{\mathrm{JSS}}(\in \{n_{\mathrm{count}}, z_{\mathrm{NN}}\})$ (counting-based observables as well as the score (=probability of the respective jet being gluon initialized) assigned by a NN) that is supposed to be sensitive to fundamental properties of QCD and hence allows to discriminate quark- from gluon-initiated jets – at least to a certain degree that results in mutual irreducible probability distributions (see Chapter 7). The motivation for measuring two classes (counting- and NN-based) of jet substructure variables instead of just one is driven by two mutually exclusive requirements: *firstly*, the variable under consideration should be "simple" in terms of interpretability, yet provide a decent sensitivity to the quark- or

[1] The term "associated" must be used carefully in order to avoid confusion with ATLAS' "track-associated" jet definition, which differs from the definition used in this analysis.

[2] Independently measured but, of course, correlated.

gluon-like nature of the hadronic shower to be mutually irreducible; *secondly*, the observable is supposed to be "as powerful as possible" in terms of quark-gluon discrimination, whereby decipherability/interpretability of the algorithm archiving this task is of minor importance.

For the first class of observables (counting-based), the cross section will be measured with respect to the track n_{trk} and Soft Drop (SD) multiplicity n_{SD}. The second measurement (NN-based), which focuses on the discrimination power between quark- and gluon-initialized jets at the price of interpretability, is done based on the prediction of a neural network that was trained similarly to a conventional quark-gluon tagger. The latter approach may seem a bit strange at first as this utilization deviates from the usual application of neural networks within the context of high-energy particle physics; however, embedded within the bigger scope of this analysis, it will become apparent that the direct measurement of a neural network score is indeed justified and well motivated.

Besides the jet substructure variable, the relative cross section is measured as a function of the *transverse momentum* p_{T}, as well as the relative *pseudorapidity* η_{rel} of the jets. The adjective, "relative", in the context of the rapidity means that not the actual pseujdorapidity value of the reconstructed jets determine the distribution they populate, but the relative orientation between the leading and the subleading jet in the detector instead. Therefore, the relative pseudorapidity is measured only in two *exclusive* bins: one of which corresponds to jets being more *central* in the detector, and a second region in *forward* direction. It is important to keep in mind that the prepositions "central" and "forward" do *not* corresponds to fixed regions or segments in the detector[3]! Whether a jet is labeled as being central or forward entirely depends on their relative orientation with respect to each other. In more concrete terms, given two jets J_1 and J_2 that both fall in the (fiducial) phase space volume, the first jet is "central" if $|\eta_{J_1}| < |\eta_{J_2}|$ which results in the second jet being labeled "forward"; *vice versa*, in case of $|\eta_{J_2}| < |\eta_{J_1}|$, jet number two is the more central one in this event.

Given the aforementioned observables (the JSS observable $\mathcal{O}_{\text{JSS}}$, the transverse momentum p_{T} and the (relative) rapidity η_{rel}), the triple-differential cross section for relative dijet production that was measured in this study is defined as

$$\frac{1}{\sigma}\frac{\mathrm{d}^3\sigma}{\mathrm{d}p_{\text{T}}\,\mathrm{d}\eta_{\text{rel}}\,\mathrm{d}\mathcal{O}_{\text{JSS}}} = \frac{1}{\sigma_{\text{tot}}}\frac{1}{\epsilon L}\frac{N^{\text{multijet}} - N^{\text{bkg}}}{\Delta p_{\text{T}}\,\Delta\eta_{\text{rel}}\,\Delta\mathcal{O}_{\text{JSS}}}, \tag{9.1}$$

whereby the numerator denotes the number of measured multijet events, N^{multijet}, in a given phase space bin (i.e., measured in finite bins of Δp_{T}, $\Delta\eta_{\text{rel}}$ and $\Delta\mathcal{O}_{\text{JSS}}$) posterior to all selection cuts and after the subtraction of contributions from all relevant background processes N^{bkg}; L being the integrated luminosity; ϵ representing the product of trigger efficiency, efficiency of the selection cuts etc.; σ is the total cross section given by the total number of multijet events integrated over the entire fiducial phase space volume.

Equation 9.1 further simplifies because the triple-differential multijet cross section is normalized by the total number of multijet event. This results in the cancellation of many systematic uncertainties to the degree of their correlation between the nominator and the denominator in the relative event rate according to Equation 9.1. For instance, the differential cross section as well as the total cross section are proportional to the luminosity and therefore *fully* correlated. Hence, the systematic uncertainty δL due to imprecise knowledge of the exact integrated luminosity drops out *completely* in this measurement. An alternative to the normalized cross section presented here would be the measurement of an absolute (event) rate that also provides information on the normalization, i.e., the absolute event yield expected in each phase space bin. Generally, absolute measurements contain more information and have an increased sensitivity to QCD effects – especially regarding measurements of the strong coupling –; however, they rely on the measured process to be well modeled in Monte Carlo. This is often not the case for the simulation of multijet events as they show substantial deviations in terms of the expected number of events – which will be shown later. For this reason, most resonance searches in which the contributions from multijet events are non-negligible (like, for instance,

[3] In terms of the absolute (pseudo-)rapidity, all jets that enter this measurement are central as only jets (tracks) with $\eta < 2.1(2.5)$ are used.

in many SUSY searches due to large missing energy because of the mismeasurement of jets and due to the appearance of neutrinos in heavy-flavor jets) rely on data-driven methods to estimate the number of events in a given signal region. Normalizing the cross section by the total event count basically gets rid of the "area" information and reduces the measurement(s) to the shape of the distributions, which are usually much better described in simulation.

Taking all of the above-mentioned into account, the normalized (inclusive) triple-differential (dijet) multijet cross section measured in this thesis can be written as

$$\frac{1}{\sigma}\frac{\mathrm{d}^3\sigma}{\mathrm{d}p_\mathrm{T}\,\mathrm{d}\eta_\mathrm{rel}\,\mathrm{d}\mathcal{O}_\mathrm{JSS}} = \frac{1}{N_\mathrm{tot}}\frac{N^\mathrm{multijet}}{\Delta p_\mathrm{T}\,\Delta\eta_\mathrm{rel}\,\Delta\mathcal{O}_\mathrm{JSS}}\,(1-r)\,, \tag{9.2}$$

with the ratio r being defined as $r := N^\mathrm{bkg}/N^\mathrm{multijet} \ll 1$. The event-yield ratio r corresponds to corrections due to processes that contribute to the measurement, but are not born from interactions of interest because final state leptons/electromagnetic showers are falsely reconstructed as hadronic jets (fakes). This correction is *very* small – even for the most dominant background process, i.e., Drell-Yan – and therefore does not affect the measurement given the degree of precision. Nonetheless, for the sake of completeness, a plausibility test of this statement based on a superficial analysis using a simplified model of the ATLAS detector will be provided later in this report.

Before closing this section for good, the attention is pointed once again to the calculation of the jet-related observables that define the phase space over which the triple-differential multijet cross section is measured. It is utterly important to keep in mind that the transverse momentum in Equation 9.2 is calculated based on the *PFlow-based* jet; otherwise, the jet energy could not be calibrated as no calibration scheme for track-based jets exist. The same is true for the relative pseudorapidity. The jet substructure variable, on the other hand, is

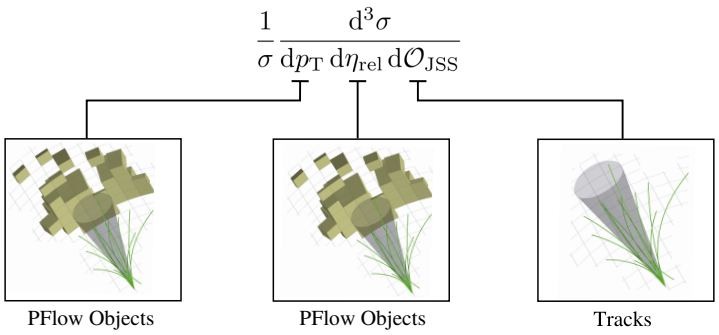


Fig. 9.1: Visual representation of the input to the different observables with respect to the triple-differential multijet cross section is measured. The transverse momentum as well as the relative pseudorapidity are calculated based on jets that have been reconstructed using PFlow objects as an input. The jet substructure variable, on the other hand, is calculated based on tracks that have been matched to the aforementioned calorimeter-based jets. (The event display in the figure shows an actual event that was recorded on August 1st, 2010 (event 118768330) with the CMS detector [25]).

calculated based on the set of *tracks* that has been *matched* to the calorimeter-based jets. This is necessary because the resolution of the calorimeter cells is far too low to measure a jet substructure observable as many particles might get absorbed in the very same cell of the calorimeter. However, for the unfolding to work, the correlation between particle and detector level must be sufficiently strong. The aforementioned inputs to the reconstruction of the cross section are summarized in Figure 9.1. Using tracks allows to unfold the JSS observable, however, it results in the measurement not being IRC-save (at the level of the jet substructure variable), i.e., the unfolded spectra presented in this study can unfortunately not be directly compared with analytic calculations in perturbative QCD (of course, it can still be compared

with Monte Carlo). Marginalizing the information of the JSS observable according to

$$\frac{1}{\sigma}\frac{\mathrm{d}^2\sigma}{\mathrm{d}p_\mathrm{T}\,\mathrm{d}\eta_\mathrm{rel}} = \frac{1}{\sigma}\int \mathrm{d}\mathcal{O}_\mathrm{JSS}\frac{\mathrm{d}^3\sigma}{\mathrm{d}p_\mathrm{T}\,\mathrm{d}\eta_\mathrm{rel}\,\mathrm{d}\mathcal{O}_\mathrm{JSS}}, \tag{9.3}$$

results in an IRC-safe double-differential multijet cross section. The cross section according to Equation 9.3 is IRC-safe because the transverse momentum as well as the relative pseudorapidity are calculated from a jet object that accounts for charged and neutral particles.

9.2 Object and Event Selection

This section lays the foundation of the entire analysis by introducing the basic event and objects selection valid for all measurements studied in this report. The starting point are the physical objects (according to Chapter 4) at detector level (9.2.1) that are reconstructed in data and simulation alike. It is followed by the reconstruction of the corresponding objects at particle level (9.2.2) that provide the truth-level information needed in order to construct the response matrix as well as to get the prior distributions in MC. Once the objects are introduced, the event selection, i.e., the phase space volume – over which the production rate is measured –, is defined for particle and detector level.

9.2.1 Detector-Level Objects

The only objects that directly enter the measurement of the triple-differential multijet cross sections presented in the thesis at hand are *hadronic* jets reconstructed from calorimeter cells, Inner Detector ($|\eta| < 2.5$) tracks with $p_\mathrm{T} > 500\,\mathrm{MeV}$ matched to the aforementioned PFLow jets, as well as – indirectly – primary vertices, which are not explicitly used in the analysis but for the purpose of the event selection solely (see track quality criteria listed below). The following paragraphs provide some background on the jets as well as the tracks used in this measurement at *detector level*.

Jets

The basic object used in this analysis are jets (as a phenomenological approach to QCD as described in Chapter 1), which are build/reconstructed from tracks that have been matched to PFlow jets at the electromagnetic energy scale using ATLAS' PFlow algorithm (see Chapter 4.4, § "PFlow objects" for a quick reminder of those objects) using the anti-k_t sequential reconstruction algorithm with a radius parameter of $R = 0.4$ (see Chapter 4.4.2). All reconstructed jets used in the measurement(s) are required to have a *minimum* transverse momentum of *at least* ($p_\mathrm{T}^\mathrm{min} =$)$250\,\mathrm{GeV}$. The reconstructed jets are further processed by applying the full jet energy calibration (see Chapter 4.4.3) to account for the non-compensating nature of the ATLAS calorimeter, dead/passive material in the detector etc. This calibration includes an area as well as a residual correction of the jets, mostly to mitigate contribution originating from pile-up; a correction of the jet energy response as a function of the transverse momentum and the pseudorapidity; followed by the global sequential calibration (which is specific to small-R jets and flattens the jet energy response for selected observables related to shower development inside the jets), flavor differences due to different partons that initialized the jets and punch-through effects; and, finally, an *in situ* calibration step, which is purely data-driven and accounts for potential biases due to the fact that all previous calibration correction factors are derived from Monte Carlo exclusively. After the full JES calibration has been applied, the *average* energy of the corrected jets is at the truth scale (see Chapter 4.4.3 for a reminder of the individual components of ATLAS' jet calibration scheme).

Tracks

Regarding the selection of tracks, *all* reconstructed *inner detector* tracks are required to pass the *tight* primary selection criteria, which are described in great detail in the following ATLAS note Ref. [197]. The tight Working Point (WP) for track selection consists of the following

individual track quality criteria that cut on the transverse momentum p_{T}, the pseudorapidity η, as well as the number of hits registered in one (or several) subsystem(s) of the inner detector:

- $p_{\mathrm{T}} > 500(400)\,\mathrm{MeV}$,
- $|\eta| < 2.5$,
- number of Pixel and SCT clusters on track ("silicon hits") ≥ 7,
- number of shared modules ≤ 1,
- number of silicon holes ≤ 2,
- number of pixel holes ≤ 1,

Tight WP (in addition to the Loose selection requirements)

- number of silicon hits ≥ 9 (if $|\eta| < 1.65$),
- number of silicon hits ≥ 11 (if $|\eta| \geq 1.65$),
- at least one hit in one of the two innermost pixel layers,
- *no* pixel holes.

Additionally to the items listed above, the event is required to contain *at least* one primary vertex with at least two valid(=tight) tracks pointing to it. In the original performance study according to Ref. [197], a lower cut on the track's momentum of 400 MeV was used. This baseline recommendation was derived using 2015 data at $\sqrt{s} = 13\,\mathrm{TeV}$. In order to account for the increased number of particles in the event due to increased pile-up interactions (which are likely to leave traces thought low-momentum tracks in the Inner Detector), the lower p_{T}-cut was increased to 500 MeV to alleviate the amount of combinatorics and to remove tracks not associated to the underlying hard scattering event. In the enumeration above, a "shared module" refers to a pixel module that has shared hits, whereby a shared hit is a hit associated with more than one reconstructed track.

Important to note, the tight selection was used mainly due to the last two items that require at least one hit in one of the two innermost pixel layers (IBL or subsequent pixel layers) as well as *no* pixel holes. The former criterion significantly improves the reconstruction of primary vertices and is therefore crucial for the reduction of pile-up. The latter cut excludes all tracks from the analysis whose intersections of the reconstructed trajectory with a sensitive detector element that not result in a registered hit. Basically, this removes all tracks with "missing hits", whereby certain regions and areas of sensors (such as, e.g., (known) inactive modules or edges on the silicon sensors) are released from this requirement.

The reason to prefer the *tight* over the (alternative) *loose* selection is the non-negligible contribution from fake tracks in case of the looser working point. Since this measurement of *jet substructure observables* is solely based on tracks associated to the PFlow jets, the number of fake tracks has to be as *small as possible* even at the cost of a reduced track reconstruction efficiency. Even more, the measured observables presented in this thesis have been chosen based on their (at least in principle) ability to distinguish between quark- and gluon-initialized jets to a certain degree that makes them mutual irreducible. Because the track multiplicity (number of tracks associated/matched to a jet) is among the most powerful discriminants for gluons and (light) quarks, most observables sensitive to quark- and gluon-related differences in the jet shapes are *strongly* correlated to the number of constituents inside the jet. Therefore, a large contribution from fake tracks would result in a bias of the entire measurement and/or in larger systematic uncertainties.

"Track-to-Jet Association"

The basic principle of matching tracks to jets build from calorimeter clusters by means of some matching scheme (ΔR or ghost-association) has been explained in detail in Chapter 4.4.4. In this analysis, the tracks are matched to the reconstructed, calorimeter-based jet using a simple ΔR-based criterion by linking all valid tracks $\{j_{\text{trk}}\}$ in the event to the calorimeter-based jet J with $\Delta R(j_{\text{trk}}, J) < 0.4$. This is sufficient as the calorimeter-based jets are reconstructed using the anti-k_t algorithm, which results in circular jet shapes (see Figure 4.2).

One aspect which is important to remember is that after the track-to-jet association has been performed, only the collection of tracks that have been matched to the PFlow jets in the event $\mathbb{J}_{\text{trk}} = \{j_{\text{trk}} \in \mathbb{R}^{(1,3)} \,|\, \Delta R(J, j_{\text{trk}}) < 0.4\}$ is used to compute the jet substructure variables $\mathcal{O}_{\text{JSS}}\left(\{j^i_{\text{trk}}\}_{i=1}^{n_{\text{trk}}}\right)$ with respect to which the differential multijet cross section is measured. This is the reason for the measurement of the JSS variable being sensitive to charged particles only. Since "track-based" jets (not to be confused with ATLAS' track-based jets that are build by directly using tracks as an input to the reconstruction algorithm) do not combine information from different subsystems of the detecto, there is no risk of double counting energies and therefore no need for a subtraction scheme. Furthermore, it should be noted that the PFLow jets (not the track-based/matched ones!) are used as a basis for the multijet event selection of the event, as well as the input for the application of the prescale factor depending on the p_{T} of the jet for the designated higher-level single jet trigger, which are defined based on the transverse momentum of jets reconstructed from PFlow objects. The calculation of the transverse momentum (for the differential cross section according to Equation 9.2) is based on the PFlow jet too and therefore contains contributions from both, charged and neutral components. This is important because the variation of the jet's momentum within the uncertainty of the jet energy scale *must* be applied to calorimeter-based jets as no corresponding calibration scheme for track-based jets exists.

Morover, it ensures that the extracted final quark/gluon fractions are IRC-safe irregardless of the measurement of the JSS observable being IRC-*un*safe. This point is very important in order to use the extracted fractions in, for instance, a JSS-based measurement of the strong coupling constant α_{S} as will be explained later in the "conclusions and future research suggestions" at the end of this report.

9.2.2 Particle-Level Objects

In order to construct the response matrix needed to unfold the measurement at detector level to particle level, truth information provided by the nominal sample (Pythia 8.230) must be used that relates an object folded with the ATLAS detector to the corresponding particle level one. Simply speaking, the very same reconstruction algorithm is executed twice: one time for the truth-level object and once again for the detector level jet that has been *matched* to the respective objects at particle level. The result is a correlation between the two, which is the basic information provided by the response matrix (see Chapter 5.1.1).

This brief Section serves the same purpose as the previous one but for particle level objects instead.

Jets

Jets at particle level are reconstructed based on stable, final state particles ($\tau > 10\,\text{mm}$) with neutrinos and muons being excluded, using the very same jet clustering algorithm(=anti-k_t) as for the detector-level objects in Section 9.2.1. Neutrinos are excluded as they only interact weakly and hence do not leave a measurable track or energy deposition in the detector volume; muons, on the other hand, are minimum ionizing particles with a dedicated spectrometer in the outermost layer of the measurement device (see Chapter 4.2). The result of the jet reconstruction is a truth/particle-level jet, associated with a set of truth-particles with known PDGID based on the Monte Carlo particle numbering scheme according to Ref. [198].

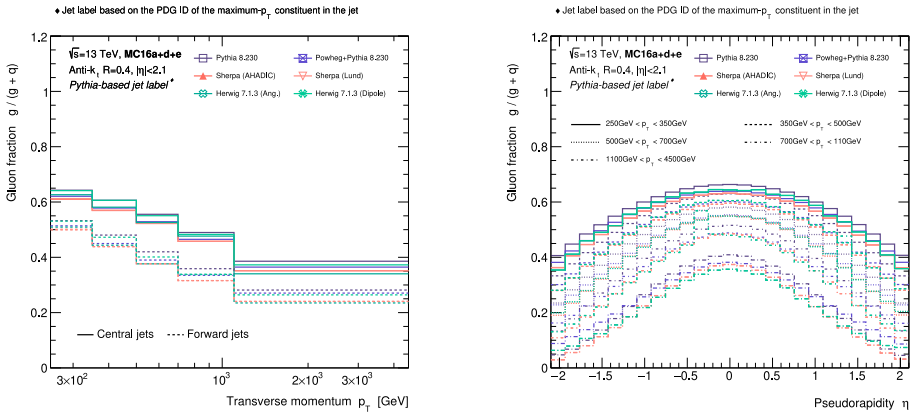
Jet-by-Jet Label

The PDGID is the basic information commonly used to assign a "flavor" (e.g., *light* quark or gluon) to the jet. This is done by directly assigning the PDGID of the *highest-p_{T} constituent* to the jet as a whole, which is conceptually problematic as has been explained in Chapter 7. For this analysis, only two categories of jet labels are defined based on whether the jet was initialized by a light quark (u(p), d(own), c(harm), s(trange), b(ottom/eauty)) or a g(luon). More precisely, a jet J is referred to as a "gluon jet" if the PDGID of the highest-p_{T} constituent $j \in J$ is '21' (PDGID of the gluon); otherwise, it is labeled as a (light) quark jet according to

$$f_J\left(\{j_i\}_{i=1}^{|J|}\right) = \begin{cases} \text{g(luon)} & \text{if } \mathrm{ID}_{\mathrm{PDG}}\left(\max\limits_{j\in J}\{p_{\mathrm{T},j}\}\right) = 21, \\ \text{q(uark)} & \text{else} \end{cases} . \tag{9.4}$$

The resulting fraction of gluon jets (according to the aforementioned jet label) in several p_{T} bins for central and forward jets is shown in Figure 9.1 for all generators for the Monte Carlo campaigns 16a, 16d and 16e, which corresponds to Run 2 data-taking conditions in simulation. Furthermore, the numerical values of the fraction of gluon-initialized jets is shown in Table 9.1 for the baseline Monte Carlo event generator, i.e., Pythia 8.230. As explained previously in Chapter 7.1.1, this definition of the jet flavor is (strictly speaking) "ill"-defined and certainly *not* collinear save as splitting the highest-p_{T} constituent into two collinear particles (for example via $g \to q\bar{q}$) may result in a different flavor being assigned to the jet. Nonetheless, this definition is the one conventionally used to assign a flavor jet-on-jet based for each jet in an event. This results in a labeled data set that was used, e.g., to train the neural network(s) presented in this report in a supervised fashion.

As expected from the multijet cross section, the fraction of quark-initialized jets continuously



(a) p_{T} *(b)* η

Plot 9.1: Fraction of gluons for several p_{T} bins subdivided into central and forward jets for all samples used in this analysis. The description "pythia-labeled jet flavor" on the canvas means that the PDGID of the leading p_{T} constituent in the jet was used to determine the flavor of the jet.

increases towards higher transverse *momenta*, which directly follows from the evolution of the proton's PDF and the increasing correlation of the parton flavor between initial and final state for increasing p_{T}. Furthermore, the central category of jets is generally "gluon enriched" compared to the forward region. This is a direct consequence of the underlying gluon PDF too and the observation that central jets are usually accompanied by higher transverse momentum and therefore larger momentum fractions according to the (LO) relation $x \sim p_{\mathrm{T}}/\sqrt{s}$.

	Transverse momentum p_{T} [GeV]				
	[250, 350)	[350, 500)	[500, 750)	[750, 1100)	[1100, 4500)
Rapidity category η_{rel}	Fraction of gluons f_g (as predicted by PYTHIA 8.230)				
Central	0.643	0.606	0.556	0.489	0.386
Forward	0.532	0.481	0.419	0.359	0.282
Combined	0.588	0.544	0.488	0.424	0.334

Table 9.1: Numerical values of the fraction of gluon-initialized jets as predicted by the baselien Monte Carlo event generator in several transverse momentum bins.

Charged Constituents

At truth level, there are generally no tracks as they are obtained from hits in the pixel sensor; instead, the equivalent to tracks at particle level are the stable, final state charged particles (the charge of the particle is part of the truth-level information available for each particle in the event record) related to the corresponding truth jet. Like the tracks, all *charged* constituents in the jet are required to have a transverse momentum above 500 MeV.

Jet-to-Jet Matching

For the purpose of constructing the response matrix that encodes the correlations between objects at truth level and those folded with detector effects (detector level), jets reconstructed at particle and detector level must be *uniquely* matched to each other. This analysis utilizes a very simple method for matching jets my means of the $\Delta R^2 = \Delta\eta^2 - \Delta\phi^2$ metric that defines a distance between the center of the jets in η–ϕ-space. Two jets J_i^{DL} and J_j^{PL} at detector and particle level, respectively, are matched if they are most close together in terms of ΔR. This is the most simple jet matching scheme, yet it is highly efficient as the respective selection (which will be introduced in an upcoming section) results in events with mostly back-to-back jets, which significantly reduces the ambiguity in the jet-to-jet assignment.

9.2.3 Good-Run Certified Data

At ATLAS, there is a strict hierarchy concerning the organization and categorization of data. In this ladder, the smallest building block is a single *pp-collision event*, which is saved at a rate of approximately 1 kHz. An ensemble of roughly 10^5 event (corresponding to something like one minute of data taking) are collected into a group referred to as one luminosity block. This luminosity block forms the smallest unit for which the luminosity is known, i.e., the luminosity is assumed to remain constant for all events in this group. The reason for this assumption is that the measurements of the *inelastic* proton cross section at the LUCID detector (see Chapter 3.3.7 for a quick review of the measurement of the luminosity at the LHC) are averaged over a period of time that corresponds exactly to one *luminosity block*. Luminosity blocks are labeled by a quality stamp that indicates whether the given block of data is valid to use in physics analyses by tagging it according to a manichaeism-like good-bad dichotomy as "bad" or "good". This information is stored in the so-called *Good Runs List* (GRL), which is applied to the event selection prior to all other cuts. A collection of luminosity blocks is further combined to what is called a full *run*. It consists of roughly 1000 luminosity blocks and therefore corresponds to several hours of data taking. Interesting to know: a single run is defined such that it is in the region of roughly one LHC fill. All runs are uniquely identified by an integer value; this will

become important in the next section as it is used to assign the correct lowest *un*prescaled higher-level single jet trigger to the jets in the event. On the next meta level, approximately five to ten runs are combined to build a *sub-period*. This intermediate step was introduced to provide a block of data for which the condition of data-taking in terms of the experimental setup remain more or less constant in time. This point will be discussed once again in the context on the "year-to-year" uncertainty that checks the closure of the distributions between different periods of data-taking – this aspect is not irrelevant for this analysis because tracks are used to reconstruct the JSS observable that are sensitive to the alignment of the detector. The penultimate object is what is know as a full *period*. A period is build from ten to five consecutive sub-periods with similar conditions. This rule is flexible; if the machine was shut down, e.g., for technical maintenance, a new period starts since the conditions of operation may have changed[4]. Last but not least, periods are combined to a *LHC Run*. One LHC Run of data-taking is equivalent to the period of time between individual *long* shut downs of the LHC for the purpose of detector upgrades and updates, e.g., Run 1, 2 (this measurement), and 3 (which is still ongoing during the time of writing this thesis).

9.2.4 Trigger Selection

As described in Chapter 3.1, throughout Run 2 conditions, the LHC operated with a bunch crossing rate of approximately 40 MHz, which corresponds to a separation in time between two proton bunches of 25 ns. Given the current technical limitations, it is not possible (and certainly will not be in the foreseeable future) to read out the detector, process all the information and store the data to non-volatile memory at a frequency of 40 MHz; hence, the rate must be *significantly* reduced to a manageable size which is currently in the order of roughly $\mathcal{O}(1\,\text{kHz})$. To put things into conceivable perspective: this means that only 0.025 ‰ out of all occurring pp collision events are kept for further analysis.

Aside from the rate reduction, the trigger entries are designed to increase sensitivity to phenomena considered interesting for physics analyses. Just "throwing a dice" for each event in order to decide whether to keep it or not would basically result in "zero" sensitivity to interesting physical phenomena. Usually, this involves searches for new, heavy particles; hence, the triggers are supposed preferentially select events within regions in phase space populated by those rare, low cross section occurrences. In case of jets, the energy and the p_{T} spectrum is steeply falling, which means the probability for obtaining an event with low-energetic jets is very high. This results in the low-threshold single jet triggers constantly firing. In order to reduce the rate of jets within a low-energy region (and simultaneously increasing the bandwidth of events with high-energetic jets), most single jet triggers are *prescaled*. Consequently, each single jet trigger (below a certain p_{T} threshold) comes with a so-called Pre-Scale factor $N_{\text{PS}} \in \mathbb{N}$ that *randomly* saves one every N_{PS} (meaning "1-in-a-N_{PS}") events passing the respective trigger threshold while discarding the rest for good. This number is *not* fixed in time, but it depends on the run condition at the time of data taking and the available bandwidth. Of course, disposing events reduces statistical precision and significance, which in turn reduces sensitivity; therefore, full sensitivity is limited to the high-p_{T} regime at which the triggers are *un*prescaled.

Each event that makes it to permanent storage passed a chain of triggers that provide a rough assignment/classification of collision events to different, not mutually exclusive categories of physical objects and interactions. Therefore, the very first step of each analysis is the selection of relevant higher-level (jet) triggers to sort/filter out the class of interesting events that meet the requirement of the event selection and for that reason are considered in the analysis.

Single-Jet Triggers and Trigger Fractions

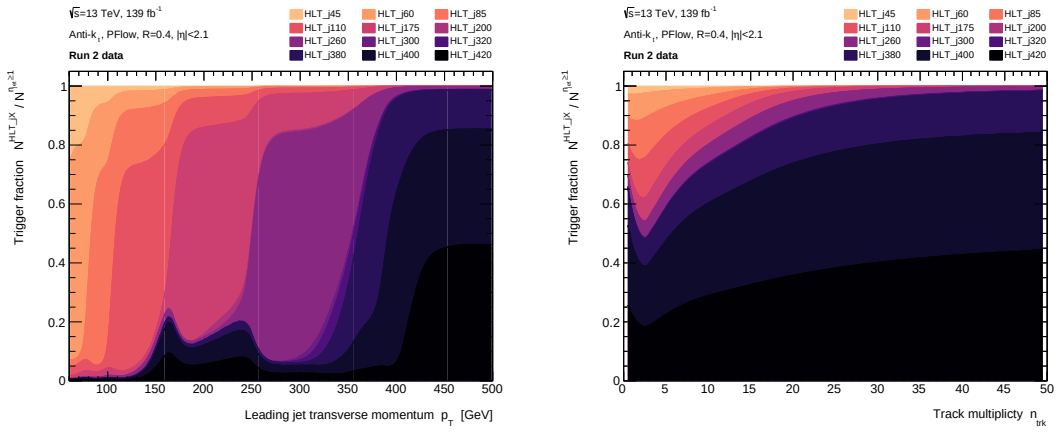
The analysis presented in this thesis is all about jets; thus, the relevant events must have been passed *at least* one higher-level single *jet* trigger. The corresponding entry in the trigger menu

[4] This is particular important for track-based measurements due to the already mentioned sensitivity to the alignment of the detector that may have slightly changed over time.

is commonly denoted as "HLT_jX", whereby "X" symbolizes the (approximate) transverse momentum threshold from which on the respective trigger operates close to full efficiency. This number only gives an approximate value for the lower trigger threshold because most triggers are not fully efficient at this momentum. Hence, the determination of a list of mutually exclusive bins in transverse momentum p_{T} in which the selected triggers are *fully* efficient will be one part of this section. This is important in order to obtain the correct cross section; otherwise, the number of measured events might underestimate the one expected from the event rate for the respective process.

This section provides a detailed step-by-step guidance through the entire trigger selection and reweighting/scaling process.

To get a first impression, it is instructive to take a look at the relevant single jet triggers



(a) Transverse momentum p_{T} *(b)* Track multiplicity n_{trk}

Plot 9.2: Percentage proportion of several higher-level single jet triggers as a function of the leading jet transverse momentum p_{T} (*left*) and the track multiplicity n_{trk} (*right*).

that come into question for this measurement as well as their fraction with respect to all triggers for a given p_{T} of the leading-p_{T} jet in the event (the trigger selection in this analysis is *always* based on the leading jet's p_{T}[5]). The fraction $f^{\mathrm{HLT\_jX}}$ of one particular trigger is simply defined as

$$f^{\mathrm{HLT\_jX}} = \frac{N^{\mathrm{HLT\_jX}}}{N^{n_{\mathrm{jet}} \geq 1}}, \tag{9.5}$$

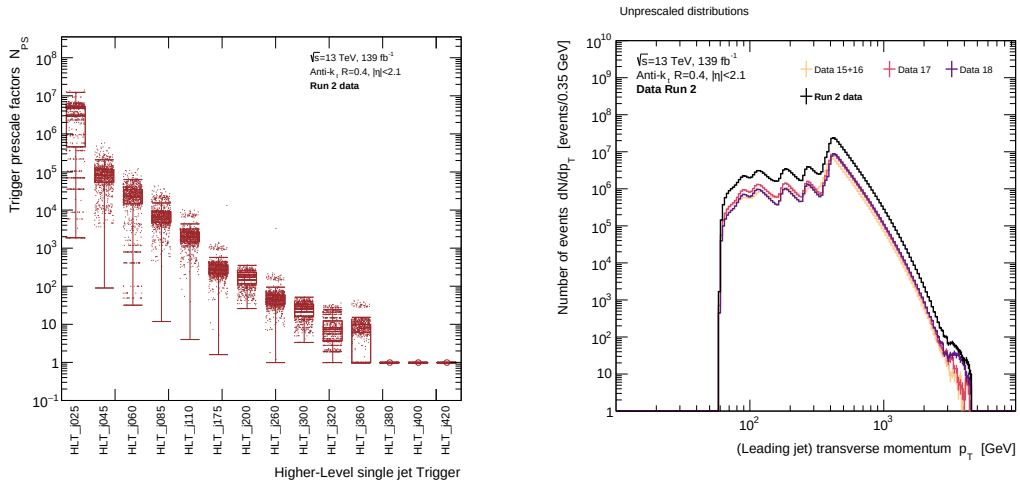
whereby $N^{n_{\mathrm{jet}} \geq 1}$ denotes the total number of events that passed *at least* one single jet trigger. By definition, the following normalization condition holds: $\sum_{\mathrm{X}} f^{\mathrm{HLT\_jX}} = 1$ (at least one trigger must be passed). The fraction of single jet triggers *versus* the transverse momentum of the leading-p_{T} jet is shown in Figure 9.2a. Figure 9.2b represents exactly the same information, but with a different representation in terms of the track multiplicity n_{trk}, i.e., the reconstructed number of charged "constituents" inside of the (leading) jet. This alternative presentation is justified since the main focus of the analysis is on jet substructure, hence, it must be ensured that the reweighted events results in meaningful distributions also in terms of internal structure of the jets. As can be seen, with increasing track multiplicity, the faction of high-p_{T} threshold triggers increases. This is expected because the number of particles increases as a function of the transverse momentum of the jet.

[5]Sometimes, the trigger threshold is defined by means of the scalar sum $H_{\mathrm{T}} = p_{\mathrm{T}}^{(1)} + p_{\mathrm{T}}^{(2)}$ of the transverse momenta of the first two leading jets in the event.

Trigger Efficiency

To gain a better feeling for the different higher-level single jet triggers, Figure 9.3a shows the prescale factors for the respective trigger that comes into question. As mentioned earlier, the prescale factors are *not* fixed, but vary in time in order to adjust to current run conditions of the experiment. The range of variation of one prescale factor is indicated by the shaded area in Figure 9.3. As can be seen from Figure 9.3a, the variation of one prescale factor is mostly constrained to the same order of magnitude, but may deviate significantly from time-to-time as indicated by the anchors and wicks of the candles. The three triggers HLT_j380, HLT_j400 and HLT_j420 are the lowest *unprescaled* triggers for the data taking years 2015, 2016; 2017, and 2018, respectively, which is why the corresponding entries in Figure 9.3a are exactly at unity, i.e., all events are triggered.

If no prescale factor is applied, the reconstructed distribution of the leading jet transverse momentum looks like the one in Figure 9.3b (mind the logarithmic scale); recovering the original, smoothly falling spectrum by applying the correct prescale factors is the objective of the reweighting procedure according to the trigger prescale factors. The first task is to



(a) Prescale factors N_{SF} for different higher-level single jet triggers. The shaded area enclosing the mean value corresponds to the total variation of the scale factors of the respective trigger.

(b) Unweighted (no trigger scale factors) spectrum of the transverse momentum for the leading p_{T} jet in the event. The different slices correspond to different years of data taking.

Plot 9.3: Prescale factors N_{PS} (*left*) and unweighted transverse momentum p_{T} distribution of the leading jet (*right*) for different higher-level single jet triggers. The last three trigger paths in Figure 9.3a are unprescaled.

define a list of higher-level single jet triggers and to ensure that they are fully efficient within their assigned *exclusive* region in p_{T}; otherwise the number of events in this p_{T} regime will be systematically underestimated. For this, the trigger efficiency of a given trigger must be measured or already known.

There are several ways to determine the trigger efficiency. One way is to simulate the full trigger chain using Monte Carlo generated events. This method works well, but here, the trigger efficiency is measured directly in data using a reference trigger.

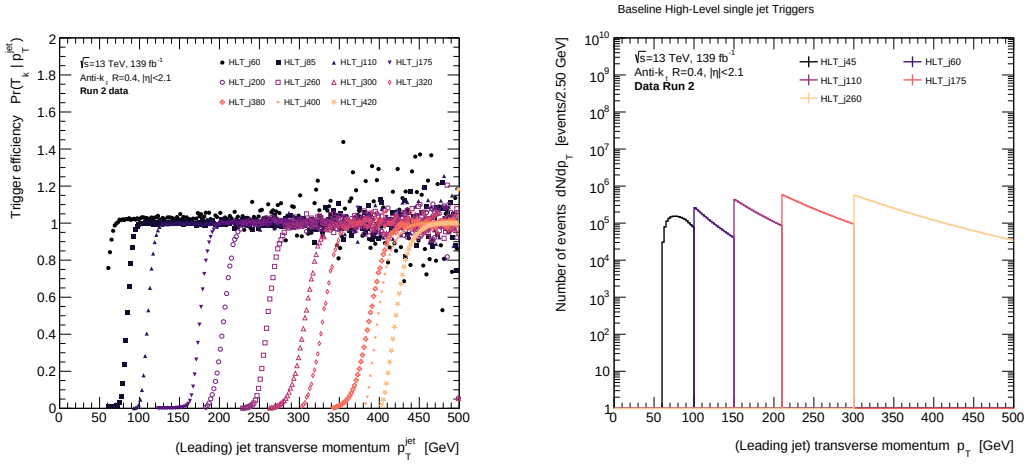
Doing an efficiency measurement in data always requires some kind of a reference – to measure the efficiency curve of a given trigger, the reference trigger can, for instance, be a lower threshold one. let $P(T_k|p_{\mathrm{T}})$ denote the probability for a trigger T_k to fire given a jet with transverse momentum p_{T}. Now, given some reference trigger T_{k-1}, the trigger efficiency $P(T_k|p_{\mathrm{T}})$ is given by

$$P(T_k|p_{\mathrm{T}}) = \frac{P(T_k \cap T_{k-1}|p_{\mathrm{T}})}{P(T_{k-1}|p_{\mathrm{T}})}, \tag{9.6}$$

whereby $P(T_k \cap T_{k-1}|p_\mathrm{T})$ is the probability for a jet with p_T to pass *both* triggers. Naively implementing this relation, however, comes with the issue of limited statistics because the prescale factor of the reference trigger is higher by definition. To avoid this problem, Equation 9.6 can be rewritten using Bayes' Theorem

$$P(p_\mathrm{T}|T_k) \propto P(T_k|p_\mathrm{T})P(p_\mathrm{T}), \tag{9.7}$$

with the normalization $P(T_k)$ being ignored because the position of the p_T threshold (which is the relevant information) does not depend on it. In Equation 9.7, $P(p_\mathrm{T})$ is the *undistorted* p_T spectrum that can be estimated(!) from a trigger with a higher prescale factor. The efficiency curve of trigger T_k is then given by the ratio of the distributions for $P(p_\mathrm{T})$ and $P(p_\mathrm{T}|T_k)$. In Equation 9.7, the normalization of the distributions was left undetermined; however, it can simply be computed by demanding the probability distributions to be properly normalized. The resulting efficiency curves based on the aforementioned prescription for a preselection (for



(a) Trigger efficiency

(b) Trigger slices

Plot 9.4: Efficiency $P(T_k|p_\mathrm{T})$ for a preselected list of higher-level single jet triggers as a function of p_T (*left*) and the corresponding p_T slices with no prescale factors applied (*right*).

reasons of clarity) of single jet triggers is shown in Figure 9.4. All curves show the expected characteristics: (almost) no efficiency for jets with a transverse momentum below the designed threshold, followed by a steep turn-on curve and finally saturation where the trigger is fully efficient – values above one are a consequence of limited statistics and normalization and can safely be ignored. The first curve in Figure 9.4b, corresponding to HLT_j45, was included to show the effect of a trigger not being fully efficient.

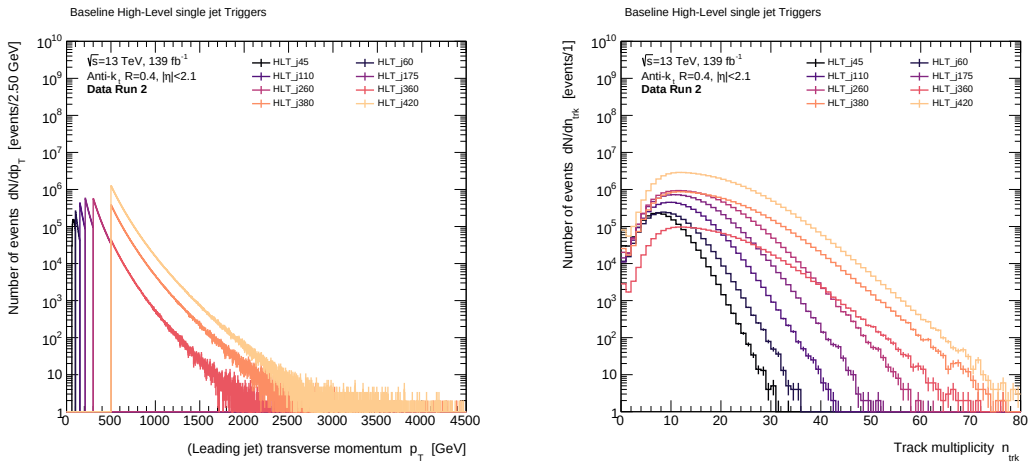
Once the trigger efficiencies of the triggers in question have been measured, all that is left to do is to read off directly from Figure 9.4a the trigger thresholds for which the selected triggers are fully efficient. The inferred exclusive p_T bins as well as the corresponding higher-level single jet trigger are summarized in Table 9.2. Above a transverse momentum of 500 GeV, all single jet triggers are *un*presclaed. However, the lowest unprescaled trigger is not the same for all years of data-taking, which must be taken into account during reweighting. Furthermore, the trigger efficiency shows a weak dependence on the pseudorapidity of the jet too, which is why two categories of higher-level jet triggers exist – one for $|\eta| < 2.1$ (central region) and another one for $|\eta| \geq 2.1$ (forward region), with the latter one showing a small shift of the threshold towards lower p_T due to increasing material budget in the forward region of the detector. Since all jets are limited to the inner part of the detector (tracker region); hence, forward triggers are excluded or rather without consequences in this analysis.

Figure 9.5 shows the result of the trigger selection for the transverse momentum (9.5a) and

Exclusive p_{T} range [GeV]	prescaled HL single jet trigger
$[100, 150)^{\dagger}$	HLT_j060
$[150, 210)^{\dagger}$	HLT_j110
$[210, 300)$	HLT_j175
$[300, 500)$	HLT_j260
$[500, \infty)$	unprescaled

Table 9.2: Exclusive p_{T} ranges associated with a higher-level single jet trigger. Jets with a transverse momentum below 250 GeV are not considered since they are exclude by the definition of the phase space selection. † Those trigger ranges are excluded due to the lower momentum cut.

the track multiplicity (9.5b). In case of the transverse momentum, the exclusive trigger ranges can be clearly seen. The three curves for $p_{\mathrm{T}} > 500\,\mathrm{GeV}$ correspond to the different data-taking campaigns 2015/16, 2017 and 2018 that differ with respect to their lowest unprescaled trigger. For the track multiplicity in Figure 9.5b, the distributions overlap of course, but an increasing average number of tracks $\langle n_{\mathrm{trk}} \rangle$ per jet for higher threshold triggers (corresponding to larger p_{T}) can be observed as expected. With the jet trigger being selected, all that is left to do is to



(a) Transverse momentum p_{T} *(b)* Track multiplicity n_{trk}

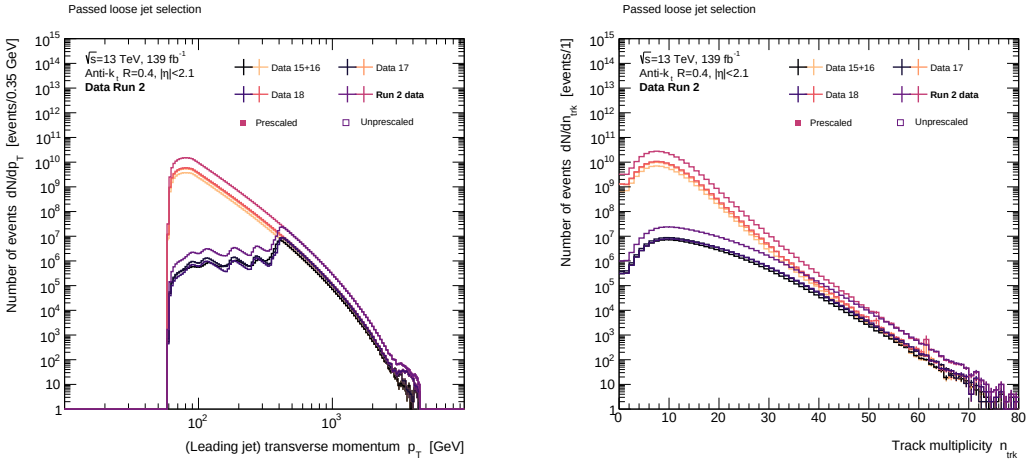
Plot 9.5: The transverse momentum p_{T} (*left*) and the track multiplicity n_{trk} (*right*) distribution for all exclusive trigger ranges with *no* prescale factors being applied.

apply the correct trigger prescale factor depending on the p_{T} of the jet. The final reweighted transverse momentum and track multiplicity distribution after selection is shown in Figure 9.6 for the transverse momentum (9.5a) and track multiplicity (9.5b).

9.2.5 Event Selection

All measurements presented in this thesis are based on multijet events recorded with the ATLAS detector. More precisely, the actual analysis makes use only of the first two leading-p_{T} jets in an event, each one populating a distribution. The measurements are *inclusive* in the sense that all events with *at least* two jets (meaning 2jet + N) that pass the basic event selection are used to determine the differential cross section.

Prior to the basic event selection that defines the (fiducial/total) phase space volume, the event must satisfy some standard cleaning criteria to ensure good data quality. As already



(a) Transverse momentum p_{T} *(b)* Track multiplicity n_{trk}

Plot 9.6: The transverse momentum p_{T} (*left*) and the track multiplicity n_{trk} (*right*) distribution *before* and *after* the prescale factors being applied.

mentioned in Section 9.2.3, the event has to be an element of a valid luminosity block, i.e., it *must* pass the GRL. This cleaning step, of course, is only applied to data since there is no corresponding variable/flag for Monte Carlo. In case of Monte Carlo, basic event cleaning is done by explicitly removing simulated events labeled as "bad" from the xAOD[6] are discarded for good. For both, data and Monte Carlo, the event is required to have *at least* one primary vertex reconstructed $n_{\mathrm{PV}} \geq 1$ with *at least* two valid, tight tracks consistent with it $n_{\mathrm{trk}} \geq 2$.

Jet Working Points

Besides the basic event cleaning cuts mentioned in the previous paragraph, the jet candidates must fulfill the ATLAS quality selection criteria. In case of small-R anti-k_t jets, there are two jet selection working points available: a *loose* selection that provides an efficiency of approximately 99.5 % for $p_{\mathrm{T}} > 20\,\mathrm{GeV}$ and 99.99 % for $p_{\mathrm{T}} > 100\,\mathrm{GeV}$ respectively (which is the standard cleaning for most ATLAS analyses); and a *tight* working point aiming for analyses sensitive to non-collision backgrounds with a jet selection efficiency of 95.0 % for a transverse momentum of $p_{\mathrm{T}} > 20\,\mathrm{GeV}$ and 99.5 % for $p_{\mathrm{T}} > 100\,\mathrm{GeV}$. Both selection working points are supposed to identify fake jets (for instance, an electron that was misidentified as a jet) and hence increase purity of the jet sample. Since the fake rate of jets significantly depends on the calorimeter noise, the characteristic signal shape is among the most powerful discriminators to identify fake jets. To this end, the actual measured shape in the calorimeter is compared to the expected shape from simulation; the quadratic difference. Additionally, the energy fractions in the different calorimeters f_{EM} and f_{HEC} has some discrimination power between fake and real jets and hence is used in the selection[7].

The loose working point is designed to keep as much jets as possible, i.e., operating at a high efficiency while maintaining a high fake jet rejection. A jet that fails *at least* one of the cuts listed below is referred to as a "BadLoose" jet:

1. $f_{\mathrm{HEC}} > 0.5$ and $|f_Q^{\mathrm{HEC}}| > 0.5$ and $\langle Q \rangle > 0.8$,
2. $|E_{\mathrm{neg}}| > 60\,\mathrm{GeV}$,

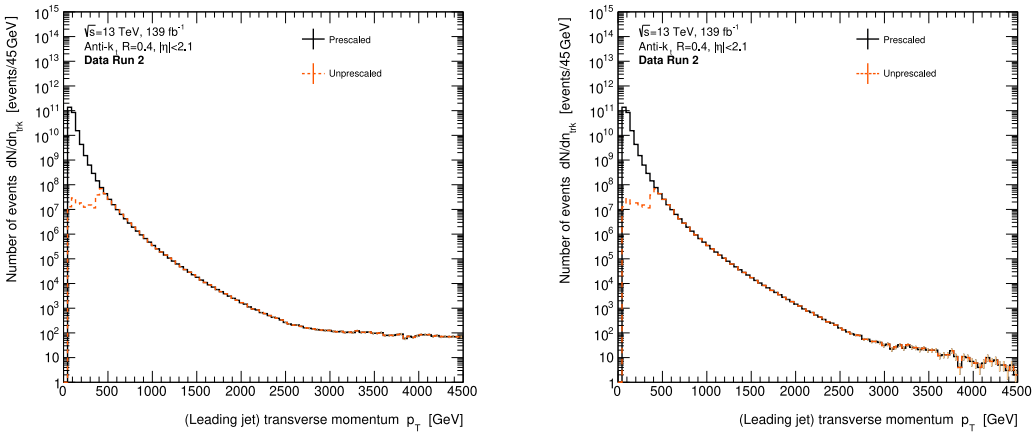
[6] The Analysis Object Data (xAOD) is a summary of the reconstructed event, which can be read by the open source analysis framework ROOT [199] as well as ATLAS' software framework (Athena [200]). As such, it is part of CERN's Event Data Model (EDM).

[7] An excellent in-dept study regarding the selection of jets at the ATLAS experiment can be found in the following reference Ref. [201].

3. $f_{\mathrm{EM}} > 0.05$ and $f_Q^{\mathrm{LAr}} > 0.8$ and $\langle Q \rangle > 0.8$ and $|\eta| < 2.8$,
4. $f_{\mathrm{max}} > 0.99$ and $|\eta| < 2.0$,
5. $f_{\mathrm{EM}} < 0.05$ and $f_{\mathrm{cf}} < 0.05$ and $|\eta| < 2.0$,
6. $f_{\mathrm{EF}} < 0.05$ and $|\eta| \geq 2.0$.

In the enumeration above, $|E_{\mathrm{neg}}|$ is the absolute value of "negative energy deposition" in the calorimeter. This may happen due sporadic noise in the calorimeter, which may result in large reconstructed energies; f_{max} denotes the maximum energy fraction in one event in any of the calorimeter layers; f_Q^{LAr} is the fraction of energy in the LAr calorimeter for jet with poor signal shape quality. The first two items are supposed to reject jets due to sporadic high noise occurring in the hadronic calorimeter of the detector. The third criterion rejects jet reconstructed due coherent noise or pathological cells in the electromagnetic part of the calorimeter. Finally, the remaining cuts aim to reduce contributions from cosmic muons, beam-induced effects and malfunction hardware. The effect of the loose jet selection described above is demonstrated by means of the p_{T} spectrum in Figure 9.7. The plateau/tail in Figure9.7a is due to the already mentioned detector noise; applying a loose jet selection as in Figure 9.7b, noise-induced calorimeter jets are efficiently rejected resulting in a smoothly, monotonically falling spectrum.

The loose selection can be extended by one additional criterion to obtain a *tight* selection



(a) "BadLoose" jets

(b) Loose jet selection applied.

Plot 9.7: Transverse momentum for "BadLoose" jets (*left*) and jets that pass the loose selection (*right*) for the prescaled and unprescaled distributions.

working point for jets. A jet that fails the tight selection is referred to as being "BadTight". The additional criterion for the tight selection includes a ratio between the fraction of charged constituents of the jet, f_{ch}, and the fraction of largest energy deposit f_{max}. A jet passes the *tight* selection, if it is not a "BadLoose" jet and additionally satisfies $f_{\mathrm{ch}}/f_{\mathrm{max}} \geq 0.1$ for $|\eta| < 2.5$, whereby the pseudorapidity cut ensures that the jet is within the region of the inner detector.

Both cross section measurements presented in this thesis have been done for the loose as well as for the tight jet selection. The tight selection, however, was found to result in larger uncertainties. Therefore, if not explicitly mentioned otherwise, the loose selection is applied to all jets.

Summary and Phase Space Volume

The event selection can be roughly divided into two parts for particle and detector level alike: *firstly*, an event cleaning is applied that removes *events* that do not pass certain data quality

flags; *secondly*, cuts applied on *the level of individual jets* that define requirements on the jets used in the analysis.

All event cleaning cuts have already been introduced in the previous Section; therefore, they should only be briefly mentioned here for clarity in bullet point form once more. All events used for the analysis *must* pass the GRL (see Section 9.2.3) with at least (inclusive measurement) *two jets* being reconstructed in the event. If the event meets the basic data quality criteria, the next step is to ensure that there is a primary vertex being present in the event with *at least* two tracks being compatible with it[8]. The latter requirement, of course, accounts for the fact that charged particles in multijet events are studied; hence, at least two tracks passing the loose/tight track selection (see Section 9.2.1) should be present in the event. Last but not least, the two leading-p_{T} jet must not be tagged as being "BadLoose".

If the event passes the basic event selection criteria, the attention focuses on the jet candidates itself. The two leading-p_{T} jets used in the analysis must be within the acceptance region of the inner detector, taking into account the jet radius $R = 0.4$ to ensure that all clustered charged particles are covered by the tracker as well. This results in a pseudorapidity cut of $|\eta| < 2.1 (= 2.5 - R(= 0.4))$ for *all* calorimeter-based jets used in the measurement. Especially at low p_{T}, there is a non-negligible contribution coming from pile-up events. Those pile-up jets are usually soft compared to the hard scatter which is accounted for by requiring the two leading jets to fulfill $p_{\mathrm{T}}^{\mathrm{lead}}/p_{\mathrm{T}}^{\mathrm{sublead}} < 1.5$ (see Ref. [202] for a motivation of this selection cut). Besides the suppression of pile-up, this cut acts a kind of a isolation criterion that results in highly isolated jets on average that *"simplifies the interpretation of the final state in terms of a* $2 \to 2$ *scattering process*" [202]. Finally, with the goal in mind to perform a measurement that is sensitive to quark-gluon difference, the measurement is conducted for central and forward jets simultaneously because the two categories are in each case gluon and quark enriched, respectively, as motivated in Section 9.2.2.

The entire event and object selection for all measurements presented in this report is

	Particle Level	*Detector Level*
Pile-up	Removed by truth information	"Removed" by selecting primary vertex
Jet Algorithm	Anti-k_t algorithm with $R = 0.4$	
Jet Requirements	$\lvert\eta\rvert < 2.1$ and $p_{\mathrm{T}} > 250\,\mathrm{GeV}$	
Jet Constituents	Only stable particles with $\tau > 10\,\mathrm{mm}$ excluding neutrinos and muons	ATLAS' PFlow objects
Input	Charged truth level constituents with $p_{\mathrm{T}} > 500\,\mathrm{MeV}$	Tracks with $p_{\mathrm{T}} > 500\,\mathrm{MeV}$, ΔR -associated to the reconstructed jet
Event Selection	No "bad" xAOD flags	GRL is passed; no "BadLoose" jets
	At least one charged constituent	At least one track
	At least two jets satisfying $p_{\mathrm{T}}^{\mathrm{lead}}/p_{\mathrm{T}}^{\mathrm{sublead}} < 1.5$	
Jet Selection	Two leading-p_{T} jets subdivided into central and forward categories	

Table 9.3: Summary of the event selection as well the input objects used for the measurement of the triple-differential multijet cross sections presented in this report.

summarized clearly in a condensed form in Table 9.3. The "input" group in Table 9.3 refers to the input that is used to compute the jet substructure variables in bins of which the triple-differential multijet cross section is measured – while the PFlow jets are only made use of for the purpose of event selection, to construct "track-associated" jets, as well as the transverse momentum p_{T} and the (relative) pseudorapidity.

In essence, the selection outlined in Table 9.3 does correspond to the fiducial/total phase space volume over which the relative rate of multijet events is measured.

It is very important to ensure that the differential cross section is determined over the *exactly*

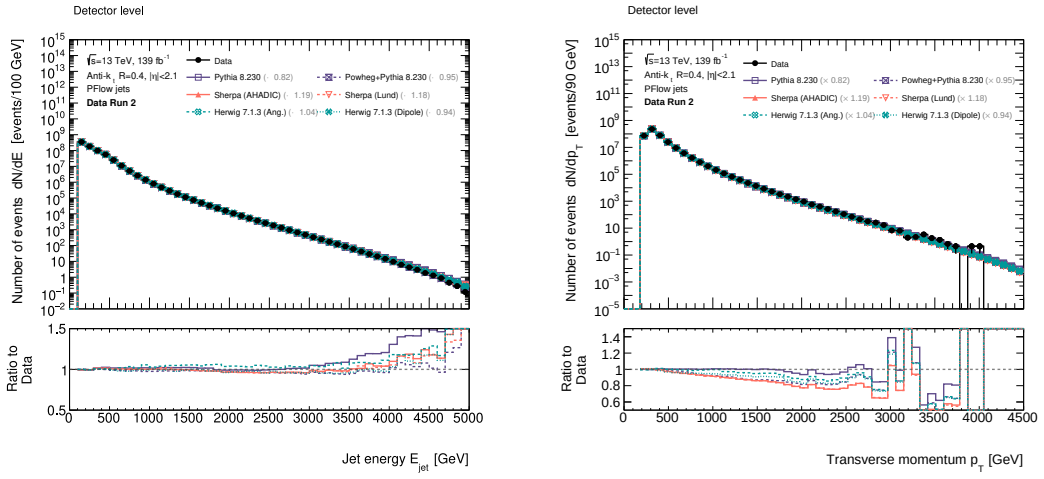
[8]This, however, does not guarantee that both PFlow contain at least one track associated with it.

same sample of jets for *all* substructure observables measured in this analysis. Modifications or additions to the selections in Table 9.3 between different measurements will result in a different composition of quark- and gluon-initialized jets of the two mixtures (central/forward jet sample), which inevitably changes the quark/gluon fraction(s). In this case, the individual measurements are not comparable with each other anymore.

9.3 Validation and Data-to-Monte-Carlo Comparison

With the completion of the previous section and the definition of the measurement's phase space volume at particle *and* detector level, all technical foundations of the analysis have been introduced. This paragraph now aims to compare data and simulation (at detector level of course) by means of a selected list of observables and their distributions in aspiration to estimate the data-to-Monte-Carlo agreement.

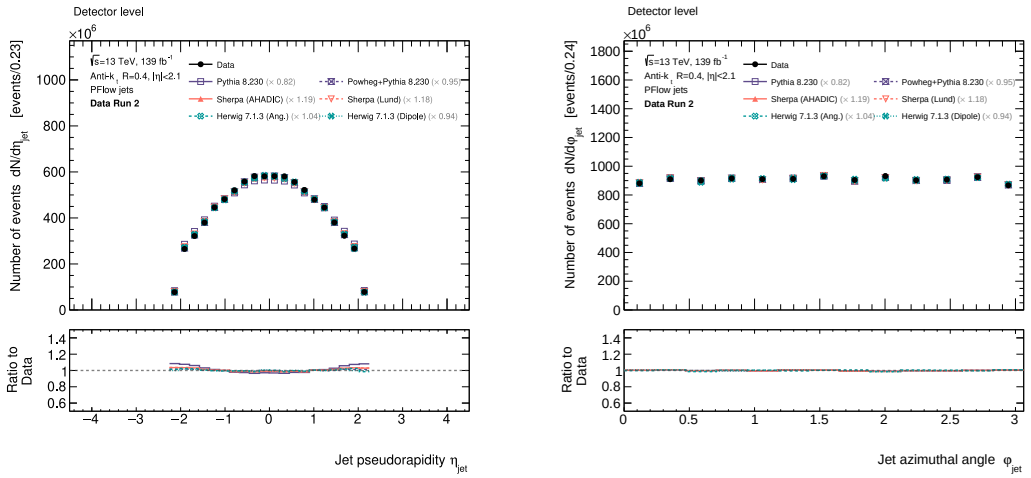
The first observables to consider are the scalar and angular components of the four-momentum of the reconstructed jet in its common representation in terms of p_{T}, E and η, ϕ. The former ones are given by the energy E and the transverse momentum p_{T} of the jet, both of which are shown in Figure 9.8 for the full Run 2 data set as well as all Monte Carlo samples listed in Table 8.1. The energy of the jet is simply given by the scalar sum of



(a) Jet energy E *(b)* Transverse momentum p_{T}

Plot 9.8: The reconstructed (detector-level) distribution of the jet energy E (*left*) and the transverse momentum p_{T} (*right*) for the full Run 2 data set and all simulated samples (for all MC campaigns) considered in this analysis.

the individual energies of each constituent inside the jet; hence, it does not depend on the angular distribution of the solitary particles and is therefore insensitive with respect to the mismodeling hereof. Accordingly, the energy distribution shown in Figure 9.8a in data and Monte Carlo shows an excellent agreement across a vast range of energies. In case of the transverse momentum p_{T} (9.8b). Contrary to the energy of the jet, the transverse momentum is indeed sensitive to the distribution of the particles constituting the jet. This manifests itself in a somewhat larger discrepancy between data and MC for this observable that, furthermore, shows a strong variation between the different generators. The next set of validation plots are the aforementioned angular distributions of the jets in Figure 9.9. Generally, all Monte Carlo samples are in good agreement with data. The pseudorapidity distributions of the jets in Figure 9.9a is constrained to $|\eta| < 2.1$ rather than the full tracker acceptance $|\eta| < 2.5$ of the Inner Detector to allow for track-to-jet association/matching. This pseudorapidity cut takes into account the radius of the jet $|\eta| < 2.5 - R(= 0.4)$ that guarantees for all charged particles – being associated to it – to lie within the acceptance region of the Inner Detector and hence to possibly (limited by the reconstruction efficiency) be reconstructed as a track which can be

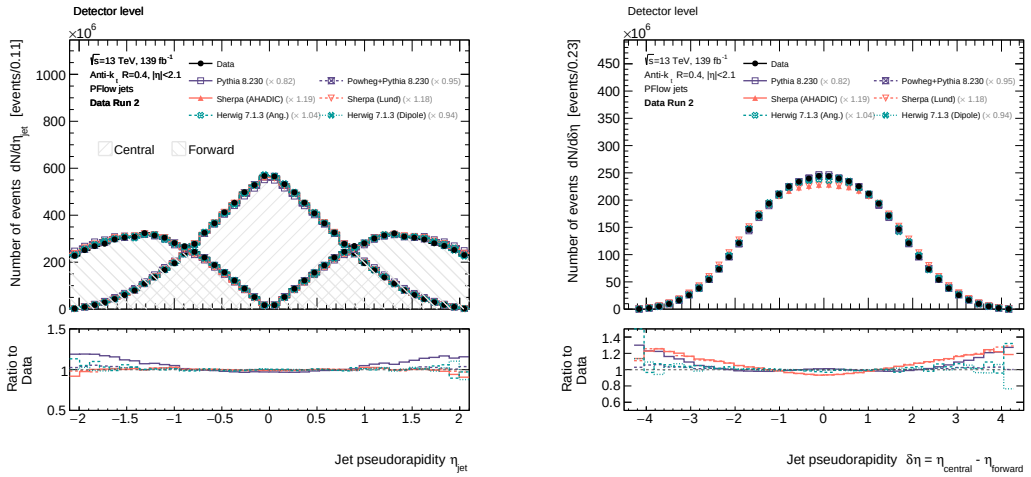


(a) Jet pseudorapidity η *(b)* Azimuthal angle ϕ

Plot 9.9: Distribution of the reconstructed jet pseudorapidity η (*left*) and azimuthal angle ϕ (*right*) at detector level.

assigned to the jet by means of a ΔR- or ghost-matching scheme.

Dwelling on the pseudorapidity a bit longer, Figure 9.10b takes a closer look at central and forward jets. As a quick reminder, a jet J_i with $i \in \{1(= \text{central}), 2(= \text{forward})\}$ (dijet selection) is called "central" if $|\eta_i| < |\eta_j|$ for $i \neq j$; otherwise, it is referred to as the forward jet. The distinction between central and forward jets plays an important role in this measurement; hence, it is important to verify that both "regions" are equally well modeled by simulation. The uncertainty in central and forward region will be different through because the material budget faced by all jets increases for jets that are more forward in *absolute* rapidity. As can



(a) Pseudorapidity distributions for "central" and "forward" jets *(b)* Pseudorapidity difference between "central" and "forward" jets

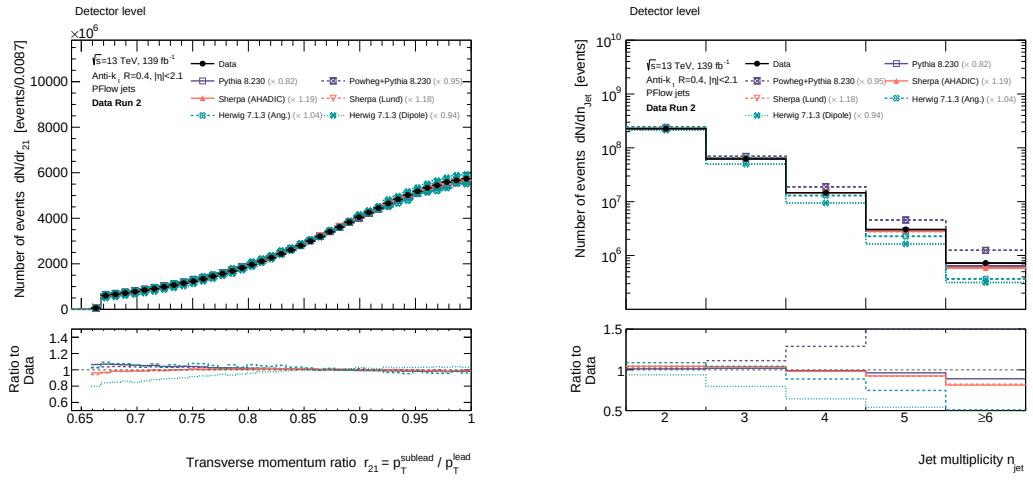
Plot 9.10: The reconstructed pseudorapidity divided into its distributions for central and forward jets (*left*) and the residual between the pseudorapidity of the more central and more forward jets, respectively, $\eta_{\text{central}} - \eta_{\text{forward}}$ on an event-on-event basis (*right*).

be seen in Figure 9.10a, there is an excellent agreement between data and Monte Carlo for central and forward jets. In case of forward jets close to the limits of the tracker acceptance, a small disagreement can be observed probably due to the above-mentioned increasing material

budget in the forward region of the detector. Figure 9.10b shows the residual between the pseudorapidity of the central and forward jet, $\delta\eta := \eta_{\text{central}} - \eta_{\text{forward}}$, in an event. This figure of merit is interesting as it is more sensitive to modeling effects than the per-jet pseudorapidity distributions because it accounts for both jets in the event. Accordingly, some characteristic and systematic deviations from data can be observed. Both PYTHIA-based (with POWHEG as matrix element generator) describe the data very well. An interesting feature is the opposite curve progression of the LO SHERPA and HERWIG samples (the NLO calculations generally agree very well with data). Since both SHERPA sample(s) utilize different hadronization models – but mostly agree in the description of $\delta\eta$ –, the origin of the observed behavior is probably due to the underlying parton shower models, i.e, the difference between a CSS and an angular-ordered shower. However, the effect may also be due the different PDF sets used for the two generator families that affect the emissions in central and forward regions due to differences between the probability of sampling a parton of a particular flavor, which affects the rapidity distribution.

The next two Figures 9.11a and 9.11b show the distribution of the transverse momentum ratio between the subleading $p_{\text{T}}^{\text{sublead}}$ and the leading $p_{\text{T}}^{\text{leading}}$ jet as well as the jet multiplicity n_{jet} in the event. The p_{T}-ratio is also part of the event selection with each multijet event in this measurement satisfying $p_{\text{T}}^{\text{lead}}/p_{\text{T}}^{\text{sublead}} < 1.5$ or $p_{\text{T}}^{\text{sublead}}/p_{\text{T}}^{\text{lead}} > 0.6\bar{6}$ accordingly; i.e. the p_{T} of the subleading jet must be at least 66 % of the leading jet in the event. This selection cut is mostly motivated by the suppression of events where the hard leading-p_{T} jet is accompanied by a second, very soft jet that most likely is not associated to the hard scatter, but originates from pile-up or noise. This helps in particular with the reduction of fake jets where a high-p_{T} jets in the event is faked by a high-p_{T} lepton with some pile-up-induced low-p_{T} jets being present. Furthermore, the cut also affects the event topology, resulting in more back-to-back jets and hence acts as an implicit isolation criterion, resulting in back-to-back jets as the p_{T} ratio is correlated with the jet thrust axis. Generally, the observable is well-described by simulation within the acceptance region – for jets beyond $p_{\text{T}}^{\text{lead}}/p_{\text{T}}^{\text{sublead}} < 1.5$, the discrepancy increases due to the limited modeling of soft QCD effects.

Figure 9.11b shows the number of jets reconstructed in the event, whereby the last



(a) p_{T} ratio for the two leading jets in the event satisfying $p_{\text{T}}^{\text{sublead}}/p_{\text{T}}^{\text{lead}} > 1/r_{21}^{\text{cut}}$

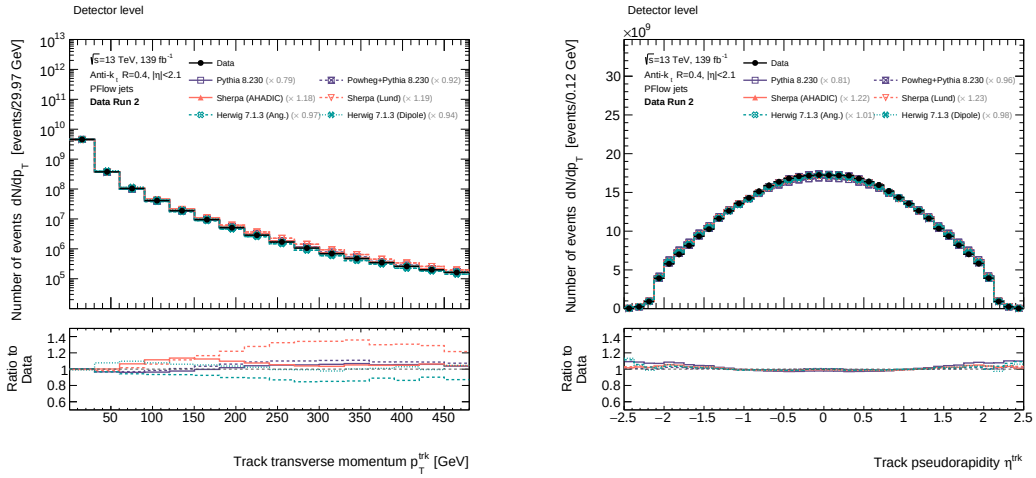
(b) Number of reconstructed jets in the event n_{jet} with the last bin being inclusive

Plot 9.11: Ratio of the leading and subleading jet's p_{T} passing the selection cuts (*left*) and the jet multiplicity n_{jet} (*right*).

multiplicity bin is inclusive, i.e., including contributions from all events with at least six jets ($n_{\text{jet}} \geq 6$) being present. The jet multiplicity is very sensitive to all kind modeling effects reaching from the matrix element calculation, parton shower and hadronization; therefore, it is a suited variable to check data-to-Monte-Carlo agreement. As expected, the disagreement between the actual measurement and simulation increases towards higher jet multiplicities due to the presence of "softer" jets, which are hard to model in simulation. The production of

additional $(2\,\text{jet} + N)$ jets particularly differs between LO and NLO matrix element calculations that are combined with a parton shower generator. The former one is a $2 \to 2$ event generator that completely relies on parton shower emissions to give rise to higher jet multiplicities in the event. This is problematic to the effect that the parton shower approach/approximation (as mention in Chapter 2.4) is based on the factorization of the cross section in the soft and collinear limit. For non-collinear splittings, the parton shower model is an approximation whose quality depends on the amount to which the initial assumption of a collinear splitting is violated. This is especially true for the choices of ordering variables that are only equivalent exactly in the collinear limit; beyond it, different ordering variables may result in (partially!) substantially different results. In case of an NLO matrix element generator, however, the situation is different as the real emission that enters the calculation at α_S^3 populates a different region in phase space than the parton shower (which is one motivation for the combination of matrix element calculations and parton shower simulation in the first place). Furthermore, the NLO matrix element is showered like the LO one; but, because the NLO sample already starts with three particles being present in the final state, the probability of having larger jet multiplicities is enhanced for higher-order calculations in general.

All validation plots considered so far – except for the jet multiplicities – are based on the (leading and subleading) *PFlow* jets. As explained previously, all substructure observables, however, are calculated from the set of *tracks* matched to those topo-cluster jets. This section is therefore concluded by inspecting the distribution of reconstructed track momenta p_T^trk as well as their pseudorapidity η^trk in Figure 9.12. There is no need to examine the corresponding



(a) Reconstructed track momentum p_T^trk *(b)* Reconstructed track pseudorapidity η^trk

Plot 9.12: Reconstructed transverse momentum p_T^trk (*left*) and pseudorapidity p_T^trk (*right*) for all *inner detector* tracks with $p_\text{T}^\text{trk} > 500\,\text{MeV}$.

figure for the azimuthal angle ϕ^trk as it is uniformly distributed by construction. Furthermore, the "mass" of the track is *constrained* to be close to the mass of the pion m_π. All distributions in Figure 9.12 are in good agreement with each other. The pseudorapidity is particularly well described as the angular resolution of the Inner Detector is excellent. Contrary to the pseudorapidity of the reconstructed jets (see Figure 9.9a), the pseudorapidity of the tracks extends over the entire acceptance range of the Inner Detector (see Chapter 3.3.4), i.e., $|\eta| < 2.5$. With the $R = 0.4$ jet being constrained to $|\eta| < 2.1 (= 2.5 - R)$, the track-to-jet matching does not loose efficiency for larger pseudorapidities.

9.4 Baseline Distributions

All distributions presented in the previous section served the purpose of validating the overall agreement between data and simulation. In the course of this measurement, plenty more figures

of merit have been studied (for reasons of clarity and comprehensibility not included in the report at hand) in order to ensure that the data is well modeled all in all. This is important as many uncertainties of the measurement are defined based on generator-to-generator variations in the target distributions; moreover, the detector-level distributions are unfolded to particle based on the response matrix derived from simulation.

This section connects to the previous one by introducing the corresponding distributions of the *transverse momentum* p_{T} as well as *relative pseudorapidity* η_{rel} along with their respective binning used in the *actual* analysis for both groups of measurements, *firstly*, the triple-differential multijet cross section with respect to a counting-based observable (track multiplicity n_{trk} and Soft Drop multiplicity n_{SD}) as well as, *secondly*, the NN-score $z_{\mathrm{E/PFN}}$ predicted by the EFN and the PFN (neural networks) respectively (the concrete implementation of both neural networks will be explained in great detail later in this thesis). The binning for the just mentioned jet substructure variables $n_{\mathrm{trk/SD}}$ and $z_{\mathrm{E/PFN}}$ are introduced in their dedicated chapters.

The selected binning for the transverse momentum and the relative pseudorapidity are shown in Table 9.4. In total, the spectrum of the transverse momentum is measured in five

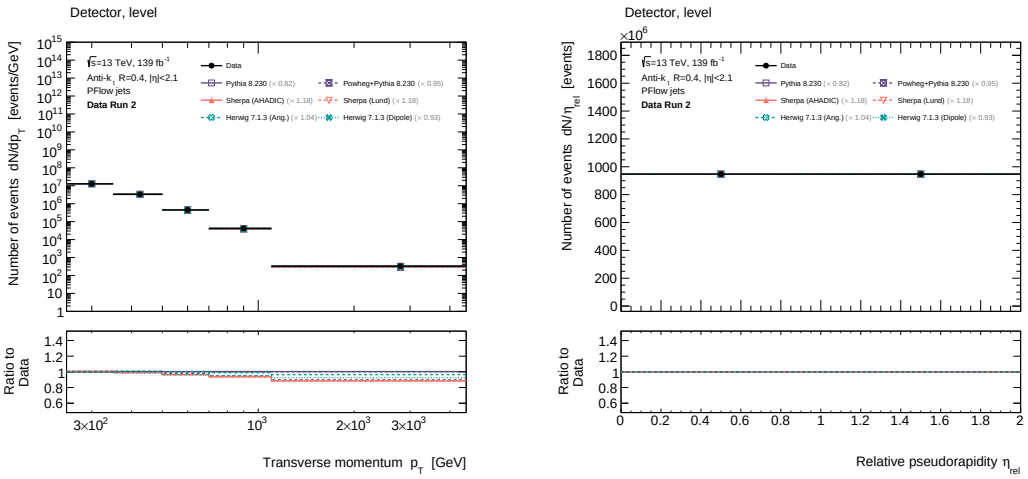
Observable	Binning					Unit
p_{T}	$[250, 350)$	$[350, 500)$	$[500, 700)$	$[700, 1100)$	$[1100, 4500)$	GeV
η_{rel}	central$(=0)$		forward$(=1)$			none

Table 9.4: Binning of the transverse momentum p_{T} as well as the relative pseudorapidity η_{rel} that is used for the *entire* analysis thought *all* cross section measurements. The assigned numbers for the relative pseudorapidity are basically arbitrary, but chosen according to the definition $\eta_{\mathrm{rel},i} = \Theta(|\eta_i| - |\eta_j|)$, whereby $i, j \in \{1, 2\}$.

exclusive (i.e., restricted to a fixed range) p_{T} bins of *non-equidistant* widths. The definition of the width of the individual bins is basically determined by the prescaled higher-level single jet triggers in the range of low p_{T}, as well as the compensation for decreasing statistical significance of high-p_{T} jets as well as the reduction of migrations due to JES/JER-related variations.

The relative pseudorapidity η_{rel} of the two leading jets in the event is simply divided into two disjoint categories, i.e., *central* and *forward*.

Both inclusive distributions are shown in Figure 9.13 for data and *all* Monte Carlo samples



(a) Transverse momentum p_{T} *(b)* Relative pseudorapidity η_{rel}

Plot 9.13: Baseline distributions of the transverse momentum p_{T} (*left*) as well as the relative pseudorapidity η_{rel} (*right*) with the appropriate binning quoted in Table 9.4 as used in all measurements through this analysis.

listed in Table 8.1. The legend entries show – besides the name/label (see Figure 8.1) of the respective Monte Carlo generator – the normalization factor for each sample with respect to the number of events measured in data according to the normalization factor. Something immediately catching one's eye are the strong variations between the different Monte Carlo event generators (even within one family), which is expected in case of multijet processes as the precision of the normalization is limited mainly do to long-range effects that are mostly described by empirical models. This discrepancy of the normalization with reference to data, however, is irrelevant in the context of this analysis because all rate measurements presented in this report are relative with respect to the total number of events in the fiducial/total volume; hence, only the "shape information" of the distributions enters.

Figure 9.13a shows the distribution of the transverse momentum for the binning stated in Table 9.4 for central and forward jets inclusively, i.e., the relative pseudorapidity has been marginalized over $P(p_{\mathrm{T}}) = \int \mathrm{d}\eta_{\mathrm{rel}}\, P(p_{\mathrm{T}}, \eta_{\mathrm{rel}})$. Both p_{T} spectra in central and forward region are very similar as the p_{T} is not very sensitive to quark/gluon-related features. Each bin was normalized by the corresponding bin width to account for non-equidistant bin sizes. One obtains a smoothly falling p_{T} spectrum whose description by simulation basically follows the characteristics already observed in the corresponding validation plot in Figure 9.8b. The selected binning of the transverse momentum is motivated by the unfolding and has been optimized by iteratively repeating the *entire* measurement several times.

The inclusive distribution of the relative pseudorapidity η_{rel} for all transverse momentum p_{T} bins is shown in Figure 9.13b. The distribution contains only two bins with the same cross section each. This is the desired property that directly follows from the definition of this observable. The number of events in each bin is identical due to the dijet selection criteria that requires at least two jets of which the leading two ones always fall in opposite rapidity categories. Trivially, the "shape" of this distribution is exactly described in Monte Carlo.

10. Uncertainties

This chapter provides a conceptual introduction to all the statistical and systematic uncertainties that are relevant for the measurements presented in this report.

The precision of the final measurement of the triple-differential multijet cross section with respect to various jet substructure observables is limited by several sources of statistical and, dominantly, systematic uncertainties. The latter one includes, for instance, the uncertainty associated to the modeling of the different production steps of the event generation like the calculation of the matrix element of the underlying subprocess (dijet production), the parton shower simulation and its various models and ordering variables, as well as the formation of color-singlet hadrons by means of the hadronization model; the uncertainty due to the calibration of the jet-energy scale; the underlying PDF for the sampling of the initial momenta of the partons inside of the proton; different "definitions" of tracks (e.g., variation of low momentum cuts etc.); variation of the cross sections due to limited knowledge of the exact pile-up condition; disabled modules in the tile calorimeter as well as the bias due to the simulation-based unfolding etc.

Regarding the organization of this chapter, each component of the total uncertainty gets its dedicated section. The fist section (10.1) outlines the calculation of the statistical uncertainties in data and Monte Carlo. The remaining sections describe several sources of systematic uncertainties. The second section (10.2) describes the calculation and composition of the modeling-related uncertainty that consists of an uncertainty related to the underlying precision of the calculation of the matrix element (10.2.1), the parton shower simulation (10.2.2) , as well as the underlying hadronization model (10.2.3). This is followed by the tracking-related uncertainties 10.3. Following that (10.4), the uncertainty due to the choice of the underlying PDF is described. Subsequently (10.5), the uncertainty due to the calibration of the energy scale of the jet is introduced. This is followed by a quick description of the uncertainty due to the limited knowledge of the exact pile-up conditions (10.6) as well as disabled modules in the tile calorimeter (10.7). After that (10.8), the uncertainty associated with the modeling-bias by means of a data-driven non-closure test is explained. Also, possible biases due to variations in the conditions of the measuring apparatus are estimated by a closure test between different runs of data taking (10.9). The calculation of the uncertainty associated to the extracted probability distributions of quarks and gluons as well as their fractions (10.10) is outlined.

10.1 Statistical Uncertainties

The statistical uncertainties (in data and MC) in each phase space bin are determined based on the execution of a fixed number of pseudo-experiments. In each *toy*-experiment, the measured number of events in each bin randomly fluctuates according to a probability density function suitable for this task. One obvious (and common) choice for the PDF would be a Gaussian, which is a frequent choice for the nuisance parameter θ of many systematic uncertainties according to

$$\rho(\theta) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{(\theta - \bar{\theta})^2}{2\sigma^2}\right), \tag{10.1}$$

whereby $\bar{\theta}$ denotes the mean value of θ. However, a Gaussian prior (according to Equation 10.1) is not suitable for a class of observables that is positively defined like, e.g., the cross section,

efficiencies etc., as it may gives rise to negative, ill-defined values. One *ad hoc* approach to solve this issue of negative values is to truncate the Gaussian at zero. This approach, however, is not well motivated as truncating the lower tail of the Gaussian distribution effectively changes the prior of the respective parameter of interest.

Another option is to use a prior that excludes negative values in the first place. A common choices is a log-normal distribution. The log-normal distribution is given by

$$\rho(\theta) = \frac{1}{\sqrt{2\pi}\ln(\kappa)} \exp\left(-\frac{\ln\theta/\bar{\theta}^2}{2\ln(\kappa)^2}\right)\frac{1}{\theta}, \tag{10.2}$$

which is a probability distribution of a random variable whose *logarithm* is normally distributed. In Equation 10.2, κ denotes the width of the log-normal distribution. For large mean values, the log-normal distribution approaches an ordinary Gaussian.

The mean value of the log-normal distribution in Equation 10.2 is just given by the measured cross section in the respective phase space bin. Accordingly, the variance is defined by the sum of the squares of the trigger prescale weights.

After the parameters of the log-normal distribution are determined, the statistical uncertainty is derived based on a sequence of '200' pseudo-experiment. In each run, the measured number of events in each phase space bin is varied/smeared according to the above-mentioned log-normal prior, which results in a total number of $n_{\text{toy}} = 200$ (toy) measurements that are subsequently unfolded to particle level. In accordance with Equation 10.2, phase space bins with a small event count will result in larger variances, which consequently results in larger statistical uncertainties. After the pseudo-experiments have been performed, the statistical uncertainty in each phase space bin is then given by one-half of the difference between the 84th and 16th quantiles of the toy sample, which corresponds to the 68 % Inter-Quantile Range (IQR) of the generated output distributions (basically the 1σ-equivalent in case of a Gaussian distribution).

In case of real collision data, the unfolding procedure, which transforms the measurement to particle level, introduces correlations between different regions of phase space. Therefore, there is a non-trivial (i.e., non-diagonal) covariance matrix for the data given by

$$\Sigma_{ij}^{\text{stat}} = \sum_{k=1}^{n_{\text{toy}}} (x_{i,k} - \langle x_i \rangle)(x_{j,k} - \langle x_j \rangle), \tag{10.3}$$

whereby the index runs over all toy measurements and x_i denotes the measured number of events in the ith phase space bin.

10.2 Modeling-Related Uncertainties

The unfolding of a measurement relies on the specification of a so-called *nominal* Monte Carlo simulation that is used to derive and construct the response matrix, which is then applied to data in order to obtain the corresponding cross section at particle level corrected for detector and acceptance effects, characteristic to the detector that was used to record the data. As there is no model that provides a perfect simulation/description of nature, the uncertainty associated with the choice of different Monte Carlo samples must be estimated. For many of today's measurements, this *model-related uncertainty* is among the most dominant sources of errors. There is no obvious, fundamentally motivated best choice that distinguishes different Monte Carlo event generators; in fact, the performance of different generators is known to vary substantially between different regions in phase space and processes. Hence, the definition of a nominal Monte Carlo sample for a measurement often relies on empirically acquired insights given the agreement with respect to data.

The specification of the various models used in the individual steps of the production chain of the event generation must be accounted for by an uncertainty. This modeling-related uncertainty is subdivided into three individual components that reflect the three main steps of the event topology simulated by Monte Carlo event generators: *firstly*, limited precision of the

calculation of the underlying subprocess (multijet production) by means of the *Matrix Element* (ME); *secondly*, the variation in the underlying model of the *Parton Shower* (PS) simulation; and, *thirdly*, the utilization of different models for the formation of color-singlet hadrons by means of *Hadronization* (Had). All three uncertainties (as well as their concrete calculation by means of variations of the triple-differential multijet cross section) will be explained below.

However, before taking a closer look at the individual modeling uncertainties, their basic calculation – which is shared by the ME-, PS- and Had-related uncertainty – will be explained in the following.

First and foremost, *there is no universally applicable definition of the modeling-related uncertainty*. All definitions, however, are somehow based on the variation of the reconstructed spectra using different models for the generation of the events from which the respective distributions are finally derived from. Furthermore, the definition of the modeling-related uncertainty should yield zero, provided that the variation in the underlying model results in no variation of the cross section at all. To introduce the definition of the modeling uncertainty that is used throughout this report, consider the following equation that was already introduced (in a slightly modified version) in Chapter 5 in the context of the unfolding procedure

$$N_i^{\mathrm{PL}} = \sum_j N_{\mathrm{DL}}^{\mathrm{multijet},j} \left(\left[R^{-1} \right] \right)_{\mathrm{DL},j}^{\mathrm{PL},i} \cdot \frac{\epsilon_{\mathrm{fake},j}}{\epsilon_{\mathrm{eff},i}}, \tag{10.4}$$

whereby $N_{\mathrm{DL}}^{\mathrm{multijet},j}$ denotes the number of multijet events at *detector* level. Besides the number of events at detector level, the total event yield at particle level N_i^{PL} depends on the response matrix R, the efficiency factor ϵ_{eff}[1] that accounts for losses due to jets that migrate out of the active phase space volume – which would result in an underestimation of the expected number of events –, as well as the fake factors ϵ_{fake} that correct the event yield for jets falsely migrating into the measurement's volume resulting in an overestimation of the cross section. All three components, response matrix R, efficiency ϵ_{eff} and fake ϵ_{fake} factor, are *model dependent*, i.e., using a different (Monte Carlo generator) model will result in different distributions for those elements. It is not clear *a priory* and not unambiguously well defined how the final model-related uncertainty is obtained from the model-dependent variations of R, ϵ_{eff} and ϵ_{fake}; there are several options at this point. Many different choices for the modeling-related uncertainty of the cross section measurement have been investigated in the course of this study, all of which yield similar results.

The final uncertainty of the matrix element, the parton shower and the hadronization model is defined by means of an *envelope* of the variations in the cross section for the response matrix and both efficiencies

$$\sigma^{\mathrm{model}} = \max \left\{ \sigma^{\mathrm{resp}}, \sigma^{\epsilon_{\mathrm{eff}}}, \sigma^{\epsilon_{\mathrm{fake}}} \right\}, \tag{10.5}$$

whereby σ^{resp} and $\sigma^{\epsilon_{\mathrm{eff/fake}}}$ denotes the bin-by-bin variation of the cross section with respect to the response matrix and the efficiency and fake factors, respectively. The variations associated with the response matrix as well as the efficiency and fake factors in Equation 10.5, respectively, are obtained by using the appropriate quantity from the *alt*ernative model while keeping the remaining ones fixed at the value measured in case of the *nom*inal Monte Carlo generator. In this case, variations in the cross section related to the response matrix, the efficiency and fake

[1] The response matrix as well as the efficiency and face factors are *always* derived from the very same Monte Carlo event generator.

factors are given by

$$(N_{\text{alt}}^{\text{nom}})_{i,\text{resp}}^{\text{PL}} = \sum_j N_{\text{nom},j}^{\text{PL}} \left([R_{\text{alt}}]^{-1}\right)_{\text{DL},j}^{\text{PL},i} \cdot \frac{\epsilon_{\text{fake},j}^{\text{nom}}}{\epsilon_{\text{eff},i}^{\text{nom}}}, \tag{10.6}$$

$$(N_{\text{alt}}^{\text{nom}})_{i,\epsilon_{\text{eff}}}^{\text{PL}} = \sum_j N_{\text{nom},j}^{\text{PL}} \left([R_{\text{nom}}]^{-1}\right)_{\text{DL},j}^{\text{PL},i} \cdot \frac{\epsilon_{\text{fake},j}^{\text{nom}}}{\epsilon_{\text{eff},i}^{\text{alt}}}, \tag{10.7}$$

$$(N_{\text{alt}}^{\text{nom}})_{i,\epsilon_{\text{fake}}}^{\text{PL}} = \sum_j N_{\text{nom},j}^{\text{PL}} \left([R_{\text{nom}}]^{-1}\right)_{\text{DL},j}^{\text{PL},i} \cdot \frac{\epsilon_{\text{fake},j}^{\text{alt}}}{\epsilon_{\text{eff},i}^{\text{nom}}}, \tag{10.8}$$

$$\tag{10.9}$$

whereby $N_{\text{nom/alt},i}^{\text{PL}}$ denotes the number of events in the ith bin of the nominal/alternative sample.

The variations are enveloped because the three elements σ^{resp} and $\sigma^{\epsilon_{\text{eff/fake}}}$ are correlated in a non-trivial way; hence, they can not be just added in quadrature, which assumes no correlation at all. Because the degree of correlation is unknown (at least in this analysis), using the maximum variation represents a common choice.

The next sections now introduce the motivation as well as the precise calculation rule for the ME-, PS- as well as Had-related uncertainty, based on the Monte Carlo event generators listed in Table 8.1.

10.2.1 Matrix Element

As explained in Chapter 2.3, the first part of the event generation is the calculation of the hard subprocess, which is calculated by means of perturbation theory because the momentum transfer between the colliding particles/partons is large. The calculation of the matrix element is a fixed-order calculation, i.e., the perturbative expansion of the scattering amplitude is terminated at some fixed order in the coupling α_{S}. The order of perturbation, which is taken into account to compute the hard subprocess, defines the precision as well as the order of the *whole* event. In this analysis, most event generators (including the baseline model) use a leading-order calculation of the matrix element because higher-order calculations are computationally expensive and time consuming, which usually results in low statistics in phase space regions with small cross sections. Furthermore, the matching and merging with the subsequent parton shower model is much more involved and introduces non-trivial event weights that further complicate the analysis. To estimate the uncertainty due to the termination of the perturbative series at some fixed order, the LO cross section must be compared to the one obtained using a higher-order calculation of the matrix element (at least NLO). It is important to ensure that both samples only differ with respect to the degree of precision of the underlying hard subprocess to ensure that the variation of the cross section is indeed due to the matrix element calculation solely.

For all measurements of the triple-differential multijet cross section presented in this report, the ME-related uncertainty is calculated based on the prescription described in Section 10.2, using the multijet samples generated from Pythia 8.230 (LO) and Powhegv2+Pythia 8.230 (NLO).

To obtain the associated uncertainty, the nominal spectra of the triple-differential multijet cross section (derived using the baseline generator Pythia 8.230 at LO) is unfolded using the response matrix derived from the Powhegv2+Pythia 8.230 sample at NLO. The unfolded spectrum is subtracted from the nominal prior at truth level; the relative uncertainty for each bin is obtained by normalizing with respect to the truth-level prior in each bin according to

$$\sigma_{\text{ME},i}^{\text{model}} = \frac{\left(N_{\text{PH+PY8}}^{\text{PYTHIA8}}\right)_i^{\text{PL}} - N_{\text{PYTHIA8},i}^{\text{PL}}}{N_{\text{PYTHIA8},i}^{\text{PL}}}, \tag{10.10}$$

whereby $N_{\text{PYTHIA8},i}^{\text{PL}}$ denotes the number of events in the ith bin of the baseline (Pythia 8.230) sample and $\left(N_{\text{PH+PY8}}^{\text{PYTHIA8}}\right)_i^{\text{PL}}$ being the number of events in the same bin, but for the

nominal sample at detector level that has been unfolded using the response matrix from POWHEGv2+PYTHIA 8.230 according to

$$\left(N^{\text{PYTHIA8}}_{\text{PH+PY8}}\right)^{\text{PL}}_{i} = \sum_{j} N^{\text{PL}}_{\text{PYTHIA8},j} \left([R_{\text{PH+PY8}}]^{-1}\right)^{\text{PL},i}_{\text{DL},j} \cdot \frac{\epsilon^{\text{PH+PY8}}_{\text{fake},j}}{\epsilon^{\text{PH+PY8}}_{\text{eff},i}}. \tag{10.11}$$

10.2.2 Parton Shower

The next modeling uncertainty is related to the parton shower simulation. All parton shower models are based on the factorization of the multi final-state cross section into a cross section of smaller final-state multiplicity multiplied by a factor that accounts for the emission of collinear radiation $\mathrm{d}\sigma_{n+1} \propto \sum_i \mathrm{d}\sigma_n P_i(z,\phi)$. With this picture in mind, it is obvious that the internal structure (i.e., the distribution of radiation) inside of a jet is basically almost completely determined by the parton simulation step. This implies that most substructure observables are *highly* sensitive to the variation of the underlying model that is used to simulate the emission of additional radiation inside of the jet. Hence, the PS-related uncertainty is expected to be among the most dominant ones (along with the uncertainty related to the calibration of the jet-energy scale and tracking) and therefore one of the limiting factors of the precision of this measurement. Furthermore, due to the underlying premises of the parton shower philosophy explained in Chapter 2.4, the uncertainty related to the PS is expected to be "noticeable" p_{T}-dependent. With increasing transverse momenta of the jets, the uncertainty is expected to decrease as in the limit of collinear radiation (with higher-p_{T} of the jets, the emissions will be on average more collinear according to $\Delta R \propto m_{\text{jet}}/p_{\text{T}}$), most parton shower models coincide; in particular, all definitions of the ordering variable for the emissions are equivalent in the collinear limit. Differences between the various models arise because the parton shower approach is also used beyond the collinear approximation, which is particularly the case for low-p_{T} jets that feature wide and soft radiation.

To obtain the uncertainty related to the parton-shower model, the triple-differential multijet cross section derived from two samples that differ with respect to their PS model must be compared. As many Monte Carlo samples listed in Table 8.1 use different approaches in the parton-shower simulation step, there are may ways to define the corresponding variation of the cross section. As already mentioned in the introduction of this chapter, several combinations have been evaluated, all of which yield comparable results. The definition, which was finally decided on, is based on the two (NLO) HERWIG samples that use an *angular-ordered* parton shower (HERWIG's default) and a parton shower approach based on a *dipole* model that models the emission of radiation similar to an elementary electric dipole antenna.

The PS-related uncertainty is defined by the difference between the particle-level spectra obtained by unfolding one of the HERWIG-based distributions using the response matrix based on the other generator and the truth-level spectra obtained from either of the HERWIG samples (meaning: unfolding one of the HERWIG-based detector-level spectra with a response matrix derived from one of the two HERWIG samples). Normalized to the truth-level prior further cancels many effects to the degree of their correlation (PDF, matrix element, hadronization etc.) such that the resulting relative uncertainty can be considered as a very good representative of the PS-related variation of the corss section. The final (relative) uncertainty due to the choice of the parton shower model is defined as an envelope of the differrent parton shower models according to

$$\sigma^{\text{model}}_{\text{PS},i} = \frac{\left(N^{\text{H7+DIPOLE}}_{\text{H7+ANGLE}}\right)^{\text{PL}}_{i} - N^{\text{PL}}_{\text{H7+ANGLE},i}}{N^{\text{PL}}_{\text{H7+ANGLE},i}}, \tag{10.12}$$

whereby $\left(N^{\text{H7+ANGLE}}_{\text{H7+DIPOLE}}\right)^{\text{PL}}_{i}$ denotes the number of events of the unfolded HERWIG spectrum (with *angular* ordered emissions) using the response matrix derived from the HERWIG samples based on a dipole model. If the response matrices of both HERWIG samples coincide, the PS-related systematic uncertainty vanishes by construction as it is expected for a proper uncertainty metric.

10.2.3 Hadronization

The last component of the simulation chain (except for the subsequent decay of the excited states) is the already mentioned hadronization step that simulates the *non*-perturbative process of the formation of color-less hadrons, which further decay to the "stable" particles measured by the detector. The variation of the cross section due to different hadronization models is expected to be less pronounced than it was the case for the PS-related one since most of the jet's internal structure is formed by the parton shower – that also has the highest sensitivity to quark-gluon-related features of the jet as the average number of radiation in a jet is roughly proportional to the color factor of quarks and gluons. Generally, the uncertainty imposed by hadronization step is expected to lay somewhere between the one connected to the matrix element and the parton shower.

The instructions for the determination of the hadronization-related uncertainty follows closely the one used for the parton shower uncertainty in the previous paragraph. For the calculation of the hadronization uncertainty, two Monte Carlos are needed that differ with respect to their hadronization model. Mainly for this purpose, the two SHERPA (version 2.2.5) samples listed in Table 8.1 have been introduced. Both generators agree in *all* event generation steps except for the hadronization model that is used. As explained in Chapter 2.5, there are two main classes of hadronization models available: the (Lund) string model as well as the cluster-based model. Each one of the two approaches to hadronization is used by one of the two SHERPA-based samples.

To obtain the final systematic uncertainty related to the variation of the triple-differential multijet cross section, the nominal Lund-based SHERPA detector-level spectrum is unfolded with the response matrices separately obtained form the SHERPA model that uses a cluster-based approach to hadronization; the uncertainty is then given by the difference between the two unfolded particle-level spectra according to

$$\sigma^{\text{model}}_{\text{Had},i} = \frac{\left(N^{\text{SHERPA+CLUSTER}}_{\text{SHERPA+LUND}}\right)^{\text{PL}}_{i} - N^{\text{PL}}_{\text{SHERPA+LUND},i}}{N^{\text{PL}}_{\text{SHERPA+LUND},i}}, \tag{10.13}$$

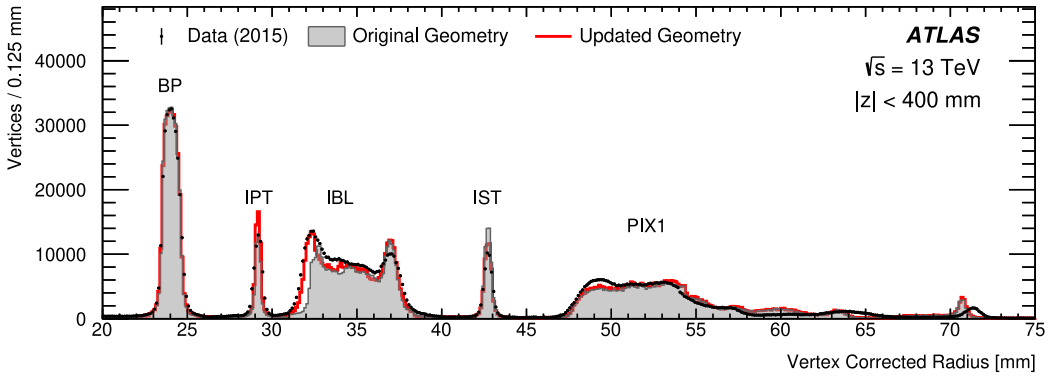
whereby $\left(N^{\text{SHERPA+CLUSTER}}_{\text{SHERPA+LUND}}\right)^{\text{PL}}_{i}$ denotes the number of events of the unfolded baseline spectrum using the response matrix derived from either of the two SHERPA samples, differing by their hadronization model.

10.3 Tracking Uncertainties

The measurement of the jet substructure observable is based on jets that have been matched to the reconstructed jet; consequently, tracking-related sources of systematic uncertainties must be taken into account. In this analysis, the tracking-related uncertainties are divided into three categories, i.e., the (global) reconstruction efficiency, the fake rate, the momentum scale of tracks as well as efficiencies within jets (dense environment).

The uncertainty associated with the reconstruction efficiency of tracks is composed of an uncertainty related to the inner detector material due to limited knowlege of the detector (referred to as inclusive efficiency uncertainty), which has been intensively studied in photon conversions as well as hadronic interactions (see Figure 10.1). The total uncertainty is 0.5 % for $|\eta| < 0.1$, growing up to a rwlative uncertainty of 2.7 % for tracks with $2.3 < |\eta| < 2.5$. The inclusive efficiency uncertainty is propagated through the measurement by randomly removing tracks in the simulation with a probability that depends on the p_{T} and η of the track.

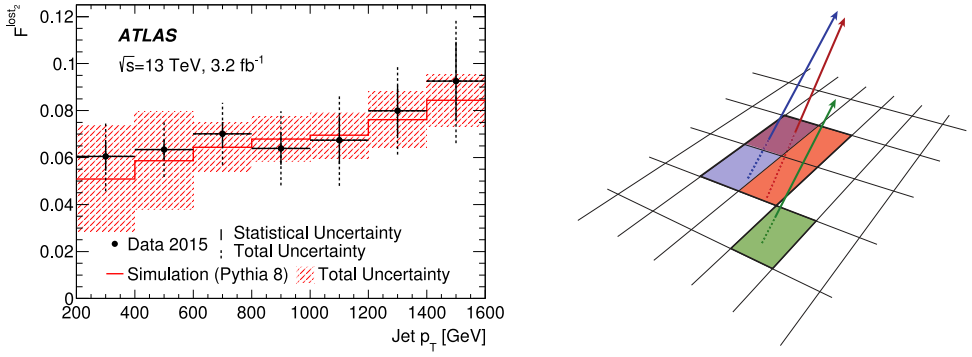
The second efficiency-related component of the tracking uncertainty accounts for clusters of pixels being compatible with multiple track candidates due to merged clusters in dense environments. This is particularly true for the region close to the center ($\Delta R \approx 0.1$) of reconstructed jets. The uncertainty related to the merging of clusters (and the resulting reduction in the reconstruction efficiency of tracks) has been determined based on the $\mathrm{d}E/\mathrm{d}x$ method as described in Ref. [44]. Following the ATLAS recommendations, in this analysis, the tracking-related efficiency in dense environments is accounted for by varying the transverse



Plot 10.1: "Comparison of the radial distribution of hadronic interaction candidates between data and simulation (original and updated simulations) for [...] 20 mm $< r <$ 75 mm [...]" (the figure as well as its caption have been adopted from Ref. 43, Fig. 10, p. 20)

momentum of the tracks within 0.4 % provided that they are within $\Delta R = 0.1$ with respect to the center of the reconstructed jet.

The next component of the tracking-related uncertainty is due to fakes that result from



(a) *(b)*

Plot 10.2: "The measured fraction of lost tracks [...] in the jet core [...] as a function of jet p_{T} for data [...] and simulation [...]." (Ref. 44, Fig. 17, p. 23) (*left*) and *"[i]llustration of [...] merged pixel cluster due to very collimated charged particles."* (Ref. 44, Fig. 2b, p. 5) (*right*).

ramdom combinations of hits due to (charged) particles that overlap in space mostly due to underlying activity (pile up) in the event. Another potential source of fakes (inside jets) is due to the already mentioned high density of tracks close to the core of jets that results in reduced resolution of individual charged particles as shown in Figure 10.2.

Concerning the parameters of the reconstructed track, the charge-over-p_{T} (q/p_{T}) is the leading source of systematic uncertainty due to detector misalignment weak modes [203]. The potential bias is corrected once per data-taking period; the total correction of charge-over-p_{T} ratio is something about 0.0001/GeV (except for $\phi \approx 0$ and $|\eta| \approx 2.5$ where the correction may reach 0.001/GeV).

10.4 PDF Uncertainties

There is a large variety of PDF set available that differ regarding the data included in the global fit, as well as the shape/model of the parameterization used for the distributions of

quarks and the gluon. Of course, the choice of the PDF set introduces an additional uncertainty to the cross section measurement that must be taken into account as the underlying PDF of the proton determines the parton content and therefore directly affects the fraction of quarks and gluons in the final state.

The baseline generator is given by PYTHIA 8.230 that uses a PDF based on the NNPDF methodology [204] which *"[...] determines the structure of the proton using contemporary methods of artificial intelligence."* [205]. Therefore, the PDF uncertainty of the triple-differential multijet cross section measurements presented in this thesis is studied mainly with respect to variations of the latest PDF set provided by the NNPDF collaboration. However, other PDF sets like the NLO-QCD CTEQ6.6 [206], or the MSTW2008 [207] have been investigated as well.

The estimation of the uncertainty associated to the PDF requires a careful treatment as each collaboration provides a prescription for their calculation. Broadly speaking, there is the so-called Hessian approach (which is, to name some examples, used for the CTEQ- and MSTW-based PDF sets) and the average over an ensemble of PDF replicas as it is used by the NNPDF-based PDFs.

The Hessian formalism studies the uncertainties through a quadratic expansion of the χ^2 function around its global minimum. The method receives its name from the Hessian matrix, i.e., the matrix of second-order derivatives of the χ^2 function *at* its minimum. The *Eigenvectors* of the Hessian matrix are then used to construct an *Eigenbasis* $\{S_\ell^\pm\}_{\ell=1}^d$ of a d-dimensional space, whereby each *Eigenvector* $S_\ell^\pm$ corresponds to a PDF that can be considered as a variation in the neighborhood of the nominal one.

In case of the NNPDF-based PDF, the calculation of errors follow a different procedure as the shapes of the density functions is parameterized by a neural network. The NNPDF fitting methodology uses a Monte Carlo representation of the original data. To estimate the uncertainty, a fixed number of replicas is generated based on the experimental data taking correlations into account. Each one of the generated data replicas N_{repl} is then used to create a new fit based on a merit that characterizes the agreement between theory and data for the eight parton PDFs. The result is a new PDF replica $\mathcal{F}[\theta_i]$ for each replica of data D_i. In order to obtain the uncertainty based on the generated list of replicas, the "central value" or the nominal PDF $\langle\mathcal{F}[\theta_i]\rangle$ must be calculated based on the average over all replicas

$$\langle\mathcal{F}[\theta]\rangle = \frac{1}{N_{\text{repl}}}\sum_{i=1}^{N_{\text{repl}}}\mathcal{F}[\theta_i]. \tag{10.14}$$

Once the central value based on all replicas have determined, the uncertainty is given by the quadratic sum of the residuals between each PDF replica and the central PDF according to Equation 10.14

$$\left(\sigma_{\text{PDF}}^{\text{NNPDF}}\right)^2 = \frac{1}{N_{\text{repl}}-1}\sum_{i=1}^{N_{\text{repl}}}\left[\mathcal{F}[\theta_i] - \langle\mathcal{F}[\theta]\rangle\right]^2, \tag{10.15}$$

which basically corresponds to the variance of the replicas with respect to the mean PDF(s).

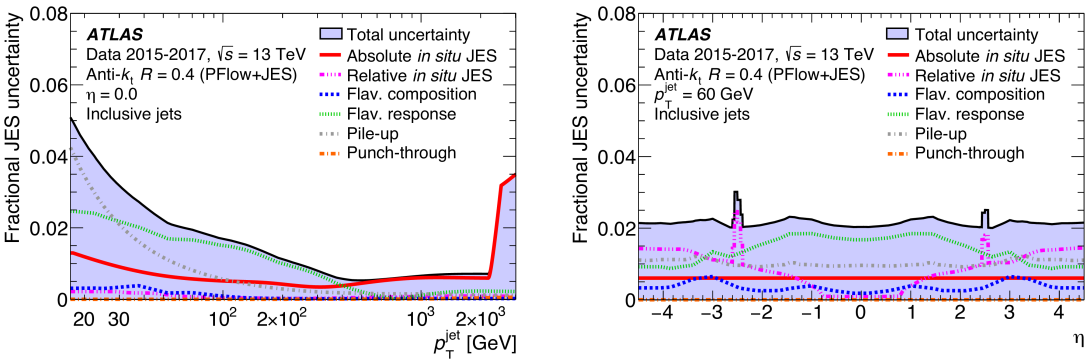
Equation 10.15 provide a prescription of how to compute the uncertainty associated to the PDF in case of the NNPDF-based approach. Both schemes are based on variations of the central or the nominal PDF that propagate into variations of the reconstructed (multi-differential) cross section. Those variations are created during the reconstruction of the cross section by applying an additional event weight $w_{\text{PDF}}^{(i)}$ for each variation. Each weight results in a complete new set of spectra that is then used to calculate the PDF-related uncertainty according to Equation 10.15, respectively.

10.5 Jet Energy Scale and Resolution (JES/JER)

The jet energy scale and jet energy resolution correction was described in detail in Chapter 4.4.3. The systematic uncertainty related to the energy scale and energy resolution of small-R jets follows the "consolidated recommendations" settled in the end of the year 2018 by the

ATLAS 'Jet Energy Calibration and Resolution' subgroup. The final (fractional) uncertainty on the jet energy scale as a function of transverse momentum p_{T} and pseudorapidity η is shown in Figure 10.3.

As a quick reminder, the calibration of the jet energy scale basically fulfills two purposes:



(a) *(b)*

Plot 10.3: Fractional jet energy scale systematic uncertainty components for anti-k_t $R = 0.4$ jets as a function of jet p_{T} at $\eta = 0$ (*left*) and as a function of η at $p_{\mathrm{T}} = 60\,\mathrm{GeV}$, reconstructed from particle-flow objects (*right*). The total uncertainty, determined as the quadrature sum of all components, is shown as a filled region topped by a solid black line. Flavor-dependent components shown here assume a dijet flavor composition (adapted from Ref. [45]).

firstly, the energy scale of the reconstructed detector-level jets does *not* correspond to the energy scale of the underlying particle-level jets that are given by the energy of all *stable* final state particles (except for neutrinos and muons) originating from the hard interaction (including activities in the jet due to the underlying event) in Monte Carlo. The calibration of the jet energy aims to correct the average energy of the reconstructed jets at detector level to match the average energy of jets at truth level. The diverging jet energies at detector and particle level are mostly due to the non-compensation nature of the ATLAS hadronic calorimeter, causing differences between the calorimeter response of electromagnetic and hadronic showers; energy losses due to inactive regions and material in the detector; energy deposits below noise thresholds; energy losses due to longitudinal particle leakage if showers are not fully contained by the calorimeters etc. Less dominant for small-R jets is the correction of the contribution from pile-up to the jet energy. *Secondly*, after the first step, the average energy scale of the reconstructed jets corresponds to truth level in simulation; however, there is still a potential risk of biasing the jets to Monte Carlo because most calibration steps are derived from simulation. To account for this effect, the second part of the energy calibration aims to bring the energy scale of jets in data and simulation into agreement.

Of course, the calibration of the energy scale is associated with an uncertainty that is referred to as jet-energy scale uncertainty. This Uncertainty basically consists of three primary sources: disagreement between data and simulation, the evaluation of the noise term from random cones in zero bias data (this one only becomes significant for jets below 100 GeV and is therefore irrelevant for this analysis) and corrections based on *in situ* dijet p_{T}-balance asymmetry measurement.

To obtain the final JES/JER-related uncertainty, the triple-differential multijet cross section is reconstructed for *each one* of the nuisance parameters that constitute the uncertainty. There are several configurations of the JES uncertainty that use different total number of nuisance parameters, depending on the degree of precision required by the analysis or the physical effects it is sensitive to. In this analysis, the configuration of the *full* set of nuisance parameters is used, which aims at the most precise jet-based measurements. In total, there are more than *one hundred* nuisance parameters. For each one of those parameters, the jet is smeared depending on their transverse momentum as well as their position in the detector in terms of the pseudorapidity. As shown in Figure 10.3, the JES uncertainty (and accordingly the

amount to which the jet is smeared) strongly depends on the p_{T} of the jets (10.3a), while the uncertainty is mostly flat as a function of the pseudorapidity with the already-mentioned exception of the gaps in the transition regions of the detector (10.3b). It is important to note that the smearing according to the JES uncertainty is applied to the four-momentum of the reconstructed jet; hence, the substructure of the jet does remain *unaffected*, which implies that the value of the jet substructure observable before and after the smearing remains unchained too. However, smearing the four-momentum may cause the jet to migrate between different regions of phase space, which may introduce off-diagonal elements in the response matrix or contribute to the efficiency and fake factors if the jet leaves or enters the active phase space volume over which the cross section is measured.

After the smearing according to the uncertainty of each nuisance parameter of the JES uncertainty has been applied, there is a variation of the triple-differential multijet cross section for each one of the respective parameter. *All* nominal spectra are unfolded using the varied response matrix and the relative difference between the truth-level (prior) distribution and the unfolded variation with respect to the nominal spectra is taken as an uncertainty for the respective nuisance parameter.

10.6 Pile-Up Uncertainties

Monte Carlo campaign production usually starts prior to the corresponding data-taking period due to the significant amount of time and computational power required to generate those samples with sufficiently high statistics. However, the exact conditions concerning the average number of interactions per bunch crossing (thenceforth referred to as μ) is usually unknown till the data has been recorded. Therefore, the simulation must be corrected afterwards to accurately model the pile-up conditions present in data as many objects (jets in particular) are sensitive to additional interactions despite pile-up mitigation techniques being applied. This is usually done by reweighting all distributions with a scale factor to mimic the effect of different pile-up conditions. The scale factor, which is used in the pile-up reweighting procedure, is varied from 1.03 to 1.06, following the official ATLAS recommendation. The uncertainty due to pile-up is then simply given by the bin-by-bin subtraction of the two histograms of the up and down variation according to

$$\sigma_i^{\mathrm{Pile-up}} = \frac{N_{\mathrm{Up},i}^{\textsc{Pythia8}} - N_{\mathrm{Down},i}^{\textsc{Pythia8}}}{N_i^{\textsc{Pythia8}}}. \tag{10.16}$$

The uncertainty due to pile-up is almost completely negligible throughout the analysis as no jets below a transverse momentum of 100 GeV are considered.

10.7 Deactivated Tile Modules

During data-taking, it may happen that certain modules of the hadronic tile calorimeter need to get deactivated/decommissioned due to technical problems like, for instance, overheating or leaking. Some of those (permanently) deactivated modules are already included in the respective Monte Carlo simulation; however, other segments of the tile calorimeter may get switched off only during specific runs in data. The tile modules that are deactivated in Run 2 (2015–2018) data and simulation are summarized in Table B.1, which can be found in the appendix of this document.

Based on the ATLAS recommendations, the transverse momentum of the jet is not corrected for jets which may have deposited energy in one of the deactivated tile modules. Instead, the impact of the deactivated tile modules on the measured event shape distributions is evaluated by repeating the measurement and vetoing events with jets directed/pointing at one of the deactivated modules in either data, Pythia (version 8.230), or both. The resulting "varied" distributions are then compared to the nominal data and Monte Carlo (not taking all deactivated tile modules into account) distributions, whereby the data is unfolded based on the response matrix that includes the effect of vetoed tile modules. The uncertainty associated

with the difference concerning deactivated tile modules as well as their impact is defined as the difference between the nominal sample and the distributions including the effect of deactivated modules in simulation and data. This uncertainty turned out to be very small in all regions of phase space.

10.8 Unfolding Bias

As described in Chapter 5, the response matrix – which encodes the correlation between particle- and detector-level information – is derived based on simulation in Monte Carlo. In the scope of this work, the baseline generator is Pythia, which is the official ATLAS recommendation for analyses involving multijet events. Since no Monte Carlo generator is expected to mimic nature perfectly well (especially in the context of QCD-related physics), biases present in the simulation may propagate to the final measurement during unfolding. This results in a different shape of the spectrum after unfolding than what would have been expected for an unbiased result. In order to account for the possible bias in the unfolded cross section measurement, an uncertainty associated with the shape of the unfolded spectrum is determined by using a data-driven closure test, which is introduced in the following paragraph before providing a mathematically rigorous description of the procedure afterwards.

The data-driven closure test is based on the comparison between the shape of the reconstructed spectra for Monte Carlo and data at detector level. Since the measured distribution of the respective observable of interest in data is based on real *pp*-collision events, the basic assumption of the data-driven closure test is that all shape differences between data and simulation observed at detector level are due to an improper modeling of the process at particle level (or because of an inappropriate modeling in the Geant4 simulation of the detector as a whole). The data-driven closure test therefore starts at detector level by reweighting the reconstructed (i.e., detector level) Monte Carlo using a smooth function such that the agreement between data and simulation improves. Following this, the reweighted truth spectrum is folded with a response matrix to obtain the corresponding distribution at detector level (which should be much closer to the distribution that has been measured in data). For the comparison with the reweighted truth-level (prior) Monte Carlo, the obtained reweighted spectra is now unfolded using the very same (nominal) response matrix as used for the actual data (including the potential bias that is supposed to be estimated). The systematic uncertainty associated with the unfolding and the bias in the shape of the truth-level distribution is given by the difference between unfolded reweighted Monte Carlo and the reweighted truth-level Monte Carlo.

Rephrasing the above-stated prescription in mathematical terms, let $d_i = d(x_i)$ denote the measured cross section of the spectrum in data in the ith phase space bin of the support x of the distribution. Similarly, $r_i = r(x_i)$ represents the folded/simulated detector level spectrum for events that pass both, truth *and* reconstructed, selections. Accordingly, $t_i = t(x_i)$ is the spectrum at particle/truth level for events that pass both selections. The correlation between truth and reconstructed events is given by the response matrix R_{ij} that contains the number of events in the simulation that fall in the reconstructed bin i and the truth bin j (this is the reason for the requirement of the jets passing both, truth- and detector-level, selections). Properly normalized, the response matrix $\widetilde{R}_{ij} = R_{ij} / \sum_k R_{kj}$ corresponds to the conditional probability for a truth-level jet in bin j to be reconstructed in bin i, i.e., $r_i = \widetilde{R}_{ij} t_j$. With the input spectra being defined, the next step is to apply the reweighting scheme such that the agreement between $d(x_i)$ and $r(x_i)$ improves. The reweighting can only be meaningful applied to events in simulation that pass both, the truth and reconstructed event selection, because this is the only configuration for which a direct comparison to data is possible. The first step is therefore to apply a *bin-by-bin* correction ϵ_i to data, whereby the *efficiency* ϵ_i is defined as the fraction of events that pass both truth- and detector-level selection $N^{\text{true}\wedge\text{reco}}$ with respect to the number of events that pass the reconstructed selection N^{reco} such that

$$\epsilon_i = \frac{N_i^{\text{true}\wedge\text{reco}}}{N_i^{\text{reco}}}, \tag{10.17}$$

gives the correction factor of the ith bin. This factor is applied *to data* bin-by-bin, which therefore results in a corrected data histogram $\widetilde{d}(x_i) := \epsilon_i d(x_i)$, ensuring that data and simulation agreeing in terms of the efficiency and fake factors. This guarantees that the uncertainty associated with the data-driven closure test is due to the unfolding step. With the reconstructed data spectrum being corrected for the efficiency factors in simulation, the next step is to compute a truth-level prior distribution $\widetilde{t}(x_i)$ such that the folded spectrum $\widetilde{r}(x_i) = \widetilde{R}_{ij}\widetilde{d}(x_i)$ closely matches the reweighted data $\widetilde{d}(x_i)$. The task now is to determine weights such that the aforementioned property is fulfilled. Assumed that the response matrix $\widetilde{R}$ is "mostly" diagonal, the reweighted truth spectrum $\widetilde{t}(x_i)$ can be obtained by simply comparing the reweighted data distribution and the reconstructed spectrum in simulation for each bin, taking the ratio between the two as a weight

$$w_i \approx \frac{\widetilde{d}(x_i)}{r(x_i)}. \tag{10.18}$$

The same weights according to Equation 10.18 are now used to obtain the corrected truth-level distribution $\widetilde{t}(x_i) = w_i t(x_i)$. The uncertainty in the ith bin associated with the data-driven *non*-closure is then given by the difference between the reweighted truth spectrum $\widetilde{t}$ and the reconstructed spectrum $\widetilde{r}(x_i)$ (given by folding $\widetilde{t}$ with $\widetilde{R}_{ij}$) unfolded with the nominal response matrix

$$\sigma_i^{\mathrm{DDNC}} = \frac{f^{\mathrm{nom}}\left(\widetilde{R}_{ij}\widetilde{t}(x_i); \widetilde{d}(x_i), r(x_i)\right) - \widetilde{t}(x_i)}{\widetilde{t}(x_i)}, \tag{10.19}$$

whereby the function f^{nom} denotes some unfolding method based on the nominal response matrix (see Chapter 5).

10.9 Year-to-Year Stability

The cross section measurements presented in this report are based on pp-collision data that has been recorded with the ATLAS detector at the LHC over a period of around three-and-a-half years. Over this time, the conditions of the accelerator as well as measuring apparatus subject to natural fluctuations regarding their operation due to external factors of influence, radiation damage etc. Moreover, the machine was switched off for maintenance work several times over the last few years. The latter point is of particular importance for this analysis because the jet substructure measurements are based on tracks that have been matched to PFlow jets. Since the reconstruction of tracks depends on the alignment of the detector within the magnetic field, possible misalignments of the device could result in biased shapes between the different shutoffs. Those changes of the experimental setup are generally hard to model in simulation; therefore, potential differences are accounted for by directly comparing the unfolded differential cross section measurement for each year $\mathrm{d}\sigma^{15+16}$, $\mathrm{d}\sigma^{17}$ and $\mathrm{d}\sigma^{18}$ with the *full* Run 2 result $\mathrm{d}\sigma^{\mathrm{Run2}}$. The uncertainty associated with the year-to-year stability of the measurement is then given by the maximum (envelope) difference between the individual years and the full measurement.

10.10 Errors on "Jet Topics"

The underlying quark and gluon distributions (see Equation 7.21) are extracted from the measured multi-differential multijet cross section over the two jet mixtures $P^k_{M_\alpha}(\mathcal{O}_{\mathrm{JSS}})$ (in central and forward region) and the reducibility factors $\kappa^k_{M_1 M_\alpha}$, which are also calculated based on the different mixtures, with the reducibility factors being given by the infimum of the likelihood ratio of the two mixtures according to Equation 7.10. Furthermore, the extracted quark/gluon fractions within the respective mixture of jets also directly derive from the measured reducibility factors of the two jet samples. The determination of the uncertainty for the underlying distributions of quark and gluon jets as well as their fractions is based on a variational approach. The procedure starts from the (unfolded) measured distributions in forward and central region ($P^k_{M_{1/2}}$ in Equation 7.19). The unfolded multijet cross section is

varied in each bin of the jet substructure observable for each component of the systematic (ME, PS, Had, JES/JER, PDF etc.) and statistical (Data and MC) uncertainty, giving rise to a set of new distributions in central and forward region from which the reducibility factors get recalculated according to Equation 7.10. This is done by assuming the systematic uncertainties to follow a Gaussian distribution with the center value of the distribution being given by the nominal cross section σ_{nom} and the variance σ_ℓ corresponding to the ℓth respective systematic uncertainty. For each component of the uncertainty, a total number of 1000 toy-experiments is performed (each one resulting in a new, varied differential cross section for the central and forward region), each one resulting in (2×)1000 reducibility factors for one systematic uncertainty. Based on those reducibility factors, the underlying quark and gluon distributions are extracted following the procedure described in Chapter 7. For each variation, the distance with respect to the nominal quark/gluon distributions (obtained using the nominal multijet cross section) are calculated (for each bin) and the 75 % and the 25 % percentiles of the total variation are taken as the uncertainty associated with the respective source of error. Furthermore, the uncertainty-related variation of the reducibility factor $\{\kappa_{ij}^{\sigma_\ell^\pm,k,m}\}_{m\in[1,N_{\text{toy}}]}$, with $N \in \mathbb{N}$ being the number of variations and $\ell \in \{\text{ME, PS, Had, JES/JER, PDF}\}$ with $\sigma^\pm$ being the corresponding upper and lower error, are used to determine the uncertainty of the extracted gluon fraction in each bin of transverse momentum k. For each reducibility factor $\kappa_{ij}^{\sigma_\ell^\pm,k,m}$, the associated fraction of gluon-like jets is determined and the difference with respect to the nominal fraction is calculated. The 75 % and 25 % percentiles are once again used as an uncertainty on the fraction of gluons.

11. Measurement of the Triple-Differential Multijet Cross Section $\frac{1}{\sigma}\frac{\mathrm{d}^3\sigma}{\mathrm{d}p_\mathrm{T}\,\mathrm{d}\eta_\mathrm{rel}\,\mathrm{d}n_\mathrm{SD/trk}}$

The previous chapters introduced the proton-proton collision data that was recorded with the ATLAS experiment at the LHC with a centre-of-mass energy of 13 TeV over a period of three years lasting from 2015 to 2018 (inclusive), which resulted in roughly $139\,\mathrm{fb}^{-1}$ of data. Additionally to the real *pp*-collision data, several jet samples derived from simulation using various state-of-the-art Monte Carlo event generators for multijet production have been presented. Those simulations account for the conditions at the LHC during data-taking in the relevant period of time in terms of beam energy, pile-up etc.; furthermore, a full GEANT4 simulation of the entire ATLAS detector was applied to all MC events to allow for comparison between real data and simulation at detector level and to construct the response matrix, which is needed for the unfolding of the multijet spectra to particle level. Afterwards, the actual event selection was introduced starting with the filtering of collision events by the trigger with signatures considered relevant for this analysis. In the context of this measurement, the signature of interest basically corresponds to the presence of at *least two* jets being present in the event. The appropriate triggers that select hadronic signatures are the higher-level single-jet triggers, which are mostly prescaled such that scale factors must be applied in order to recover the correct spectra. Once the reweighting schemes in data and simulation have been established, the data-to-MC agreement was examined for a selected list of observables like, for instance, the invariant mass, the energy and the pseudorapidity of the reconstructed detector-level jets, as well as some global event observables like the jet multiplicity. Moreover, the transverse momentum p_T and the (relative) rapidity η_rel distribution along with their respective binning (used through this analysis) have been introduced.

This chapter finally turns the attention to the actual measurement of the first class (out of two) of the triple-differential multijet cross section according to

$$\frac{1}{\sigma}\frac{\mathrm{d}^3\sigma}{\mathrm{d}p_\mathrm{T}\,\mathrm{d}\eta_\mathrm{rel}\,\mathrm{d}n_\mathrm{count}} = \frac{1}{N_\mathrm{tot}}\frac{N^\mathrm{multijet}}{\Delta p_\mathrm{T}\,\Delta\eta_\mathrm{rel}\,\Delta n_\mathrm{count}}, \tag{11.1}$$

with respect to a *counting-based* jet substructure observable n_count that is given by the *track multiplicity* n_trk (counting all *tight* tracks with $p_\mathrm{T} > 500\,\mathrm{MeV}$) and the *SD multiplicity* n_SD for different configurations of the Soft Drop algorithm.

The first section (11.1) of this chapter formally introduces the SD multiplicity n_SD along with the parameters it depends on. Thereupon, in the next section (11.2), the different settings of the SD multiplicity in terms of its parameters, which have been measured in the context of this analysis, are introduced along with a motivation of the respective configuration of the SD algorithm with regard to the extraction of the underlying quark/gluon distributions and fractions. To complete the list of counting-based substructure observables, the track multiplicity n_trk and its relation/correlation with n_SD is shown (11.3). Afterwards (11.4), the measured spectra in data *and* simulation are presented for all counting-based observables introduced above. Following the usual analysis workflow, the subsequent section (11.5) takes a look at the response matrix as well as the efficiency and fake factors for the different regions in phase space for all observables as a preparation for the following unfolding step. Finally, this chapter is concluded (11.6) by the final *unfolded* triple-differential multijet cross section for the different counting-based observables including all sources of systematic and statistical

uncertainties introduced in Chapter 10.

11.1 Soft Drop Multiplicity n_{SD}

The observable in question *counts* the number of splittings/branchings inside of a jet by iteratively traversing through the history of angular-ordered emissions. The angular-ordered history of the emissions inside of the jet is obtained by reclustering the jet's constituents using the C/A clustering algorithm (see Chapter 4.4). Following the definition used in ATLAS' recently published *"[m]easurement of the Lund jet plane using charged particles [...]"* [208], an emission is defined to be the softer sub-jet of the proto-jet pair obtained by reversely declustering the clustering history of jet. Additionally, each emission is required to satisfy the Soft Drop (SD) condition, which is why the observable is henceforth referred to as *SD multiplicity* n_{SD}, i.e., the number of valid SD splittings inside of a jet.

The SD multiplicity depends on a number of parameters that allow to control its sensitivity to soft and wide-angled radiation and thereby changing the sensitivity to quark and gluon jets, as well as perturbative and non-perturbative effects. Most of those parameters directly inherit from the underlying SD algorithm; hence, gaining a basic intuition for the "Soft Drop declustering" is among the first objectives of this chapter. This includes a quick, superficial review of the "Mass Drop Tagger" (MDT), which is a predecessor and a special case of the SD declustering scheme introduced in this section. This will be followed by a generalization of the SD algorithm to obtain a prescription suited to count the number of valid SD emissions inside of a jet by taking its entire clustering history into account. This algorithm is referred to as "Iterated Soft Drop" (ISD). With the ISD being introduced, all necessary technical prerequisites are established such that the subsequent section mainly focus on gaining an intuition of the substructure observable in the context of quark-gluon discrimination. This includes a simple analysis of the observable at leading-logarithm accuracy (which has been adopted from the original publication that proposed the SD multiplicity in Ref. [26]) to accommodate for the statement made in Chapter 7.3 regarding the mutual irreducibility of Poisson-like distributed (counting) random variables. This is important to conceptually demonstrate that the extraction of the underlying quark and gluon distributions as well as the quark/gluon fractions based on the SD multiplicity is indeed justified.

After the observable has been introduced, all subsequent sections are dedicated to its measurement with the ATLAS detector.

11.1.1 "Mass- and Soft-Drop"

Over the last few years, probing the substructure of jets has gained significant importance regarding its potential to, e.g., discriminate quark- and gluon-initialized jets or other particles. Techniques that try to identify and classify jets originating from different fundamental processes or particles (with the aim of discriminating them from background) are usually referred to as *(jet) taggers*. Common examples are the already mentioned quark-gluon discriminators, top, W/Z or even Higgs taggers etc. Besides that, a better understanding of the internal structure of jets, i.e., the characteristic features associated to specific underlying processes, becoming increasingly important in order to efficiently mitigate contributions from, for example, pile-up contamination and other underlying activity in the event. Methods that use jet substructure information to remove soft activity inside the active area of the jet (which is not associated to the hard scatter of the event) are referred to as *(jet) groomers*.

Besides the growing relevance concerning the phenomenological aspects of jet substructure with regard to experimental applications, analytic calculations based on perturbative QCD of variables (that capture certain features of jets), likewise, has made significant progress in the same period of time. Those *first-principle* studies have made a major contribution to the theoretical understanding of the key characteristics of many substructure methods as well as their implications regarding jet tagging and jet grooming in equal measure.

As part of this effort, special attention has been paid to the jet mass and its characteristic features regarding the most commonly used taggers and groomers, i.e., pruning [209, 210], trimming [211] and the aforementioned Mass Drop Tagger (MDT) [212]. All previously mentioned

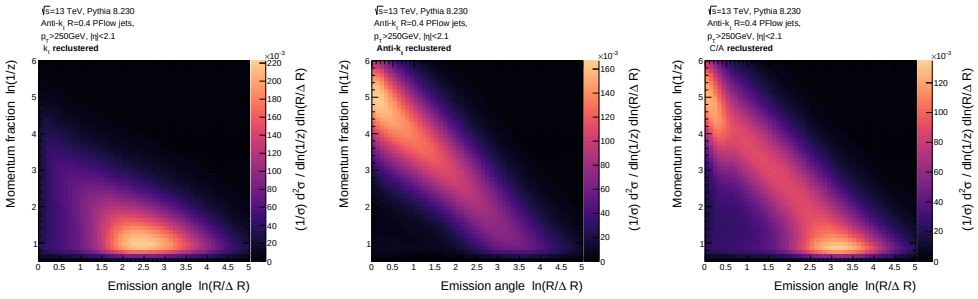
methods share a common philosophy, that is, they try to identify *sub*-jets within the original jet that is interpreted as a proxy for radiation emitted during the formation of the jet (see introduction to parton showers in Chapter 2.4). If the momentum fraction of the sub-jet with respect to the jet's (global) momentum is below some lower threshold value, it is most-likely due underlying soft activity in the event or hadronization effects and consequently removed from the jet. This, of course, must be done with great care as it may significantly affects the shape of the mass distribution (as well as other substructure observables); the objective of the grooming procedure therefore is to keep the contributions from signal to the distribution mostly unaltered while reducing those due to background processes.

While being conceptually similar, all three groomers/taggers mentioned above differ considerably regarding the degree to which soft radiation is removed from the original jets. It is a matter of common knowledge that pruning as well as trimming act (quite) "aggressively" on signal and background alike.

All three methods have been subject of intensive studies regarding their analytic behavior with respect to logarithmic resummation. As a result, those investigations lead to the development of a *modified* version of the MDT (mMDT), which contrast the other methods by the property of only involving single logarithms (no Sudakov double logarithms) and being free of the complex class of non-global logarithms [213], which introduce *highly* non-trivial correlations between jets and the event as a *whole*. This property makes the mMDT ideally suited for the analytic calculation of jet-related properties as it significantly increases the sensitivity of the final distributions to perturbative(="calculable") effects.

The concrete implementation of the mMDT algorithm is not subject of this section because the methodically very similar *Soft Drop* declustering scheme is explained in detail hereinafter. As it will be pointed out later, the mMDT can be recovered from the SD procedure given an appropriate configuration of its parameters.

The *Soft Drop declustering* method can be considered as a generalization of the aforemen-



(a) k_t *(b)* Anti-k_t *(c)* C/A

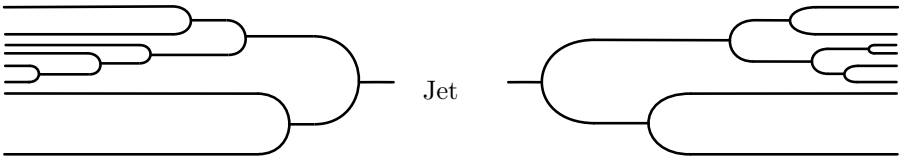
Plot 11.1: The Lund (emission) jet plane for all jets with $p_{\mathrm{T}} > 250\,\mathrm{GeV}$ for different recombination schemes according to the k_t (*left*) the anti-k_t (*middle*) and the Cambridge-Aachen C/A (*right*) jet clustering algorithm. (Keep in mind that the basic input to all figures is a anti-k_t jet.)

tioned mMDT that removes wide-angle soft radiation from a jet to mitigate contamination due to initial state radiation, underlying event activity and multiple hadron scatterings that give rise to pile-up. To cut a long story short, Soft Drop declustering removes soft emissions inside of a jet of radius R (radius parameter of the sequential recombination algorithm) based on the following condition

$$[0, 0.5] \ni z := \frac{\min\left(p_{\mathrm{T}j_1}, p_{\mathrm{T}j_2}\right)}{p_{\mathrm{T}j_1} + p_{\mathrm{T}j_2}} > z_{\mathrm{cut}} \left(\frac{\Delta R_{12}}{R}\right)^{\beta}, \tag{11.2}$$

whereby z_{cut} defines a *lower* threshold on the momentum fraction of the softer splitting daughter – which is assumed to correspond to the emission –, ΔR_{12} the emission angle between the two daughters and $\beta \in \mathbb{R}$ corresponding to a free parameter that allows to control the degree of grooming as a function of the radiation angle.

In further detail, the input to the declustering algorithm is a jet that has been reconstructed using some clustering algorithm (see Chapter 4.4.2). If the jet was reconstructed with *any* other clustering scheme besides the Cambridge-Aachen one (in this analysis, all jets are reconstructed using the anti-k_t algorithm), the constituents of the jet *must* be reclustered using the C/A algorithm to obtain a pairwise clustering tree with an angular-ordered structure[1]. Due to the already stated benefits, jets are typically reconstructed using the anti-k_t clustering algorithm (see Chapter 4.4.2), which collects inputs according to decreasing p_{T} of the emission. This results in very regular jet shapes, but the clustering history does not represent the intuitive idea of radiation being emitted during the showering step with decreasing collimation angles due to color coherence and Heisenberg's uncertainty principle. The actual SD declustering is therefore very often preceded by an *re*clustering step that results in an angular ordered history of the emissions. This situation is illustrated in Figure 11.1 for all sequential clustering algorithms introduced in Chapter 4.4.2. A jet (its constituents) that was reconstructed using the anti-k_t algorithm is reclustered using the Cambridge-Aachen algorithm. It is very important to keep



(a) anti-k_t *(b)* C/A

Fig. 11.1: Two clustering histories according to different clustering algorithms of the *very same* jet: anti-*kt* (*left*) and C/A (*right*).

in mind that the reclustering step does *not* change the overall jet (global jet-related properties are not affected by the reclustering), but results in a *re*interpretation of the clustering history that resembles the expectations regarding the order in which radiation is emitted during shower formation. Accordingly, all jet substructure observables that do not explicitly depend on the emission history of the jet are invariant with respect to the reclustering step. This can be seen by directly comparing the two clustering diagrams in Figure 11.1a and Figure 11.1b that show the clustering history of the *very same* jet according to the anti-k_t (*left*) and the C/A (*right*) algorithm, respectively. Both diagrams start from the exact same configuration of inputs (compare left- and right-most configuration in Figure 11.1), but the cluster sequence differs with respect to both schemes.

After an angular-ordered emission history of the jet is available, the SD declustering algorithm traverses backwards throughout the clustering tree and checks the pair(s) of proto-jets for the SD condition according to Equation 11.2. The individual steps of the SD algorithm read as follows:

Definition 3 (**Soft Drop Declustering [214]**).

1. *Break the jet J into two sub-jets by undoing the last stage of C/A clustering. Label the resulting two sub-jets as j_1 and j_2.*

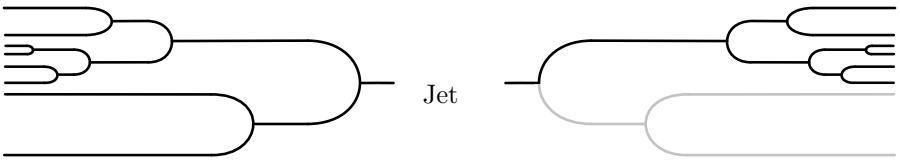
2. *If the sub-jets passes the soft drop condition [(see Equation 11.2)], then deem j to be the final soft-drop jet.*

3. *Otherwise, redefine j to be equal to sub-jet with larger p_{T} and iterate the procedure*

[1]This is, as a matter of fact, the recommended procedure because jets that have been clustered using the anti-k_t algorithm are guaranteed to contain all their constituents within the defined jet radius R, i.e., $\Delta R_j \leq R$. In case of the C/A (or the k_t) algorithm, however, the jet shape is highly irregular as can been seen in Figure 11.2a (4.2a), which may give rise to constituents with $\Delta R_i > R$, causing undesired behavior in the SD condition in accordance with Equation 11.2.

4. *If j is a singleton and can no longer be declustered, then one can either remove j from consideration ("tagging mode") or leave j as the final soft-drop jet ("grooming mode").*

It is important to note that the SD algorithm is *terminated* if the sub-jets pass the SD condition (step 2) and does *not* iteratively traverses throughout the entire clustering tree – this is the main difference with respect to the iterative version that is used to reconstruct the SD multiplicity n_{SD}, which is introduced in the next section. The result of the SD declustering is shown in Figure 11.2 pictographically. The SD algorithm can be categorized into a *tagging*



(a) C/A clustered *(b)* SD groomed

Fig. 11.2: A C/A reclustered/reconstructed jet (*left*) and its groomed counterpart using Soft Drop declustering (*right*).

($\beta < 0$) and a *grooming* ($\beta > 0$) mode of operation depending on the chosen value for β, which controls the sensitivity to the angular components of the emissions. To better understand the effect of different configurations in $(z_{\mathrm{cut}}, \beta)$-space of the SD algorithm, it is beneficial to transform Equation 11.2 to a different representation that allows to easily visualized the (valid) emission plane. The (primary) Lund Jet Plane (LJP) [215] is perfectly suited to study the emission phase space within jets by showing the correlation between the emission angles $\ln R/\Delta R_{12}$ and the momentum fraction $\ln 1/z$. Rearranging Equation 11.2 in terms of $\ln 1/z$ and $\ln R/\Delta R_{12}$, the SD condition can be interpreted as plane equation that defines a (selection) area/boundary of valid emissions in the LJP according to

$$\ln \frac{1}{z} < \beta \ln \frac{R}{\Delta R_{12}} + \ln \frac{1}{z_{\mathrm{cut}}}. \tag{11.3}$$

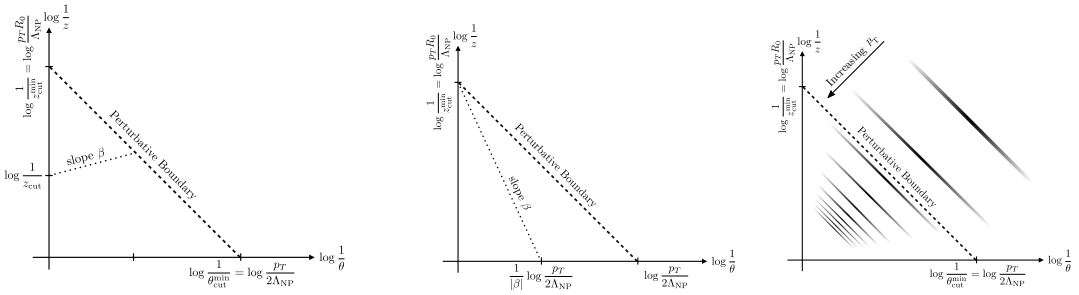
In Equation 11.3, β corresponds to the slope of the decision boundary with z_{cut} being the (lower) axis intercept. This decision boundary is shown in Figure 11.3 for $\beta > 0$ (11.3a) and for $\beta < 0$ (11.3b) along with the perturbative boundary above which the emission plane is mostly dominated by hadronization. The perturbative limit is (roughly) defined by the energy scale/parameter $\Lambda_{\mathrm{QCD}} \sim 200\,\mathrm{MeV}$, i.e., if the k_t of the emission is close to the hadronization scale $k_t \sim \Lambda_{\mathrm{QCD}}$. With $k_t = p_{\mathrm{T}} \Delta R \overset{!}{=} \Lambda_{\mathrm{QCD}}$, the two limits are then given by $\ln 1/\Delta R \sim \ln \Lambda_{\mathrm{QCD}}/p_{\mathrm{T}}$ and $\ln 1/z \sim \ln p_{\mathrm{T}} \Delta R/\Lambda_{\mathrm{QCD}}$.

The next section extends the SD algorithm by iterating through the entire clustering history of the reconstructed jet. This will also introduce the final instructions for calculating the Soft Drop multiplicity n_{SD}, which is measured in this analysis.

11.1.2 "Iterated Soft Drop"

The Soft Drop declustering algorithm introduced in Section 11.1.1 terminates as soon as a hard branching in the clustering history of the jet was found. The *Soft Drop multiplicity* n_{SD} is measured based on a modified version of standard SD algorithm in which the Soft Drop condition (Equation 11.2) is *Iteratively* (ISD) applied to all branchings/nodes in the clustering history of the jet down to a lower angular threshold cut θ_{cut} on the emission angle[2]. The lower angular cut is an additional parameter with respect to the "standard" version of the SD

[2]Keep in mind that the jet has been reclustered using C/A; hence, all emission are ordered in *decreasing* emission angles, i.e., if the emission angle of one emission is below θ_{cut}, this accounts for all following emissions as well.



(a) $\beta > 0$ *(b)* $\beta < 0$ *(c)* k_t^{cut}

Fig. 11.3: Decision boundary in $(\ln 1/\Delta R, \ln 1/z)$-space as defined by the SD condition according to Equation 11.2 and 11.3 for $\beta > 0$ (*left*), $\beta < 0$ (*middle*) as well as a lower k_t cut (*right*) including the perturbative limit (adapted and modified from Ref. 26, Fig. 8, p. 16).

algorithm and is freely to choose. It can be used to reduce the sensitivity of the ISD-based counting observable n_{SD} to hadronization effects by taking the perturbative boundary in the emission plane in Figure 11.3 into account. As shown before, requiring the emissions' k_t to be above the hadronization scale Λ_{QCD}, implies an angular cut of $\ln 1/\theta_{\text{cut}} \sim \ln \Lambda_{\text{QCD}}/p_{\text{T}}$, which can directly visualized in the emission plane.

With this additional parameter, the iterated version of the Soft Drop declustering algorithm reads as follows:

Definition 4 (Iterated SD Declustering/ SD multiplicity n_{SD} [26]).

1. *"Set the counter [n_{SD}] equal to 0. Traverse the C/A clustering tree, beginning at the trunk [(hard splitting)] and sequentially examining each branching.*

2. *Upon arriving at a branching into sub-jets [j_1] and [j_2], check whether the branching angle satisfies*

$$\Delta R_{12} > \theta_{\text{cut}}. \tag{11.4}$$

   If not, the algorithm terminates [because all subsequent emissions are below θ_{cut} due to the angular ordering of the jet clustering.]

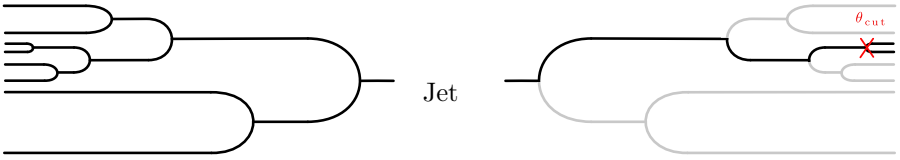
3. *If $\Delta R_{12} > \theta_{\text{cut}}$, then check whether the [S]oft [D]rop condition [(see Equation 11.2)] is satisfied [...] If not, return to Step 2[.] on the harder of sub-jets [j_1] and [j_2].*

4. *If the soft drop condition is satisfied, define*

$$z_n := z, \quad \theta_n := \Delta R_{12}. \tag{11.5}$$

   Then increment [$n_{\text{SD}} \to n_{\text{SD}} + 1$] and return to Step 2[.] on the harder of sub-jets [j_1] and [j_2]."

Figure 11.4a shows an example for a possible clustering diagram of a C/A reclustered jet and the iterated Soft Drop declustering applied to it. In each splitting node in Figure 11.4b, the SD multiplicity n_{SD} (counter) is incremented by one until the emission angle falls below the angular cut θ_{cut}, in which case the algorithm terminates. In case of $\theta_{\text{cut}} = 0$, the algorithm iterates through the entire clustering tree of the jet checking all emissions against the SD condition according to Equation 11.2.



(a) C/A clustered *(b)* ISD

Fig. 11.4: A C/A jet (*left*) and the ISD algorithm applied (*right*).

11.1.3 Soft Drop Multiplicity n_{SD}

In summary, the SD multiplicity n_{SD} is based on the iterative version of the Soft Drop algorithm (see Section 11.1.1 and 11.1.2). For *each* branching node in the clustering history of the C/A (reclustered) jet, the SD condition according to Equation 11.2 is evaluated for the two current proto-jets. Given the Heaviside function Θ, the SD multiplicity n_{SD} can compactly be written as

$$n_{\mathrm{SD}}(z_{\mathrm{cut}}, \beta, \theta_{\mathrm{cut}}) = \sum_{i=1}^{N_{\mathrm{nodes}}} \Theta\left\{ z_i - z_{\mathrm{cut}} \left(\frac{\Delta R_i}{R} \right)^{\beta} \right\} \Theta\left\{ \Delta R_i - \theta_{\mathrm{cut}} \right\}, \tag{11.6}$$

with N_{nodes} being the number of branching nodes in the clustering history of the declustered jet and z_i and ΔR_i denoting the momentum fraction as well as the emission angle of the ith splitting in the jet.

To reduce the effect on hadronization effects (and to increase sensitivity to the perturbative regime in phase space), a modified version of Equation 11.6 is introduced that includes an *explicit* cut on the k_t of the emissions according to $p_{\mathrm{T}}^{\mathrm{emission}} \Delta R_i > k_t^{\mathrm{cut}}$, whereby $p_{\mathrm{T}}^{\mathrm{emission}} := \min\left(p_{\mathrm{T},j_1}, p_{\mathrm{T},j_1}\right)$ returns the transverse momentum of the softer *proto*-jet that is assumed to correspond to the emission, with respect to the jet's trunk.

The transverse momentum of the emission k_t is given by $k_t \sim p_{\mathrm{T}j1}^{\mathrm{emission}} \Delta R(j_1, j_2) = z p_{\mathrm{T}j_{\mathrm{m}}} \Delta R(j_1, j_2)$, whereby $p_{\mathrm{T}j_{\mathrm{m}}}$ denotes the transverse momentum of the "mother" of j_1 and j_2, i.e., $j_m \to j_1\, j_2$. The k_t-cut can be interpreted in terms of the $(\ln 1/z, \ln 1/\Delta R)$ emission plane in Figure 11.3. Requiring $k_t > k_t^{\mathrm{cut}}$ and transforming into an appropriate representation, the lower k_t-cut can be interpreted as a decision boundary according to

$$\ln \frac{1}{z} < (-1) \ln \frac{1}{\Delta R} + \ln \frac{p_{\mathrm{T}}^{\mathrm{emission}}}{k_t^{\mathrm{cut}}}, \tag{11.7}$$

which – in contrast to Equation 11.3 – is variable with respect to the lower cut on the momentum fraction of the proto jet. For high-p_{T} jets, the trunk (usually) is of high momentum too, which implies a larger phase space. For low(er)-p_{T} jet emissions, the right-hand side of Equation 11.7 *decreases* and puts a tighter constrain on the region of possible emissions. In general, the effect of a k_t-cut corresponds to setting $\beta = -1$ as well as choosing a variable lower momentum cut $z_{\mathrm{cut}}(p_{\mathrm{T}}) = k_t^{\mathrm{cut}}/p_{\mathrm{T}}$ as can directly be extracted by comparing ters and coefficients in Equation 11.7 with Equation 11.3.

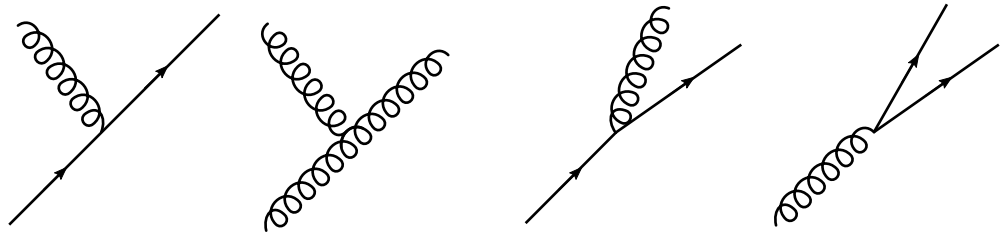
11.1.4 LL Calculation[3]

The previous section introduced the SD multiplicity n_{SD}, i.e, the *counting-based* substructure observable that is used to gain sensitivity to quark- and gluon-like jets already at the level of the differential multijet cross section measurement. As was already anticipated (but not proven) in Chapter 7 in the context of "topic" extraction, the probability distribution over the SD multiplicity of a jet follows a Poisson distribution (at LL), which results in the underlying quark/gluon $P_{q/g}^{k}(n_{\mathrm{SD}})$ distributions being mutually irreducible and therefore in general extractable using the unmixing method introduced in Chapter 7.2. This section

[3]The following section is not necessary for the general understanding of the analysis presented in this thesis and may be safely skipped by the reader.

will make up for the missed proof of the Poisson-like nature of the SD multiplicity (many counting-based observables in general) by providing a step-by-step derivation of the normalized cross section $(1/\sigma)\ \mathrm{d}\sigma^{\mathrm{LL}}/\mathrm{d}n_{\mathrm{SD}}$ at leading-logarithm accuracy. The major part of the following statements have been adopted (either verbatim or in spirit) from the original publication *"Casimir Meets Poisson: Improved Quark/Gluon Discrimination with Counting Observables"* by Christopher Frye *et al.* that can be found in Ref. [26].

For a calculation at LL accuracy, "flavor-changing" splittings like the ones shown in Figure

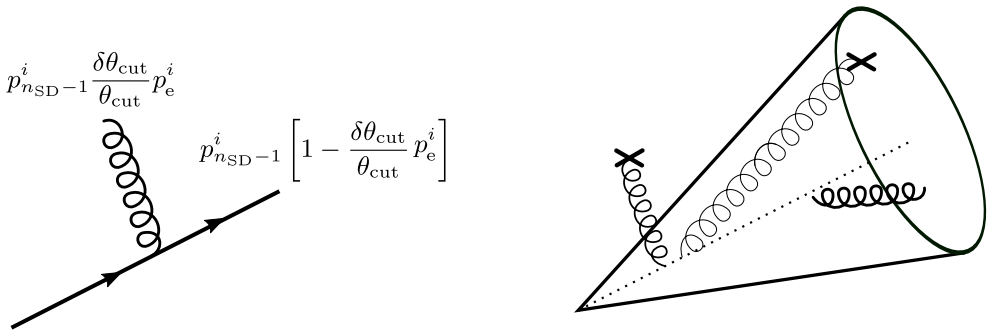


(a) LL+NLL *(b)* LL+NLL *(c)* NLL *(d)* NLL

Fig. 11.5: Different parton splittings that contribute only at LL (11.5a, 11.5b) and NLL (11.5a–11.5d) accuracy

11.5c and 11.5d are *not* present, but only soft and collinear emissions of gluons as shown in Figure 11.5a and 11.5b contribute. Because the emissions are very soft ($z \ll 1$), they can be visualized as a sequence of emissions along the jet's trunk (see Figure 11.6b) such that energy losses of the (main) radiator are negligible and the process can be visualized as a consecutive sequence of colored emissions.

To obtain the distribution over n_{SD}, the probability $p^i_{n_{\mathrm{SD}}}(\theta_{\mathrm{cut}})$ for some flavor i to observe



(a) Emission tree *(b)* Two emissions

Fig. 11.6: Probability of a valid and invalid SD emission (*left*) as well as an illustrative representation of gluon emissions at LL (*right*).

n_{SD} SD emissions for a lower angular cut θ_{cut} in the ISD algorithm (see Section 11.1.2) must be computed – the dependency of the other SD parameters, β and z_{cut}, will be implicitly included through the SD condition that enters the integral. Of course, the probability distribution must satisfy the normalization condition $\sum_{n_{\mathrm{SD}}=1}^{\infty} p^i_{n_{\mathrm{SD}}}(\theta_{\mathrm{cut}}) = 1$, whereby a sum instead of an integral was used because the support of $p^i_{n_{\mathrm{SD}}}$ is discrete (counting observable) $\mathrm{supp}(p^i_{n_{\mathrm{SD}}}) \subset \mathbb{N} \cup \{0\}$.

The probability distribution $p^i_{n_{\mathrm{SD}}}(\theta_{\mathrm{cut}})$ can be obtained by a set of evolution equations that describe the dependence of $p^i_{n_{\mathrm{SD}}}$ with respect to changes in the lower angular cut, i.e., the change of the observed number of SD emissions n_{SD} for $\theta_{\mathrm{cut}} - \delta\theta_{\mathrm{cut}}$. To get the evolution equations, consider a shift in the angular cut $\theta_{\mathrm{cut}} \to \theta_{\mathrm{cut}} - \delta\theta_{\mathrm{cut}}$. If there is a valid SD emission in the interval $[\theta_{\mathrm{cut}} - \delta\theta_{\mathrm{cut}}, \theta_{\mathrm{cut}}]$, the SD multiplicity will increment by one $n_{\mathrm{SD}} \to n_{\mathrm{SD}} + 1$; otherwise, it remains unchanged $n_{\mathrm{SD}} \to n_{\mathrm{SD}}$.

The statement above can be expressed mathematically by a "decision" tree according to

Figure 11.6a that give the probability of having a *valid* emission, i.e., passing the SD condition according to Equation 11.2, within the interval $[\theta_{\text{cut}} - \delta\theta_{\text{cut}}, \theta_{\text{cut}}]$

$$\begin{aligned} p^i_{n_{\text{SD}}}(\theta_{\text{cut}} - \delta\theta_{\text{cut}}) = \; & p^i_{n_{\text{SD}}-1}(\theta_{\text{cut}}) \frac{\delta\theta_{\text{cut}}}{\theta_{\text{cut}}} p^i_{\text{pass}} \\ & + p^i_{n_{\text{SD}}-1}(\theta_{\text{cut}}) \left[1 - \frac{\delta\theta_{\text{cut}}}{\theta_{\text{cut}}} p^i_{\text{pass}}\right]. \end{aligned} \tag{11.8}$$

For an emission to be valid, it must satisfy the SD condition (henceforth refereed to as *(valid) SD emission*); hence, the probability p^i_{pass} must vanish if there was an emission *not* passing the SD condition. This is archived by the theta function

$$\Theta_{\text{SD}}(z, \theta) := \Theta\left(z - z_{\text{cut}} \left(\theta/R\right)^{\beta}\right), \tag{11.9}$$

which vanishes if its argument is below zero(=SD condition fails). With the Heaviside step function and the splitting kernels $P_{i\to i}$ for no changing flavors (at LL accuracy), the probability according to Equation 11.8 can be written as

$$\begin{aligned} p^i_{n_{\text{SD}}}(\theta_{\text{cut}} - \delta\theta_{\text{cut}}) = p^i_{n_{\text{SD}}-1}(\theta_{\text{cut}}) \frac{\delta\theta_{\text{cut}}}{\theta_{\text{cut}}} \int_0^{1/2} \mathrm{d}z \, \frac{\alpha_{\text{S}}(z\theta_{\text{cut}} p_{\text{T}})}{\pi} P_{i\to i}(z)\Theta(z, \theta_{\text{cut}}) \\ + p^i_{n_{\text{SD}}}(\theta_{\text{cut}}) \left[1 - \frac{\delta\theta_{\text{cut}}}{\theta_{\text{cut}}} \int_0^{1/2} \mathrm{d}z \, \frac{\alpha_{\text{S}}(z\theta_{\text{cut}} p_{\text{T}})}{\pi} P_{i\to i}(z)\Theta(z, \theta_{\text{cut}})\right], \end{aligned} \tag{11.10}$$

whereby the strong coupling α_{S} runs with the 1-loop β-function at LL accuracy. Rearranging Equation 11.10 results in the following linear first-order differential equation

$$\frac{\mathrm{d}p^i_{n_{\text{SD}}}}{\mathrm{d}\theta_{\text{cut}}} = \frac{p^i_{n_{\text{SD}}}(\theta_{\text{cut}}) - p^i_{n_{\text{SD}}-1}(\theta_{\text{cut}})}{\theta_{\text{cut}}} \int_0^{1/2} \mathrm{d}z \, \frac{\alpha_{\text{S}}(z\theta_{\text{cut}} p_{\text{T}})}{\pi} P_{i\to i}(z)\Theta(z, \theta_{\text{cut}}), \tag{11.11}$$

which describes the change of the probability distribution $p^i_{n_{\text{SD}}}$ for variations in θ_{cut}. Furthermore, there is a boundary condition $p^i_0(R) = \delta_{n_{\text{SD}},0}$ to Equation 11.11 that constraints all emissions to be within the jet's radius R as shown in Figure 11.6b. The probability, of having no (valid) emissions at all between R and θ_{cut} is simply given by

$$p^i_0(\theta_{\text{cut}}) = e^{-I_{i\to i}(\theta_{\text{cut}}, R)}, \tag{11.12}$$

with $I_{i\to i}$ being defined as

$$I_{i\to i}(\theta_1, \theta_2) = \int_{\theta_1}^{\theta_2} \frac{\mathrm{d}\theta}{\theta} \int_0^{1/2} \mathrm{d}z \, \frac{\alpha_{\text{S}}(z\theta_{\text{cut}} p_{\text{T}})}{\pi} P_{i\to i}\Theta_{\text{SD}}(z, \theta), \tag{11.13}$$

whereby Equation 11.12 basically corresponds to the Sudakov form factor that was introduced in Chapter 2.4.1. Now, the probability of observing *at least* one SD emission $p^i_{n_{\text{SD}}>0}$ can be written as a recursive relation in terms of the probability of already measured $n_{\text{SD}} - 1$ emissions

$$p^i_{n_{\text{SD}}>0}(\theta_{\text{cut}}) = \int_{\theta_{\text{cut}}}^{R} \frac{\mathrm{d}\theta}{\theta} \, e^{-I_{i\to i}(\theta_{\text{cut}}, \theta)} \left\{ \int_0^{1/2} \mathrm{d}z \, \frac{\alpha_{\text{S}}(z\theta_{\text{cut}} p_{\text{T}})}{\pi} P_{i\to i}(z)\Theta_{\text{SD}}(z, \theta) \right\} p^i_{n_{\text{SD}}-1}(\theta). \tag{11.14}$$

Equation 11.14 can be used iteratively to obtain higher SD multiplicities. To illustrate the iterative character, consider the first step of the recursion with only one SD emission $n_{\text{SD}} = 1$

$$\begin{aligned} p^i_{n_{\text{SD}}=1}(\theta_{\text{cut}}) = & \int_{\theta_{\text{cut}}}^{R} \frac{\mathrm{d}\theta}{\theta} \, e^{-I_{i\to i}(\theta_{\text{cut}}, \theta)} \\ & \times \left\{ \int_0^{1/2} \mathrm{d}z \, \frac{\alpha_{\text{S}}(z\theta_{\text{cut}} p_{\text{T}})}{\pi} P_{i\to i}(z)\Theta_{\text{SD}}(z, \theta) \right\} e^{-I_{i\to i}(\theta, R)} \qquad (11.15) \\ = & \, e^{-I_{i\to i}(\theta_{\text{cut}}, R)} I_{i\to i}(\theta_{\text{cut}}, R). \qquad (11.16) \end{aligned}$$

The same can be done recursively for all values of n_{SD} according to $p^i_{n_{\mathrm{SD}}}(\theta_{\mathrm{cut}}) = e^{-I_{i\to i}(\theta_{\mathrm{cut}},R)} \prod_{n=1}^{n_{\mathrm{SD}}} p^i_n$, which results in the following expression

$$p^i_{n_{\mathrm{SD}}}(\theta_{\mathrm{cut}}) = \frac{1}{n_{\mathrm{SD}}!} \left[I_{i\to i}(\theta_{\mathrm{cut}}, R)\right]^{n_{\mathrm{SD}}} e^{-I_{i\to i}(\theta_{\mathrm{cut}},R)}, \tag{11.17}$$

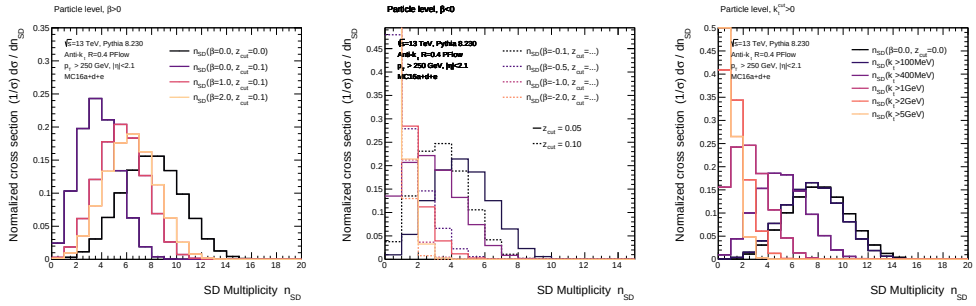
which finally shows that the SD multiplicity is described by a Poisson distribution with a mean value of $I_{i\to i}(\theta_{\mathrm{cut}}, R)$ at LL accuracy.

11.2 Preparations for Measuring n_{SD}

The previous sections introduced the SD multiplicity as a "simple" example for a counting-based jet substructure observable that does not suffer from Casimir scaling and can be used as a quark/gluon discriminator. The observable depends on a number of parameters (z_{cut}, β, θ_{cut}) that allow to increase (or decrease) sensitivity to momentum and/or angular components of the radiation. This section, therefore, introduces the two configurations of the SD multiplicity that have been measured at the ATLAS experiment within the context of this analysis.

Before presenting the configurations of the SD multiplicity measured in this analysis, it is important to gain an intuition of how the different parameters of the ISD algorithm affect the distribution of the SD multiplicity as well as the discrimination power between quark- and gluon-initialized showers. This was (partly) already done in Section 11.1.1 (see Figure 11.3); however, the following considerations focus on the *actual* distributions derived from the nominal Monte Carlo sample at *particle level* for various settings of the ISD algorithm.

A small selection of distributions for different configurations of the SD multiplicity is shown in Figure 11.2, subdivided into distributions obtained for $\beta \geq 0$ (11.2a), $\beta < 0$ (11.2b) and a non-vanishing k_t-cut (11.2c), with the latter one basically corresponding to $\beta < 0$ with a p_{T}-dependent momentum fraction $z_{\mathrm{cut}} = z_{\mathrm{cut}}(p_{\mathrm{cut}})$ (see Equation 11.7). The distributions in



(a) $\beta \geq 0$ *(b)* $\beta < 0$ *(c)* $k_t^{\mathrm{cut}} > 0$

Plot 11.2: Distribution of the SD multiplicity n_{SD} for different configurations of the I(SD) algorithm: positive (*left*) and negative (*middle*) values of β for $z_{\mathrm{cut}} = 0.1$ and $k_t^{\mathrm{cut}} = 0$, as well as for a set of k_t (*right*) cuts with $\beta = z_{\mathrm{cut}} = 0$.

Figure 11.2 are *inclusive* in transverse momentum p_{T}, pseudorapidity as well as (light-)quarks and gluons (the individual quark/gluon distributions will be studied later).

Starting with $\beta \geq 0$ in Figure 11.2a, the distribution corresponding to the default configuration ($z_{\mathrm{cut}} = \beta = 0$), of course, comes with the largest number of emissions on average because all emissions are counted. Increasing the momentum fraction to $z_{\mathrm{cut}} = 0.1$ while keeping the observable insensitive to the emission angle ($\beta = 0$) of the radiation results in an *aggressive* grooming of the jet that gets rid of all soft emissions with a momentum fraction below 10 % with respect to the harder branch of the splitting. Gradually increasing sensitivity to the angular components of the radiation up to $\beta \to 2$ loosens the SD condition such that collinear emissions are more likely to pass even if the momentum fraction is below the aforementioned 10 %. For $\beta \to \infty$, the SD multiplicity corresponds to the default configuration independent of the chosen value for z_{cut}.

For $\beta < 0$ in Figure 11.2b, the situation is different. There are two distributions for each β value: one for $z_{\text{cut}} = 0.005$ (solid line) and one for $z_{\text{cut}} = 0.01$ (dashed line). For negative β, very collinear emissions are removed from the jet, which results in the distributions shifting towards smaller multiplicity values for β being more negative; increasing z_{cut}, additionally tightens the SD condition and results in less (valid) SD emissions on average.

Last but not least, the distributions for a k_t^{cut}-cut in Figure 11.2c look similar to the ones for $\beta < 0$ as expected from Equation 11.7. The variable momentum cut, however, loosens the SD condition, resulting in broader distributions compared to the corresponding spectra in case of $\beta < 0$ as the variable lower momentum cut allows for more freedom.

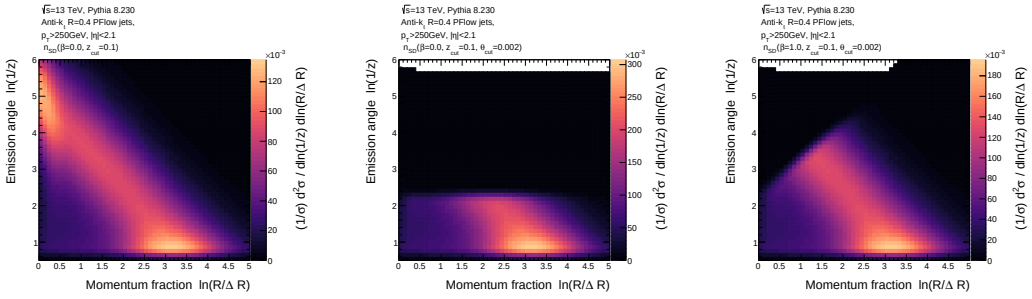
Now, turning to the configurations of the SD multiplicity relevant for this analysis. In total, *two different configurations of the SD multiplicity are measured*:

- $n_{\text{SD}}(z_{\text{cut}} = 0.1, \beta = 1.0, \theta_{\text{cut}} = 0.002)$,
- $n_{\text{SD}}(z_{\text{cut}} = 0.0, \beta = 0.0, k_t \geq 4.0\,\text{GeV})$,

whereby $k_t^{\text{cut}} = 0\,\text{GeV}$ is implicitly assumed if not explicitly mentioned otherwise[4].

As described in the previous section, the different configurations of the ISD algorithm can be interpreted as a "cut" or decision boundary in emission phase space. According to Equation 11.3, β controls the "slope" of the lower decision boundary and thus regulates the sensitivity to the emission angle. The lower cut on the momentum fraction, on the other hand, determines the sensitivity to soft radiation. In the context of this analysis, the SD multiplicity was measured for a lower cut on the momentum fraction of $z_{\text{cut}} = 0.1$ and $\beta = 1$. In case of $\beta = 0$ (without any further cuts), the SD condition basically replicates the mMDT and does not depend on the emission angle at all. This is (as already mentioned) a very aggressive form of grooming that does not count emissions whose momentum fraction is below 10 % with respect to the jet's trunk independent of its emission angle. This condition is loosened for $\beta > 0$ provided that the constituents of the jet are constrained to be contained within the jet radius $\Delta R \leq R$, which is the case for this measurement because all jets used are reconstructed using the anti-k_t algorithm. In this case, the term $(\Delta R/R)^{\beta} < 1$ relaxes the $z_{\text{cut}} > z_{\text{cut}}(\Delta R/R)^{\beta}$; for very collinear radiation $\Delta R \to 0 (= \theta_{\text{cut}})$, the SD condition becomes $z > 0$ and therefore counts all emissions. In the limit of $\beta \to \infty$, the SD multiplicity counts all emissions inside of the jet. The different emission planes corresponding to the different configurations of the SD algorithm are visualized in a Lund diagram in Figure 11.4.

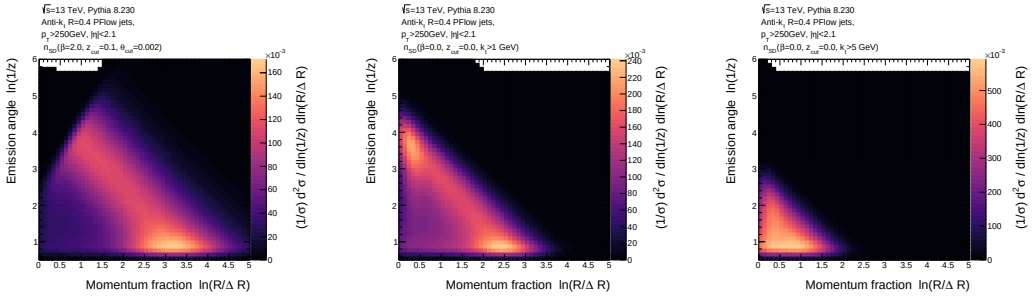
The second configuration of the SD multiplicity includes a lower cut on the $k_t \sim p_{\text{T}}\Delta R$



(a) $z_{\text{cut}} = 0, \beta = 0$ *(b)* $z_{\text{cut}} = 0.1, \beta = 0$ *(c)* $z_{\text{cut}} = 0.1, \beta = 1$

of the emissions in the C/A reclustered jet. As shown by Equation 11.7, the effect of cutting on the radiation's k_t is mathematically equivalent to setting $\beta = -1$ and to let the lower cut on the momentum fraction change dynamically as a function of p_{T} via $z_{\text{cut}}(p_{\text{T}}) = p_{\text{T}}/k_t^{\text{cut}}$ for a k_t cut of $k_t^{\text{cut}} = 1\,\text{GeV}$. This is exemplary shown in Figure 11.4b and Figure 11.4c that shows the valid emission plane for two values of the softer jet's momentum. As can be seen, for harder emissions, the phase space of the emissions in the Lund diagram increases. Increasing

[4]The selection of configurations of the SD multiplicity is mostly driven with reference to possible applications in Monte Carlo tuning as well as the reduction of the sensitivity to systematic uncertainties.



(a) $z_{\text{cut}} = 0.1, \beta = 2$ *(b)* $z_{\text{cut}} = p_{\text{T}}/1\,\text{GeV}, \beta = -1$ *(c)* $z_{\text{cut}} = p_{\text{T}}/5\,\text{GeV}, \beta = -1$

Plot 11.4: Lund jet plane for the different configurations of the SD algorithm used in the measurements of the substructure observables.

k_t^{cut}, generally decreases the sensitivity of the measurement to hadronization effects. A large variety of different values for k_t^{cut} has been studied starting from the non-perturbative energy scale $\Lambda_{\text{QCD}} \sim 200\,\text{MeV}$ up to 5 GeV. While increasing the k_t-cut results in better data-to-MC agreement, the non-continuous nature of n_{SD} results in significant "loss" in shape information, meaning, only a small number of bins does have a probability mass different from zero. With the loss of shape information, however, the sensitivity to quark-gluon-related features is reduced as well. Therefore the lower k_t-cut was chosen to be $k_t^{\text{cut}} = 1\,\text{GeV}$, which implies that – on average – the SD multiplicity for jets with $p_{\text{T}} \gtrapprox 200\,\text{GeV}$ is non-zero.

11.3 Track Multiplicity n_{trk}^{500}

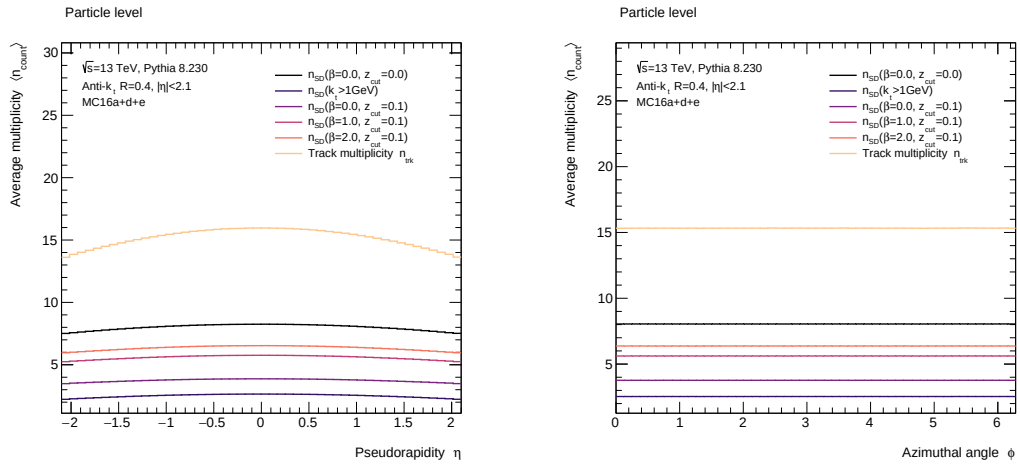
Besides the number of valid SD splittings introduced in the previous section, an additional counting observable is measured that is given by the total number of tracks inside[5] of the reconstructed jet, henceforth referred to as *track multiplicity* n_{trk}. All tracks that contribute to this substructure observable have to fulfill the basic track quality criteria at ATLAS (see Chapter 9.2.5); furthermore, to suppress contributions from pile-up and multiple scattering, the track's transverse momentum must be above 500 MeV; otherwise, the track is discarded. In the following, the notations n_{trk}^{500} and n_{trk} are used interchangeably without risk of ambiguity since no other track multiplicity with different momentum thresholds was studied within the course of this measurement.

The multiplicity of tracks associated to a jet can be considered as the number of charged constituents/particles (or at least proportional to it as there are other effect that may give rise to charge particles like, e.g., photon conversion). Referring to previous explanations, it is therefore evident that the number of tracks in a jet is expected to be sensitive to quark-gluon-related features. Indeed, it is a well-known fact that the track multiplicity is among the most powerful (non-multivariate) quark-gluon discriminants available. However, it comes with the disadvantage of not being collinear safe (without lower momentum threshold, it is neither infrared nor collinear safe) because the collinear splitting of a particle increments the multiplicity and is therefore sensitive to collinear splittings. In case of this measurement, the property of n_{trk} of not being IRC-safe is of minor significance since the usage of tracks results in (IR)C-*un*safe substructure distributions anyway.

With the track multiplicity being introduced, the set of counting-based jet substructure observables is complete. The remaining part of this section is therefore dedicated to study the relation between the different multiplicities as well as some basic distributions with the intention to provide additional information and intuition.

Firstly, the correlation between the *average* multiplicity value with the pseudorapidity η as well as the azimuthal angle ϕ of the jets is investigated. Starting with the latter one,

[5] As explained in Chapter 4.4.2, the jets are not directly reconstructed from charged particles' trajectories, but from tracks in the detector that have been matched (assigned) to the jet.



(a) Average n_{count} *versus* η *(b)* Average n_{count} *versus* ϕ

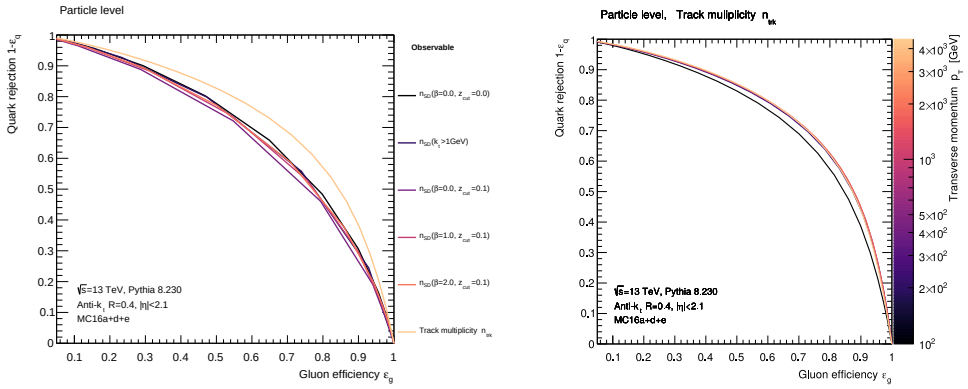
Plot 11.5: Average counting-based jet substructure observable n_{count} as a function of the jet's pseudorapidity η (*left*) and azimuthal angle ϕ (*right*).

the corresponding Figure 11.5b shows the desired features, i.e., a mostly flat distribution, which indicates that the average multiplicity $\langle n_{\mathrm{count}} \rangle$ and the azimuthal angle ϕ are apparently *un*correlated. This is in line with the fraction of quark- and gluon- initialized jets not being correlated with ϕ at all. The average multiplicity $\langle n_{\mathrm{count}} \rangle$ varies significantly between the different observables, which is expected since different configurations of the SD algorithm correspond to different levels of grooming.

Figure 11.5a shows the average multiplicity $\langle n_{\mathrm{count}} \rangle$ as a function of the jet's pseudorapidity η, revealing a (more or less) clear correlation between the two observables. This should not come as a surprise because the decision of measuring the triple-differential multijet cross section in different (pseudo)rapidity bins is based on exactly this insight. The correlation is a consequence of quark-gluon-related features in the hadron's underlying PDF that results in the central region of the detector being gluon enriched. Since all counting-based observables measured in this analysis are sensitive to quark-gluon-related features, this property translates to the average multiplicity, resulting in a correlation with η. Due to the significantly varying quark/gluon fraction as a function of the jet's p_{T}, one expects a noticeable dependence/correlation between the average number of emissions and the pseudorapidity. Consistently, the central region $\eta \approx 0$ shows a larger average for all counting-based observables due to the larger color charge of the gluon, with results in more emissions (on average) if initialized by a gluon.

It should be noted once again that all distributions in Figure 11.5 are inclusive in transverse momentum p_{T}, i.e., they are dominated by *low*-momentum jets due to the strong p_{T}-dependence of the multijet cross section.

To get an impression of the discrimination power of the different counting-based substructure observables, Figure 11.6a shows the Receiver Operating Characteristic (ROC) curve for the different observables that are used in the cross section measurement. As expected, the track multiplicity provides the best separation between quark and gluons. The quark-gluon discrimination power is very similar for all configurations of the SD algorithm, being relatively close to n_{trk} because it is not limited by Casimir scaling. The kinks visible in the ROC curve that are based on the SD multiplicity are due to discrete nature of the observable, which also applies to the track multiplicity; however, because the number of bins is larger in case of n_{trk}, the resulting ROC curve is smoother.



(a) ROC curve n_{count} *(b)* ROC curve n_{trk}

Plot 11.6: ROC curves for several counting-based observables (*left*) as well as the discrimination performance for the example of n_{trk} as a function of p_{T} (*right*).

11.4 Detector-Level Distributions

With the different counting-based jet substructure observables being introduced, this section presents the actual measurement of the triple-differential multijet cross section with respect to the track n_{trk} as well as SD multiplicity n_{SD} using pp-collision events recorded with the ATLAS detector as well as events generated in Monte Carlo with the full detector simulation applied. Moreover, all particle/truth-level distributions are reconstructed along with the response matrix, subsequently used to unfold the measurement to particle level and to derive uncertainties.

Prior to carrying out the actual, i.e., final measurement, plenty of parameters need to be optimized in order to obtain optimal performance regarding the unfolding, uncertainties etc. First and foremost, the binning of the different counting-based observables must be optimized with the objective to obtain response matrices *as diagonal as possible* while maintaining *as much information* (concerning the shape of the probability distributions) *as possible* because the shape difference between quark- and gluon-initialized jets determines the discrimination power between the two categories. This is a challenging optimization problem – especially in case of a jet substructure measurement! Generally, there is no ideal solution, but only a trade-off between diagonality of the response matrix and the information contained in the shape of the distribution[6]. Optimizing the binning for the different observables, therefore, requires to iterate over the *entire* chain of the measurement (reconstruction, unfolding to particle level and evaluation of the systematic and statistical uncertainties) several times until an adequate configuration has been found. This procedure (including all trials and tribulations) can not be fully outlined at this point, which is why some results must be anticipated here without providing further background.

In order to provide a concise presentation of all counting-based observables measured in this analysis, each substructure observable is dealt with in a dedicated section of their own.

11.4.1 Track Multiplicity n_{trk}^{500}

The first substructure observable to consider is the number of tracks associated to the jet that can be interpreted as a proxy for the number of charged constituents. The final binning for

[6] This can easily be understood by considering both extreme cases where the binning is infinitely small (i.e., continuous) and only one single bin. In the former case, there are significant migrations (between bins) as difference between particle and detector level are always present due to distortions caused by the detector. Consequently, the uncertainty on the measured distributions will be very large too. In case of a coarse binning, however, the effect of migrations can be considerably reduced at the expense of loosing information, but gaining high precision in turn.

the track multiplicity that was used in this analysis is summarized in Table 11.1. As one can see, the binning of the track multiplicity (all counting-based observables) is comparatively coarse; this is mostly driven by the tracking-related variations that induce migrations between multiplicity bins[7]. The binning shown in Table 11.1 does *not* change as a function of the

Observable	Binning	Unit
n_{trk}^{500}	$[0,5)$ $[5,10)$ $[10,15)$ $[15,20)$ $[20,25)$ $[25,30)$ $[30,35)$ $[35,40)$ $[40,45)$ $[45,50)$ $[50,80]$	none

Table 11.1: Binning of the track multiplicity n_{trk} used thought all cross section measurements in this analysis.

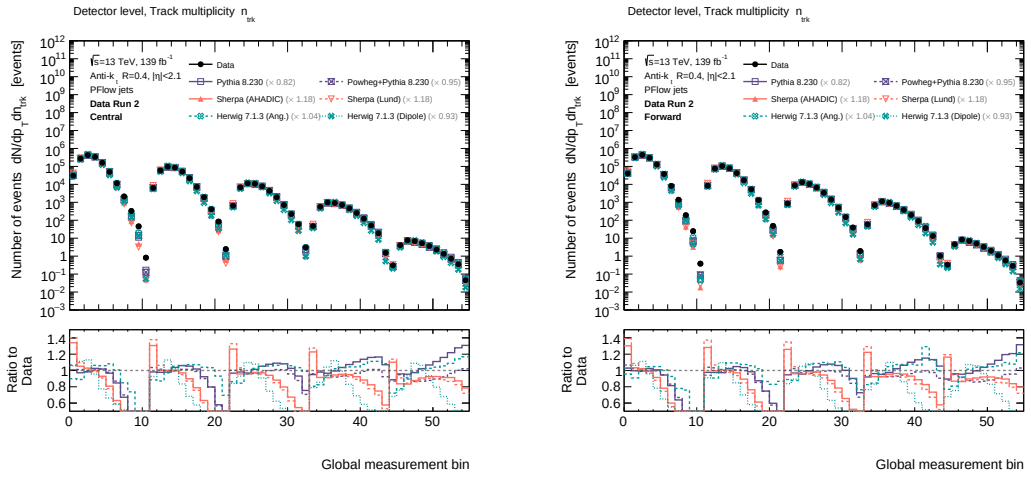
jet's transverse momentum, i.e., the very same binning is used for all p_{T} bins (as well as both pseudorapidity bins) listed in Table 9.4. There are, however, several argument that speak in favor of varying the binning of the JSS observables for different p_{T} regimes instead, most importantly, to account for the significant change in the shape of the distributions in different regions in phase space (increasing average multiplicity and broadening of the distributions with increasing transverse momentum of the jet for counting-based observables). For instance, some bins with high probability mass in the low-p_{T} regime might be spaciously populated for high-p_{T} jets (because the mean of the distribution increases with p_{T}), which result in larger systematic uncertainties in this bin due to migrations and increased statistical uncertainties. Using a variable binning for the JSS observables, however, represents a *considerable* additional effort concerning the implementation of the analysis since the aforementioned optimization problem must be performed in each p_{T} bin individually[8]. To avoid this problem, the evolution of the differential cross section $(1/\sigma)\,\mathrm{d}\sigma/\mathrm{d}\mathcal{O}_{\mathrm{JSS}}$ with transverse momentum p_{T} is accounted for in another way, i.e., by subsequently merging adjacent bins. This important point will be covered in great detail in an upcoming chapter that uses the final measurements of the triple-differential multijet cross section to extract the underlying quark/gluon distributions using the "jet topics" method introduced in Chapter 7.

The resulting spectra on detector level for central and forward jets are shown in Figure 11.7. A separate presentation of the distributions for central and forward jets was chosen for reasons of clarity; one should keep in mind, however, that the cross section is measured in *both pseudorapidity bins simultaneously*, i.e., there are migrations between the two pseudorapidity categories too. This will become immediately apparent as soon as the response matrix is shown in the upcoming section. Figure 11.7 shows the number of events in *each* phase space bin $[p_{\mathrm{T}}, \eta, n_{\mathrm{trk}}]$ measured for the full Run 2 data set as well as for all Monte Carlo samples for the corresponding simulation campaigns *at detector level*. The distributions of all Monte Carlo samples are normalized/scaled by a factor to match the total number of events measured in data. As mentioned before, this is only done for comparing the shape of the distributions at detector level. The corresponding factor by which the Monte Carlo spectrum was scaled is shown in the figure's legend in parenthesis attached to the label of the respective MC generator (see Figure 8.1), similar two what was done in the previous section.

In the representation of the measured spectra shown in Figure 11.7, the non-equidistant binning of the observable expected from Table 11.1 is *not* explicitly recognizable. In this representation, each phase space bin, in which the number of multijet events is measured, is assigned to a *unique* integer value (henceforth referred to as *global bin index* or *global*

[7]Keep in mind that especially the tracking-related variations associated with inefficiency in the reconstruction of tracks in the Inner Detector randomly remove tracks according to some probability that depends on the transverse momentum as well as the pseudorapidity. This may give rise to considerable migration that must be accounted for by means of an increased bin width.

[8]Honesty demands that the need for a variable binning of the JSS observable was not anticipated from the very beginning when designing the analysis, but was an insight gained after a significant fraction of the analysis code (which, in general, is very modular) has already been written. Implementing this new feature would have given rise to considerable additional expenses regarding evaluation and debugging and has therefore discarded for good, but should certainly be considered in subsequent studies following this one.



(a) Central jets *(b)* Forward jets

Plot 11.7: Full measurement of the number of multijet events in different regions in phase space with respect to the track multiplicity for central (*left*) and forward (*right*) jets. The divided presentation of the multijet spectra for central and forward jets only serves the purpose of better visualization; indeed, the number of multijet events is *simultaneously* measured in both pseudorapidity regions and therefore accounts for migrations between the two.

measurement bin) by a bijective map according to

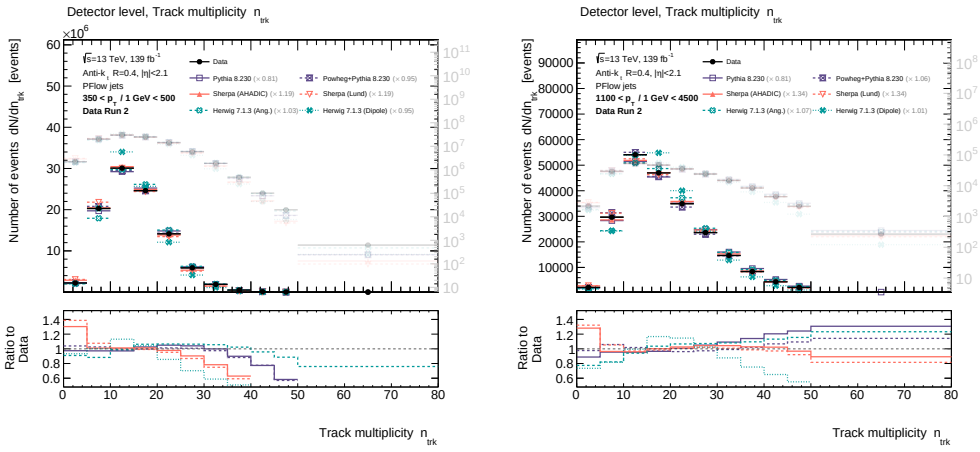
$$I_{i_{\mathcal{O}_{\mathrm{JSS}}},i_{p_{\mathrm{T}}},i_\eta} := I(i_{p_{\mathrm{T}}}, i_\eta, i_{\mathcal{O}_{\mathrm{JSS}}}) = i_{\mathcal{O}_{\mathrm{JSS}}} + N^{\mathrm{bins}}_{\mathcal{O}_{\mathrm{JSS}}} \cdot (i_{p_{\mathrm{T}}} + i_\eta \cdot N^{\mathrm{bins}}_{p_{\mathrm{T}}}), \tag{11.18}$$

whereby $i_{p_{\mathrm{T}}}, i_\eta, i_{\mathcal{O}_{\mathrm{JSS}}} \in [0, N^{\mathrm{bins}}_{p_{\mathrm{T}},\eta,\mathcal{O}_{\mathrm{JSS}}}]$ denotes the current bin number in terms of pseudorapidity, transverse momentum p_{T} and the substructure observable $\mathcal{O}_{\mathrm{JSS}}$, with $N^{\mathrm{bins}}_{\mathcal{O}_{\mathrm{JSS}}}$ and $N^{\mathrm{bins}}_{p_{\mathrm{T}}}$ being the total number of bins of the latter two. The pseudorapidity only comes with two bins (central(=0) and forward(=1)). The function according to Equation 11.18 allows to transform the 3-dimensional cross section (which would require a 4-dimensional space to be visualized) onto a 1-dimensional space, significantly simplifying visualization as well as further processing of the *entire* measurement. With this transformation applied, each repeating pattern in Figure 11.7a and Figure 11.7b corresponds to a different bin in transverse momentum for central (*left*) and forward jets (*right*), respectively, for which the track multiplicity was measured[9].

Of course, the distributions in Figure 11.7 can also be transformed into a representation in which the non-equidistant binning of the substructure observable is explicitly visible like in Figure 11.8.

Figure 11.8 shows the measured spectra of the track multiplicity for two different p_{T} bins for the full Run 2 data and all Monte Carlo samples listed in Table 8.1 in a linear (left axis) as

[9] A concrete example might be beneficial to better understand the transformation according to Equation 11.18. Assumed a truth-level jet with a transverse momentum of $p_{\mathrm{T}} = 360\,\mathrm{GeV}$ (bin $i_{p_{\mathrm{T}}} = 1$ as this bin number corresponds to $p_{\mathrm{T}} \in [350\,\mathrm{GeV}, 500\,\mathrm{GeV}]$ according to Table 9.4) was reconstructed in the central region (bin number $i_\eta = 0$ as there are two pseudorapidity bins) with a total number of $n_{\mathrm{trk}} = 21$ tracks (bin $i_{\mathcal{O}_{\mathrm{trk}}} = 4$ because the 4th bin corresponds to $n_{\mathrm{trk}} \in [20, 25]$ referring to Table 11.1). According to Equation 11.18, this jet will contribute to bin $I^{\mathrm{PL}}_{4,1,0} = 15 (= 4 + 11 \cdot (1 + 0 \cdot 5))$ in the *global* representation. Now, consider the very same jet but at detector level. Due to distortion in the reconstruction of the jet's energy in the calorimeter system of the detector, the energy of the truth-level jet is measured to be only 340 GeV at detector level. Because the position measurement is very precise, the jet is still correctly identified as being central. In this (not unreasonable) example, the jet will be assigned to a different global bin index at detector level, i.e., $I^{\mathrm{DL}}_{4,0,0} = 4 (= 4 + 11 \cdot (0 + 0 \cdot 5))$ because the jet now falls in $i_{p_{\mathrm{T}}} = 0$, which corresponds to $p_{\mathrm{T}} \in [250\,\mathrm{GeV}, 350\,\mathrm{GeV}]$. In the response matrix, the correlation between I^{DL} and I^{PL} is shown; for the example at hand, the jet will contribute to the *off*-diagonal elements since the jet migrated between two p_{T} bins.



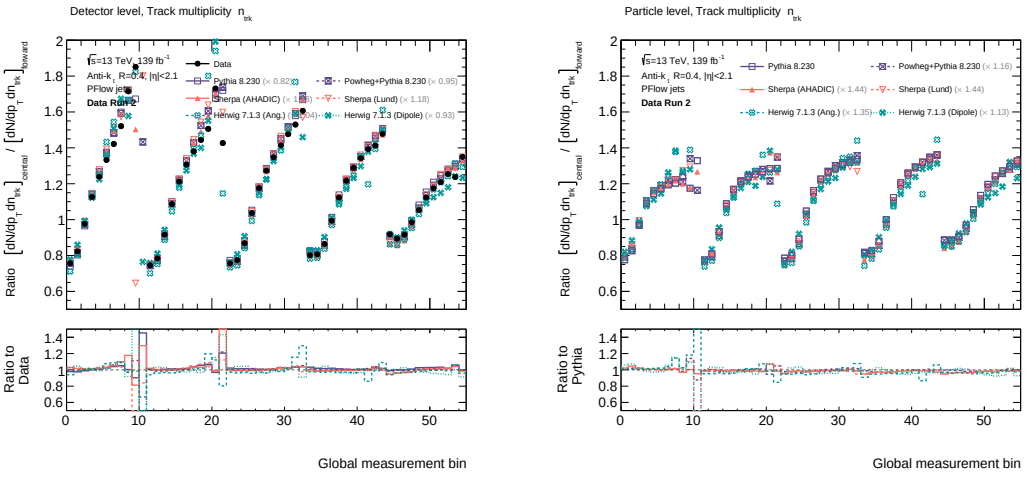
(a) $p_{\mathrm{T}} \in [250, 350)$ *(b)* $p_{\mathrm{T}} \in [1100, 4500)$

Plot 11.8: Single-differential distributions of the total number of multijet events measured in each track multiplicity bin for different p_{T} (in GeV) regimes.

well as a logarithmic (right, faint axis) scale to illustrate the number of events in the highest multiplicity bins in the tails of the distributions with low production rates. The lower plot once again shows the ratio between all simulation-based spectra with respect to the distributions measured in data. As one can see, the shape of the distributions significantly changes between different p_{T} regimes, with the average number of tracks moving towards higher multiplicities for increasing transverse momenta.

The production rate of multijet events is exponentially falling with increasing transverse momentum, which is expected from theoretical calculations of the multijet cross section. Looking at the individual track multiplicity distributions for each p_{T} bin, two main features can be observed that are shared by *all* counting based observables considered in this analysis: *firstly*, the average/mean of the (track) multiplicity distributions increases towards higher transverse momenta; *secondly*, the distributions broaden in company with the rising mean value of the number of emissions.

Concerning the data to simulation resemblance, a considerable variation regarding the



(a) Particle level *(b)* Detector level

Plot 11.9: Ratio between the spectra for central and forward jets at particle level (*right*) and detector level (*right*). In case of the particle level distributions, the ratio was taken with respect to the nominal sample (Pythia) as no data exists at truth level.

degree of agreement for the different Monte Carlo generators can be observed. Overall, the MC-based predictions of the relative number of events is in good agreement with the spectrum observed in data, with deviations below 10 % close to the bulk of the distributions up to 50 % (and more) in regions dominated by hadronization effects, soft QCD or limited statistics. For the low-p_{T} regime, the high-multiplicity bins are almost completely left unpopulated because the average number of tracks is much lower (see the statements above concerning the problem of non-p_{T}-dependent binning of the JSS observables). In the high-p_{T} regime, however, the tail of the underlying n_{trk} distribution evolves towards larger multiplicities, which are most likely to contain a significant fraction of gluon-initialized jets as the average number of particles is larger compared to quarks. Since all state-of-the-art Monte Carlo generators (including Pythia of course) have been extensively tuned to quark-enriched samples from LEP data, the overall data-to-MC-agreement is expected to be better for regions in phase space (such as those given by smaller track multiplicities) that are dominated by quark-initialized jets. The actual situation is, needless to say, more involved because the different Monte Carlo generators have been tuned to data to varying degrees. Nonetheless, the given statement captures one aspect the explanation.

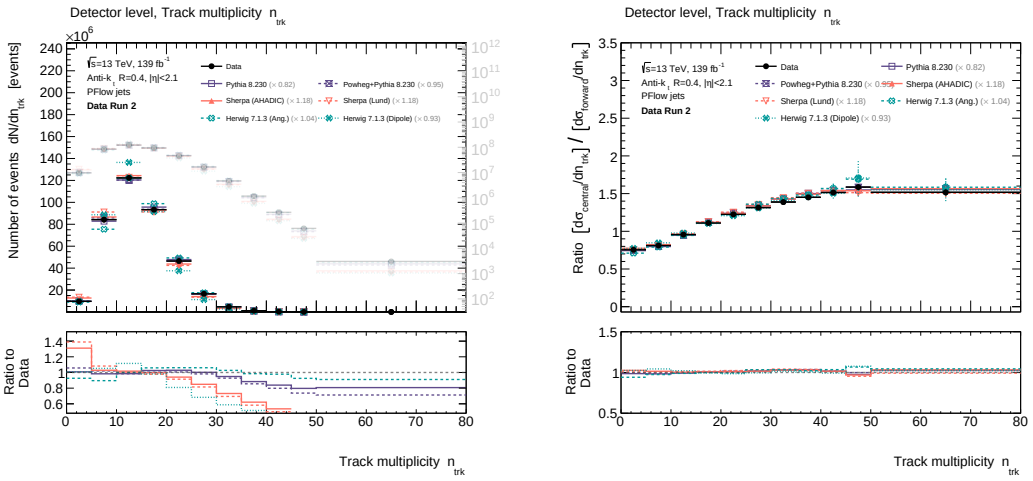
The multi-differential cross section is measured in two regions that are defined by the relative pseudorapidity between the two leading-p_{T} jets in the event. The motivation behind this analysis strategy is to obtain a measurement of the relative multijet production rate over a sample that is quark- and gluon-enriched, respectively. Comparing the corresponding spectra for central jets in Figure 11.7a and forward jets in Figure 11.7b *by eye*, it proves to be pretty difficult to identify striking differences between the two spectra. It is therefore beneficial to consider the ratio between the distributions corresponding to the central and forward region. The result is shown in Figure 11.9 for particle (11.9a) and detector (11.9b) level, respectively. The ratio(s) in Figure 11.9 reveal an interesting structure between central and forward jets – which will play a crucial role for the extraction of the jet topics according to the method described in Chapter 7. For each p_{T} bin, the ratio is (approximately) *monotonically* increasing with the (track) multiplicity. This is a clear indication of the two samples (defined by the central and forward region) giving rise to different mixtures/compositions of jets, i.e., different fractions of quarks and gluons. Because the central region is gluon-enriched due to the structure of the underlying PDF, the bulk of the distribution is at higher multiplicity values compared to the forward regions that is "dominated" by quark-initialized jets. The limits of those ratios ($r^{\mathrm{central}}_{\mathrm{forward}} \to \min / \max$) – simply speaking – determine the "anchor" bins/regions that are used to extract the underlying quark/gluon distributions and fractions as explained in Chapter 7[10]. Regarding the features of the ratio with respect to different transverse momenta, a pronounced difference between particle and detector level can be observed. In case of the truth-level distributions, the p_{T}-dependence is much weaker compared to the corresponding figure at detector level. This does not come as a surprise because multiplicities are expected to be sensitive to detector effects due to, e.g., varying material budget or different responses of the calorimeter for quark- or gluon-initialized showers. Those effects are corrected for by the unfolding procedure, described in the upcoming section.

Figure 11.10 shows the spectrum of the track multiplicity as well as the ratio between the measured distributions in central and forward region *inclusive* in transverse momentum and pseudorapidity. The corresponding distributions can simply be calculated by marginalizing the transverse momentum as well as the pseudorapidity categories, which corresponds to the summation of the individual multiplicity spectra in each p_{T} and pseudorapidity bin according to

$$\frac{\mathrm{d}N}{\mathrm{d}n_{\mathrm{trk}}} = \sum_{j=0}^{N^{\mathrm{bins}}_{p_{\mathrm{T}}}} \sum_{k=0}^{N^{\mathrm{bins}}_{\eta_{\mathrm{rel}}}} \left[\frac{N}{\Delta p_{\mathrm{T}} \Delta \eta_{\mathrm{rel}} \Delta n_{\mathrm{trk}}} \right]_{jk} . \tag{11.19}$$

Because the cross section of multijet events is exponentially falling with transverse momentum of the jets, the resulting track multiplicity distribution in Figure 11.10a is mostly determined by the "low"-p_{T} regime as can be verified in direct comparison with Figure 11.8a. Nonetheless,

[10] The anchor bins are determined based on the ratio of the *unfolded* spectra in central and forward region.

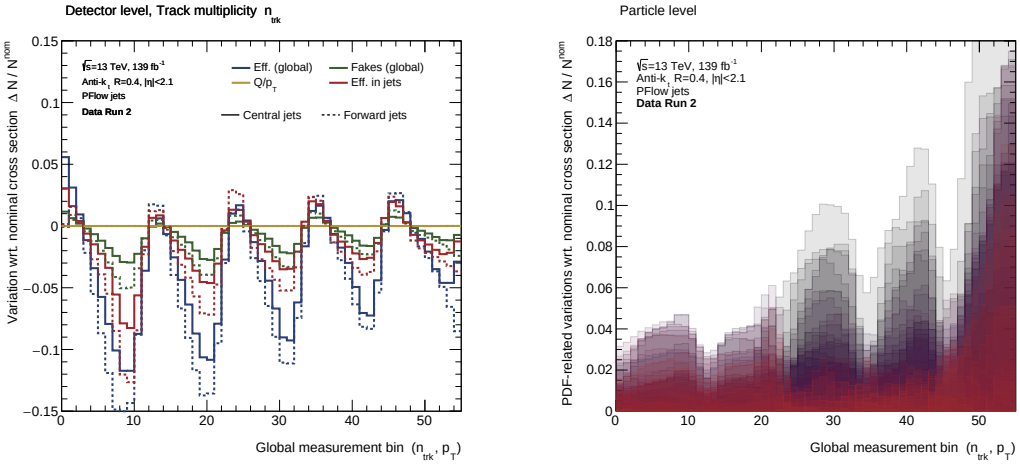


(a) *(b)*

Plot 11.10: Reconstructed track multiplicity distributions n_{trk} in data and simulation inclusive in transverse momentum and pseudorapidity (*left*) and the ratio between the spectra measured in central and forward region (*right*).

a significant contribution to the tail of the distribution (see faint curve) due to high-p_{T} jets can be observed in the logarithmic representation, which is mainly populated by jets in the high-p_{T} domain. The rightmost Figure 11.10b shows the ratio of the multijet spectra measured in central and forward region at detector level inclusive in transverse momentum – in contrast to Figure 11.9 that shows the cross section ratio for individual p_{T} bins. In this representation, the non-equidistant binning is, once again, explicitly visible. With reference to explanations already provided in context of Figure 11.9, the curve progression in Figure 11.10b can be well understood by recalling that the central and forward region have different compositions of quark and gluon jets due to the underlying PDF of the proton. Because quark jets are (on average) harder and composed of a smaller number of particles, the left-most limit is dominated by quark-like jets, with the rightmost part of the distribution being gluon-enriched accordingly. This also provides a possible explanation for the progression of the data-to-MC agreement of the ratio with respect to n_{trk}.

As explained in Chapter 10, the systematic uncertainties of the cross section measurement are determined by fully propagating corresponding variations of the respective nominal spectra through the unfolding procedure. Therefore, each systematic uncertainty comes with a set of variations of the triple-differential multijet cross section at detector and/or particle level. Figure 11.11 shows two examples of those variations that correspond to variations due to tracking- (11.11a) and PDF-related (11.11b) uncertainties. The total tracking-related uncertainty consists of several components that manifest as different variations of the nominal cross section that are shown in Figure 11.11a for central (solid lines) and forward (dashed lines) jets. First and foremost, the variation of the cross section is systematically larger in case of jets more forward in the detector. This is expected as forward jets see (on average) an increased material budget. The different components of the tracking-related variations have been described in Chapter 7; broadly speaking, they are divided into four categories: one of which accounts for global/inclusive track reconstruction efficiency; for fake jets mostly due to combinatorics and pile-up; efficiency in dense environments inside jets; as well as an uncertainty on the Q/p_{T} term. As expected, the latter one is completely negligible – at least in this analysis. In case of the other components, the degree of variation varies significantly between different multiplicity bins up to 10 % with respect to the nominal cross section. It is important to keep in mind that the variations shown in Figure 11.11a do *not* correspond to the final tracking-related uncertainty as they are on detector level. In order to obtain the corresponding uncertainty, the tracking-related variations must be propagated throughout the unfolding procedure described



(a) Tracking-related variations

(b) PDF-related variations

Plot 11.11: Variations of the nominal spectra at detector level due to tracking- (*left*) and PDF-related (*right*) uncertainties.

in the upcoming section.

Figure 11.11b shows the variations of the nominal cross section due to variations of the baseline PDF set.

11.4.2 SD Multiplicity

The next observable that has been measured is the number of (valid) Soft Drop emissions inside of the jet. As explained previously, two configurations of the ISD algorithm have been studied: *firstly*, one that uses the default configuration of the algorithm ($\beta = z_{\text{cut}} = 0$), but applies a lower cut on the transverse momentum of the emission of $k^{\text{cut}} = 400\,\text{MeV}$, which means that only emissions/splittings with $k_t > 400\,\text{MeV}$ are taken into account (this basically corresponds to only counting perturbative splittings); *secondly*, no k_t-cut being applied (explicitly) but with $\beta = 1.0$, $z_{\text{cut}} = 0.1$ and a lower cut on the emission angle of $\theta_{\text{cut}} = 0.002$.

Before turning the attention to the reconstructed distributions, Table 11.2 shows the binning of both configurations of the SD multiplicity. As in case of the track multiplicity, the

Observable	Binning	Unit
$n_{\text{SD}}(k_t > 400\,\text{MeV})$	$[0,2)$ $[2,3)$ $[3,4)$ $[4,5)$ $[5,6)$ $[6,7)$ $[7,8)$ $[8,9)$ $[9,10)$ $[10,12)$ $[12,15)$ $[15,20)$	none
$n_{\text{SD}}(z_{\text{cut}} = 0.1, \beta = 1, \theta_{\text{cut}} = 0.002)$	$[0,3)$ $[3,4)$ $[4,5)$ $[5,7)$ $[6,7)$ $[7,8)$ $[8,9)$ $[9,12)$ $[12,20)$	none

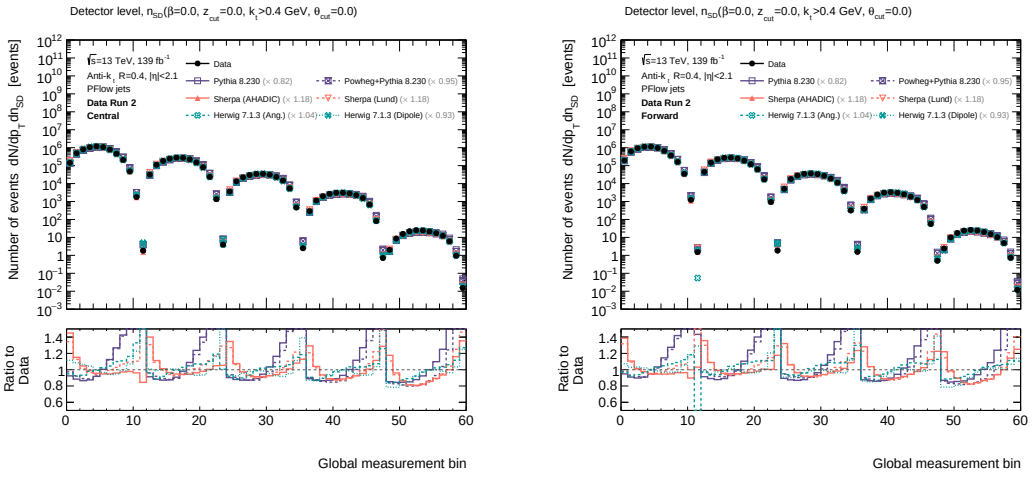
Table 11.2: Binning of the two different configurations of the SD multiplicity n_{SD} used thought all cross section measurements in this analysis.

binning was optimized to minimize migrations between bins while keeping as much information regarding the shape of the distributions as possible.

Following the structure of Section 11.4.1 for the track multiplicity, Figure 11.12 shows the *full* measurement of the SD multiplicity $n_{\text{SD}}(z_{\text{cut}} = 0, \beta = 0, k_t > 400\,\text{MeV})$ (henceforth notated as $n_{\text{SD}}(k_t > 400\,\text{MeV})$) for central (11.12a) and forward (11.12b) jets. In this context, the attention is – again – explicitly drawn to the fact that the cross section was measured and unfolded *simultaneously* in central and forward region. As a quick reminder, the effect of a lower cut on the emissions' momentum is shown in Figure 11.4b and Figure 11.4c, respectively,

for the Lund jet plane.

In the previous configuration, the SD emission inside of the C/A reclustered jet was only



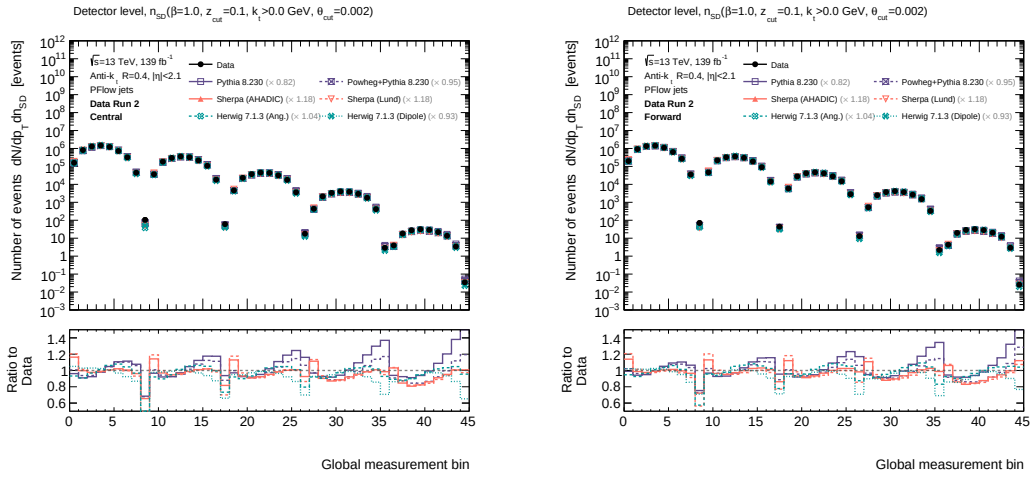
(a) Central jets *(b)* Forward jets

Plot 11.12: Full measurement of the number of multijet events in different regions in phase space for central (*left*) and forward (*right*) jets with respect to the transverse momentum p_{T} and the SD multiplicity $n_{\mathrm{SD}}(k_t > 400\,\mathrm{MeV})$.

considered valid (and therefore counted) if the k_t of the emission was above 400 MeV. The second configuration of the SD multiplicity uses a lower cut on the momentum fraction of the emissions of $z_{\mathrm{cut}} = 10\,\%$ with respect to the p_{T} of the jet's trunk that is loosened by taking the angular components of the emissions inside of the jet into account by setting $\beta = 1.0$ in Equation 11.2. Furthermore, very collinear radiation is removed by using a lower cut on the emission angle of $\theta_{\mathrm{cut}} = 0.002$. The corresponding emission plane is shown in Figure 11.3c. The motivation for measuring this configuration in addition to the one that was shown in the previous paragraph is given by the circumstance of $n_{\mathrm{SD}}(\beta = 1.0, z_{\mathrm{cut}} = 0.1)$ being one of ATLAS' default settings of the SD algorithm, which will play an important role in a potential measurement of α_{S} based on the results derived in this analysis, which will be further developed in the proposal of "future research suggestions" in the end of this report. Furthermore, a second configuration of the SD multiplicity was measured to investigate the effect of different binnings (see Table 11.2) on the extracted quark/gluon distributions and fractions. The corresponding spectra for central and forward jets for $n_{\mathrm{SD}}(\beta = 1.0, z_{\mathrm{cut}} = 0.1)$ are shown in Figure 11.13.

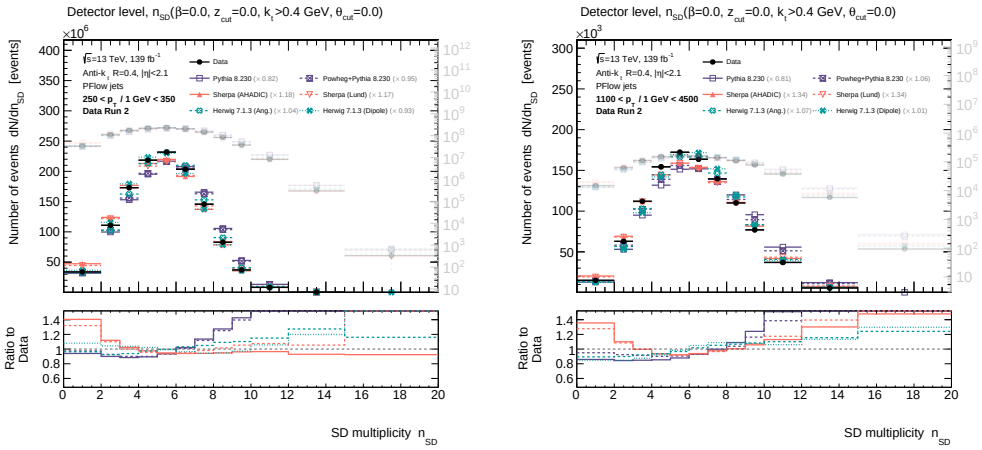
In Figure 11.12 and Figure 11.13, like for all corresponding figures, the total number of events measured in each phase space bin was normalized by the total bin width $\Delta w^{ijk}_{\mathrm{bin}} = \Delta^i \mathcal{O}_{\mathrm{JSS}} \Delta^j p_{\mathrm{T}} \Delta^k \eta$, whereby $\Delta^n \Pi = \Pi^{n+1} - \Pi^n$ with $\Pi \in [\mathcal{O}_{\mathrm{JSS}}, p_{\mathrm{T}}, \eta]$ denotes the nth bin's width. This way, the "correct" or expected shape of the distributions is recovered, taking the different bin widths into account.

As for the track multiplicity, two examples of the single-differential multijet cross section (inclusive in central and forward region) are shown in Figure 11.15 and Figure 11.15 for the two configurations of the SD algorithm for two bins (low and high p_{T}) in transverse momentum. The aforementioned correlation between the average number of emissions and the width of the SD multiplicity distribution with the transverse momentum of the jet can nicely be observed going from Figure 11.14a to Figure 11.14b, which corresponds to an increase in p_{T}. Regarding the data-to-MC agreement, most simulated samples agree well for most regions of phase space with a discrepancy below 10 %. Besides that, Figure 11.14 and Figure 11.15 illustrate the previously stated argument of using different binning for the SD multiplicity (instead of a fixed one) to account for the evolution of the underlying probability distribution with transverse momentum of the jets. Because the same binning was used for all p_{T} bins in this measurement, the binning of the SD multiplicity represents a compromise between the diagonality of the



(a) Central jets *(b)* Forward jets

Plot 11.13: Full measurement of the number of multijet events in different regions in phase space for central (*left*) and forward (*right*) jets with respect to the transverse momentum p_{T} and the SD multiplicity $n_{\mathrm{SD}}(z_{\mathrm{cut}} = 0.1, \beta = 1, \theta_{\mathrm{cut}} = 0.002)$.

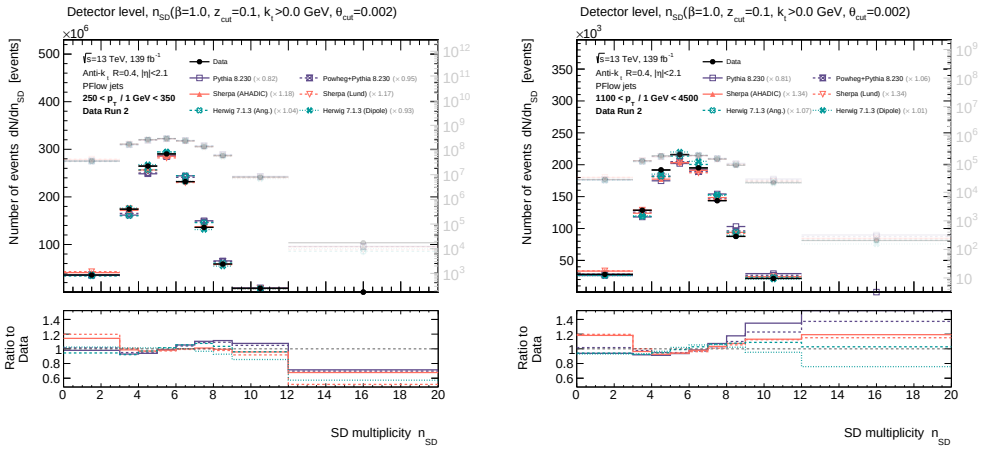


(a) $p_{\mathrm{T}} \in [250, 350)$ *(b)* $p_{\mathrm{T}} \in [1100, 4500)$

Plot 11.14: Single-differential distributions of the total number of multijet events measured in each SD multiplicity bin of $n_{\mathrm{SD}}(k_t > 400\,\mathrm{MeV})$ for different p_{T} (in GeV) bins.

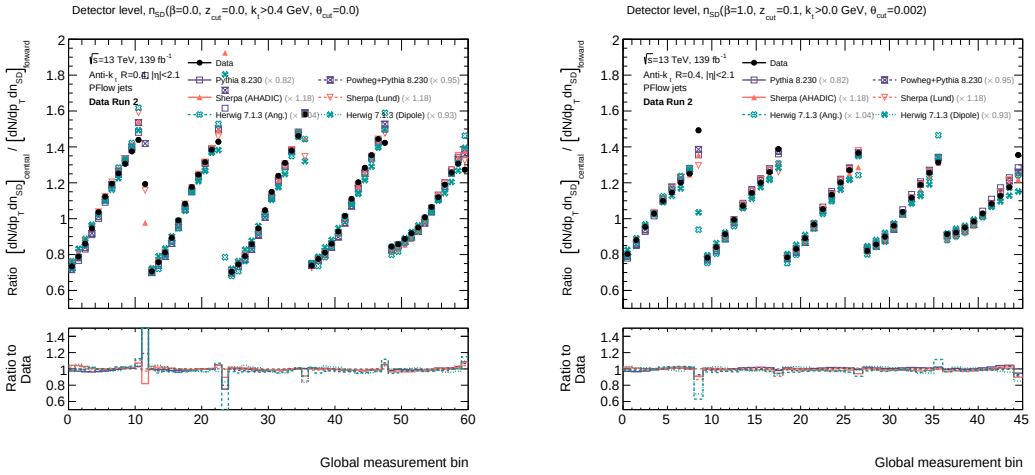
response matrix over the entire p_{T} range and the information encoded in the shape of the distribution. The sub-optimal binning of the multiplicities (this statement also holds true for the measurement of the NN-score presented in the upcoming chapter) poses a potential problem for the extraction of the underlying quark/gluon distributions in the context of the determination of the anchor regions/bins. This complication is accounted for by subsequently merging adjacent bins – this point will be discussed in great detail in the corresponding sections of this report.

Like in case of n_{trk}, the superficial comparison of the measured spectra in central and forward region does not reveal a striking difference between the two (jet) samples. The ratio between the two distributions, shows that the SD multiplicity is indeed sensitive to the differences regarding the fraction of quark and gluon jets of the two mixtures over which the cross section is measured. This is shown in Figure 11.16 for the measured spectra of $n_{\mathrm{SD}}(k_t > 400\,\mathrm{MeV})$ (11.16a) and $n_{\mathrm{SD}}(z_{\mathrm{cut}} = 0.1, \beta = 1, \theta_{\mathrm{cut}} = 0.002)$ (11.16b). The agreement



(a) $p_{\mathrm{T}} \in [250, 350)$ *(b)* $p_{\mathrm{T}} \in [1100, 4500)$

Plot 11.15: Single-differential distributions of the total number of multijet events measured in each SD multiplicity bin of $n_{\mathrm{SD}}(z_{\mathrm{cut}} = 0.1, \beta = 1, \theta_{\mathrm{cut}} = 0.002)$ for different p_{T} (in GeV) bins.



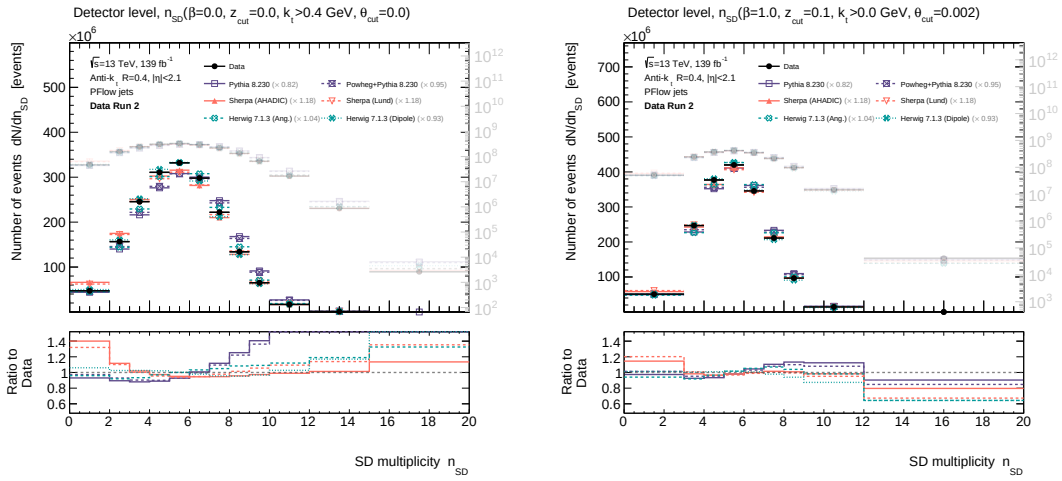
(a) $n_{\mathrm{SD}}(k_t > 400\,\mathrm{MeV})$ *(b)* $n_{\mathrm{SD}}(z_{\mathrm{cut}} = 0.1, \beta = 1, \theta_{\mathrm{cut}} = 0.002)$

Plot 11.16: Ratio between the spectra for central and forward jets of $n_{\mathrm{SD}}(k_t > 400\,\mathrm{MeV})$ (*left*) and $n_{\mathrm{SD}}(z_{\mathrm{cut}} = 0.1, \beta = 1, \theta_{\mathrm{cut}} = 0.002)$ (*right*) at detector level.

between data and simulation is excellent as most effects in central and forward region are highly correlated among each other and therefore cancel to the degree of their correlation; they only vary concerning their composition of quark and gluon jets. The spikes that can be observed in the ratio plots in Figure 11.16 are mostly due to very low statistics in the respective upper and lower limits of phase space. The curve progression of the ratio as a function of the (global) measurement bin closely resembles those of the track multiplicity in Figure 11.9, but with a better agreement between particle and detector level. The better agreement between truth and reco level is mostly due to the coarser binning of the distribution's support, but also because the SD multiplicity is generally less sensitive to soft-physics effects (at least for the configurations used in this analysis).

Figure 11.17 shows a summary of the measurement of the two configurations n_{SD} at detector level in terms of the inclusive (over all p_{T} bins and pseudorapidity categories) single-differential spectrum.

As it was the case for the track multiplicity, the density of the multijet cross section quickly

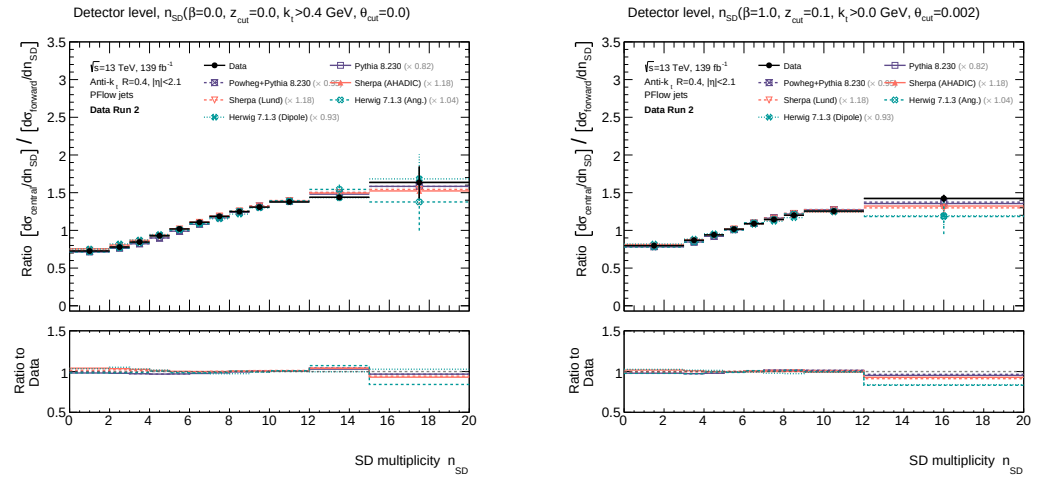


(a) $n_{\mathrm{SD}}(k_t > 400\,\mathrm{MeV})$ *(b)* $n_{\mathrm{SD}}(z_{\mathrm{cut}} = 0.1, \beta = 1, \theta_{\mathrm{cut}} = 0.002)$

Plot 11.17: Reconstructed SD multiplicity distributions n_{SD} in data inclusive in transverse momentum and relative pseudorapidity for $n_{\mathrm{SD}}(k_t > 400\,\mathrm{MeV})$ (*left*) and $n_{\mathrm{SD}}(z_{\mathrm{cut}} = 0.1, \beta = 1, \theta_{\mathrm{cut}} = 0.002)$ (*right*).

falls with the transverse momentum of the jets, which is why the bulk of the distribution in Figure 11.17 is mostly due to low-p_{T} jets. However, the logarithmic representation of the spectrum clearly shows contributions from higher-p_{T} jets that are more likely to populate tails of the distribution corresponding larger multiplicities (pay attention to the fainted curve and corresponding scale). The average number of SD emissions $\langle n_{\mathrm{SD}} \rangle$, of course, strongly depends on the transverse momentum, which will be studied once the measurement got unfolded to particle level. The inclusive ratio in Figure 11.18 nicely shows the different fraction of quark and gluon jets in central and forward regions for the various SD multiplicity bins. The highest n_{SD} bin exhibits significant deviations between the different samples due to statistical limitations.

Last but not least, like it was done in case of the track multiplicity, the variations of the cross

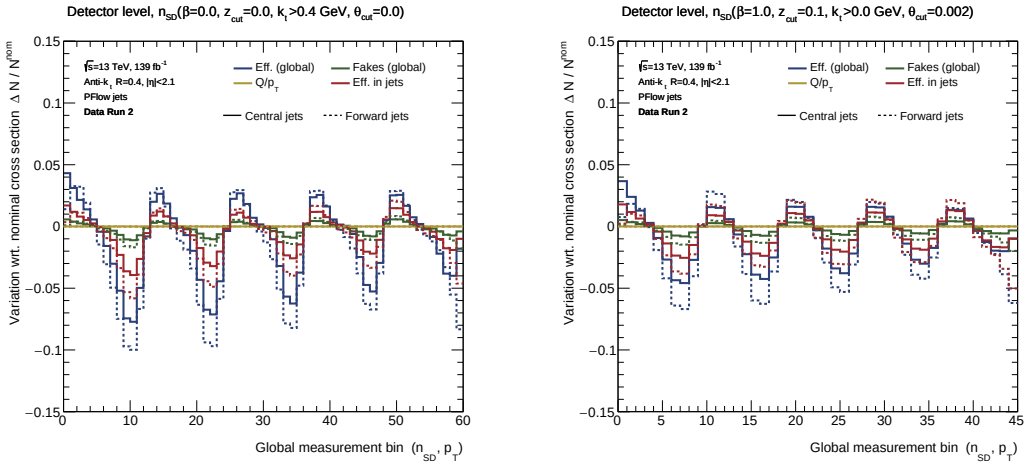


(a) $n_{\mathrm{SD}}(k_t > 400\,\mathrm{MeV})$ *(b)* $n_{\mathrm{SD}}(z_{\mathrm{cut}} = 0.1, \beta = 1, \theta_{\mathrm{cut}} = 0.002)$

Plot 11.18: Ratio between reconstructed spectra in central and forward region in data inclusive in transverse momentum and relative pseudorapidity for $n_{\mathrm{SD}}(k_t > 400\,\mathrm{MeV})$ (*left*) and $n_{\mathrm{SD}}(z_{\mathrm{cut}} = 0.1, \beta = 1, \theta_{\mathrm{cut}} = 0.002)$ (*right*) at detector level.

section related to the tracking uncertainties are shown in Figure 11.19 for $n_{\mathrm{SD}}(k_t > 400\,\mathrm{MeV})$

(*left*) and $n_{\mathrm{SD}}(z_{\mathrm{cut}} = 0.1, \beta = 1, \theta_{\mathrm{cut}} = 0.002)$ (*right*). As expected, the tracking-related



(a) Tracking-related variations

(b) PDF-related variations

Plot 11.19: Variations of the nominal spectra at detector level due to tracking-related uncertainties for $n_{\mathrm{SD}}(k_t > 400\,\mathrm{MeV})$ (*left*) and $n_{\mathrm{SD}}(z_{\mathrm{cut}} = 0.1, \beta = 1, \theta_{\mathrm{cut}} = 0.002)$ (*right*).

variations are overall smaller compared to the corresponding graphs of the track multiplicity in Figure 11.11a. The variations associated with the PDF are basically the same and therefore not shown. Generally, most statements provided in case of the track multiplicity in Section 11.3 still apply to the SD multiplicity and are therefore left without any further comment. The only aspect to note at this point is how the coaser binning of the multiplicity reduces the sensitivity to tracking-related uncertainties.

11.5 Unfolding of the Cross Section Measurement

The previous section presented the measured triple-differential multijet spectra in *real* pp-collision data and Monte Carlo simulation for the track multiplicity as well as several configurations of the SD multiplicity introduced in Section 11.1.3. Furthermore, the ratio between the spectra *simultaneously* measured in the central and forward pseudorapidity region was examined, which confirmed that all counting-based observables studied in this analysis are sensitive to the different composition of quarks and gluons in the two orthogonal samples of jets given by the central and forward category. This fact will be come important once the underlying quark and gluon distributions as well as their corresponding fractions are extracted.

The measured distributions in Section 11.4 are – for the most part – at *detector level*, i.e., the obtained spectra are "distorted" due to limited acceptance and finite resolution of the ATLAS detector. As described in Chapter 5, the procedure of detector unfolding aims to "deconvolut" the effect of the measuring apparatus by transforming the measurement to particle/hadron level by means of the response matrix that captures the correlations between particle- and detector-level objects. Moreover, the data must be corrected for acceptance and efficiency losses due to migrations of jets that enter or leave the phase space volume over which the triple-differential multijet cross section is measured. After unfolding, the cross section is at particle level and can directly be compared to predictions of a Monte Carlo generator or explicit theory calculations, whereby the latter case requires an IRC-safe observable/measurement.

The purpose of this section is to provide all relevant information regarding the unfolding of the triple-differential multijet cross section with respect to the various counting-based observables introduced in the previous section (technical aspects in this section also apply to the NN-based observables). In this context, the main emphasis is on the diagonality of the response/unfolding matrix as well as the efficiency and fake factors in the different regions of

phase space. Generally, this section intends to provide convincing arguments to the reader that the unfolding of all substructure observables is well under control.

Regarding the structure of this section, the arrangement of results basically resembles the order of Section 11.4, i.e., each counting-based observable that has been measured gets its own brief section showing the response matrix and the efficiency/fake factors along with some general comments and comparisons between the different variables.

11.5.1 Background Processes[11]

As briefly mentioned (but not explicitly demonstrated) in the final paragraph of Chapter 9.1, contributions from other sources besides multijet production to the measured cross section are completely negligible for all regions in phase space considered in this analysis. The reason for this measurement being (almost) entirely insensitive to background processes is that *all* real jets (i.e., phenomenological signatures of quarks and gluons in the experiment) are considered a signal *by definition*. Hence, the only possible source of backgrounds to this measurement is due to other particles being misidentified and falsely reconstructed as hadronic jets (with the exception of hadronic jets from tau decays). This may happen as there is a "continuous" transition between electromagnetic and hadronic showers. Especially electromagnetic showers initialized by a very high-p_{T} electron or photon may penetrate into the hadronic calorimeter, which may result in fake (hadronic) jets. Furthermore, taus are likely to decay into hadrons ($\Gamma_{\tau\to\mathrm{hadrons}}/\Gamma_{\mathrm{tot}} \approx 65\,\%$), and since tau-lepton identification usually operates at smaller efficiencies, hadronic remnants of a tau decay might get "wrongly" reconstructed as hadronic jets originating from the hard interaction.

This section is supposed to support the statement of the (dominant) background contribution, i.e., Drell-Yan in this case, being completely negligible by directly comparing the event yield of the multijet and Drell-Yan production for several regions in phase space for all counting-based observables. This comparison is done by means of a very simplified and superficial analysis based on events generated at LO using MadGraph5_aMC@NLO (matrix element) and Pythia8 (subsequent parton showering, hadronization and decay to final state particles) with a simplified simulation of the ATLAS detector based on *Delphes*, which allows to *"perform[...] a fast multipurpose detector response simulation"* [216, 217, 218].

In Equation 5.7, the number of observed multijet events is unfolded and corrected for acceptance effects to obtain the cross section at particle level. The underlying assumption, however, is that contributions from other processes to the measured event yield are negligible in *all regions of phase space*. In general, the expected contribution from other processes must be subtracted from the observed event count *prior* to the unfolding step as shown in Equation 11.20 below

$$N_i^{\mathrm{PL}} = \sum_j \left(N_{\mathrm{DL}}^{\mathrm{multijet},j} - N_{\mathrm{DL}}^{\mathrm{bkg},j} \right) \left(\left[R^{-1} \right] \right)_{\mathrm{DL},j}^{\mathrm{PL},i} \cdot \frac{\epsilon_{\mathrm{fake},j}}{\epsilon_{\mathrm{eff},i}} \tag{11.20}$$

$$= \sum_j N_{\mathrm{DL}}^{\mathrm{multijet},j} \left(1 - r_j\right) \left(\left[R^{-1} \right] \right)_{\mathrm{DL},j}^{\mathrm{PL},i} \cdot \frac{\epsilon_{\mathrm{fake},j}}{\epsilon_{\mathrm{eff},i}}, \tag{11.21}$$

whereby the ratio r being defined as $r_j := N_{\mathrm{DL}}^{\mathrm{bkg},j} / N_{\mathrm{DL}}^{\mathrm{multijet},j}$. The number of events due to background processes is usually estimated by means of Monte Carlo simulations and/or extrapolations from well-defined and well-understood control regions.

In case of the multijet cross section measurement, there is almost no *irreducible* (i.e., processes that result in an identical constellation of particles in the final state at hadron level, but emerge from non-signal related interactions) background because all (real) jets are considered as a signal. There are, however, background contributions from *reducible* processes where the objects of interest, i.e., jets, are faked by other particles like, for instance, leptons.

[11] There is no background. All jets are signal by definition. Background my only arise from mistakenly reconstruct other objects (like leptons) as being a hadronic jet. To the extent that this statement is sufficient, the reader may safely skip this section without having to fear any disadvantages. This section is only supposed to support the statement of contributions from Drell-Yan being completely negligible in each phase space bin of the measurement.

In principle, each process whose final state particles might be falsely identified as hadronic jets may contribute to the observed event yield to a varying degree given the region of phase space. Hence, the question of whether events of non-related processes "pollute" the measurement is out of the question; the question to be answered is, whether this contribution is significant ($r_j \sim \mathcal{O}(1)$) and therefore need to be properly accounted for.

At the LHC, the most dominant background contribution to the multijet rate is due to Drell-Yan production. It is therefore sufficient to estimate the event yield of this process compared to the number of multijet events in order to check whether the contribution from background is significant. Of course, there are other processes that contribute – e.g., processes involving photons in the final state –; however, many of those processes are at higher-order and therefore negligible compared to Drell-Yan, or less likely to fake jets as, e.g., the identification of photons is very efficient.

Hard Subprocess In order to estimate the contribution of the dominant source of background, a total number of $10 \cdot 10^6$ ($10 \cdot 10^6$) generator level events have generated for multijet (Drell-Yan) using MadGraph5_aMC@NLO. Using MadGraph5's application programming interface, multijet events are produced via the command:

```
# MadGraph command multijet production
generate p p > j j
```

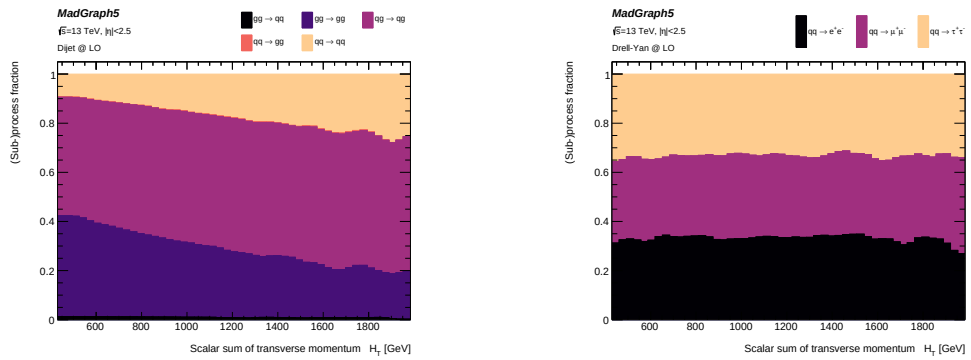
The above statement results in di-"jet" events ($2 \to 2$ process), i.e., two partons in the final state that are either gluons and/or (light) quarks (top quarks are excluded). For the simulation of Drell-Yan events,the following command was used:

```
# MadGraph command Drell-Yan production
define ell+ = e+ mu+ ta+
define ell- = e- mu- ta-
generate p p > ell+ ell-
```

The first two lines define the set of possible final state particles. In this setting, the mediator couples to electrons, muons as well as taus.

For both processes shown above, the very same PDF (NN23LO1) was used to sample the initial momenta for the incoming partons. Additionally, several (loose) generator-level cuts (like, e.g., a minimum p_T for the final state particles, constraining the pseudorapidity in anticipation of ATLAS' tacker acceptance etc.) have been applied to enhance the statistics in regions in phase space considered relevant for this analysis.

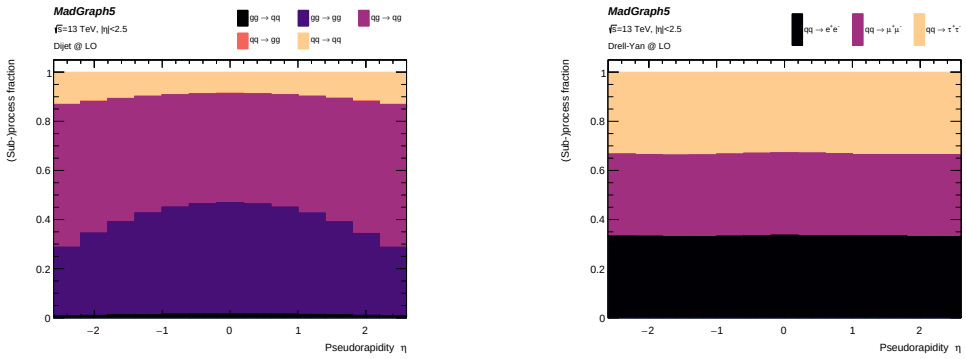
As part of the evaluation process, several kinematic distributions for various figures of merit



(a) Multijet *(b)* Drell-Yan

Plot 11.20: Fraction of sub-processes for multijet (*left*) and Drell-Yan (*right*) as a function of the scalar p_T sum H_T of both final state particles at generator level.

have been studies and compared[12]. As a small selection, consider Figure 11.20a that shows the fraction of different sub-processes for multijet (11.20b) and Drell-Yan (11.20b) production as a function of the scalar sum of the transverse momentum of the two particles in the final state $H_{\mathrm{T}} = p_{\mathrm{T}}^{(1)} + p_{\mathrm{T}}^{(2)}$. In case of multijet events, the dominant (sub-)processes contributing



(a) Multijet *(b)* Drell-Yan

Plot 11.21: Fraction of sub-processes for multijet (*left*) and Drell-Yan (*right*) as a function of the pseudorapidity η of both final state particles.

to the cross section are $gg \to gg$ and $qg \to qg$ for the chosen range of H_{T}. This is expected from the theoretical calculation of the corresponding channels/diagrams. Furthermore, the fraction of gluons in the final state decreases as a function of H_{T}, which is expected from the multijet cross section – this feature of the multijet cross section (i.e., decreasing number of gluon-initialized jets in the final state) will play an important role later once the underlying quark/gluon fractions are extracted – because with increasing p_{T} of the jet, the momentum fraction $x \propto p_{\mathrm{T}}/\sqrt{s}$ increase with an increased probability of finding/sampling a quark according to the proton's PDF.

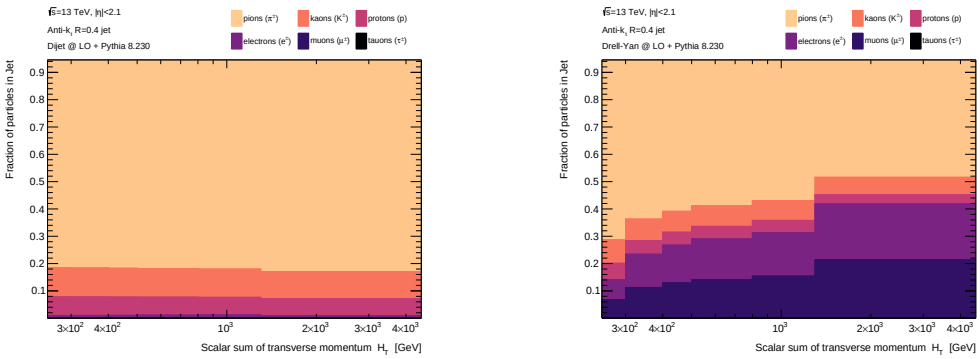
The corresponding figure for Drell-Yan-based events shows that the probability of generating an electrons, muons or taus is basically the same, which is expected from lepton universality given the range of energies.

Figure 11.21 also shows the fraction of different sub-processes but as a function of the particles' pseudorapidity (within the tracker acceptance of the ATLAS detector). In case of multijet production, a clear correlation between the "degree of contribution" and the respective process can be observed. After the previous chapters and section, this should not come as a surprise anymore since the observed characteristics is the exact same reason for measuring the triple-differential multijet cross section in central an forward region, i.e., the central region is enriched by gluons, while the forward region is quark enriched.

Parton Shower and Hadronization So far, the final state particles are at generator level at LO. To increase the particle (and jet) multiplicity in the final state and to improve the accuracy of the estimation of the cross section, a parton shower simulation (see Chapter 2.4) is applied to all generated events, using MadGraph's interface to Pythia 8.230. In case of time-like showers, the same PDF as for the matrix element calculation has been used. The configuration of the parton shower algorithm corresponds to Pythia's default tune and parton shower model, which is based on a sequence of k_t-ordered emissions. Besides the parton shower model, Pythia's Lund-string-based hadronization model (see Chapter 2.5) was used to simulate the fundamentally non-perturbative process of the formation of color-singlet states as well as the subsequent decay to stable final state hadrons (according to Pythia's decay tables).

[12]The generated events are saved in the Les Houches Event (LHE) format [182] that was analyzed using a tool provided by Ref. [219].

After the hadronization step, the outgoing particles of the calculation of the matrix element (hard subprocess) have been showered (parton shower) and are decayed (hadronization) to stable particles that can be measured in a detector. Figure 11.22 shows the fraction of the (stable) final state, *charged* particles for multijet (11.22a) and Drell-Yan (11.22b) events associated to jets that have been reconstructed at *particle level* using the anti-k_t clustering algorithm with $R = 0.4$. As expected, in case of multijet events, the final state is dominated



(a) Multijet *(b)* Drell-Yan

Plot 11.22: Fraction of (charged) particles in reconstructed jets at hadron level for multijet (*left*) and Drell-Yan (*right*) as a function of the scalar p_{T} sum H_{T}.

by charged mesons (mostly pions followed by kaons) and protons. However, also charged leptons are involved in hadronic showers. In case of Drell-Yan, the contribution of charged leptons is – of course – considerably higher. Taus are not directly reconstructed in Figure 11.22b due to their short lifetime, however, they contribute due to hadronic and leptonic decays.

Detector Simulation and Object Reconstruction So far, the generated events are at particle (or hadron) level – prior to the parton-shower step they are at *generator level*. At this stage, the background contribution to the multijet cross section due to Drell-Yan production would be very small because the latter one does not contain any jets – aside from jets originating from time-like (initial state) parton shower simulation (which are usually very soft in p_{T}) and hadronic tau decays. As mentioned previously, Drell-Yan production becomes a potential source of background as soon as the final state leptons might be incorrectly reconstructed and identified as jets of hadronic origin. The main reason for this to happen is the distortion of the particles' four-momenta once they traversing the measuring apparatus due to finite resolution, dead material etc. To estimate the effect of a reducible background source, a simulation of the detector is therefore unavoidable.

All samples listed in Table 8.1 have a full GEANT4 simulation of ATLAS detector applied, which is indispensable in case of the actual cross section measurement and the degree of precision that is aimed for. A full GEANT4 simulation, however, is very(!) time consuming and excessive for the modest demands required by this simple background-related study[13]. Fortunately, there is an alternative available, *Delphes*, that allows to perform simulations of the ATLAS detector at a level of precision completely sufficient for the modest demands required by this superficial study. *"Delphes[.]"*, to directly quote the authors, *"has been developed[.] for a fast simulation of a general-purpose collider experiment."* [220]. The fast simulation provided by *Delphes* includes the most crucial experimental features supposed to be accounted for in a detector simulation, such as the geometry of both central and forward detectors; the effect of magnetic fields on tracks; the reconstruction of photons, leptons, jets, b-jets, τ-jets(!) and missing transverse energy; a lepton isolation; a trigger emulation [220]; to name the most

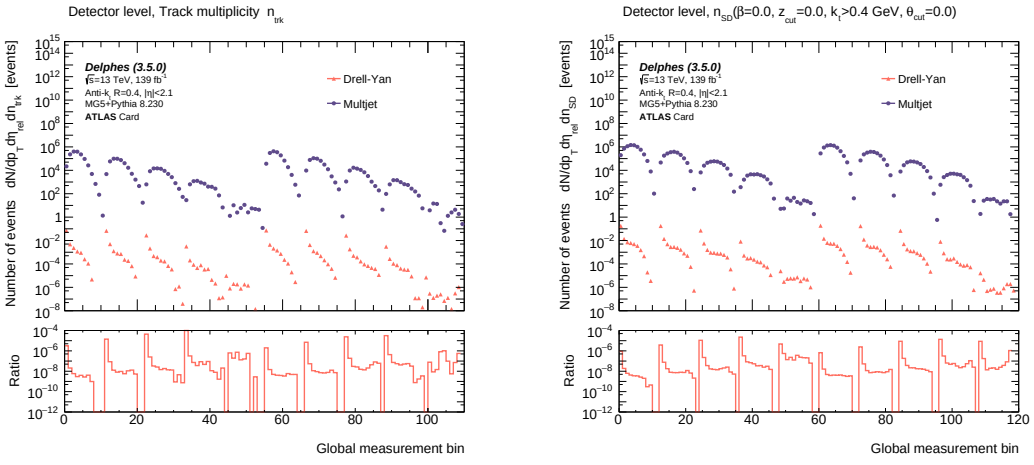
[13]Keep in mind that the estimated number of background-induced events does not enter the unfolding procedure – otherwise, a full GEANT4-based simulation of the background would be unavoidable. The objective of this section is to illustrate that there is no need account for background contributions in the first place.

important ones. Despite being great for simple and quick analyses, the precision and accuracy of detector simulations using *Delphes* has some limitations. Stating the *Delphes* authors literally again: *"[the] [d]etector geometry is idealized, being uniform, symmetric around the beam axis, and having no cracks [(particularly said for ATLAS' famous and beloved pseudorapidity crack ...)] nor dead material. Secondary interactions, multiple scatterings, photon conversion and bremsstrahlung are also neglected[.]"* [220].

Delphes' detector layout follows a generic multipurpose design like the one used by, e.g., ATLAS and CMS. It consists of several detector layers, starting with a tracking system most close to the beam line, surrounded by an electromagnetic and a hadron calorimeters (each with a central region and two endcaps) and two forward calorimeters. The last detector component is a muon spectrometer. To animate the detector, different detector cards can be loaded that *"allow[.] [to study] a large spectrum of running conditions by modifying basic detector parameters, including calorimeter and tracking coverage and resolution, thresholds or jet algorithm parameters[.]"* [220]. Of course, for this background related analysis, *Delphes* has been configured to mimic the ATLAS detector by using the available ATLAS detector card(s).

Once the simplified detector has been configured, the events generated in the previous step are passed through the simplified, *Delphes*-based ATLAS detector. The kinematics variables, i.e., the four-vector components of the final-state particles generated in the previous step are then smeared according to the tunable resolution of the different detector components. Because the measurement of the jet substructure observable is based on track, the reconstruction of charged particles' trajectories is of particular interest. In *Delphes*, each stable, charged particle with a transverse momentum above the threshold defined by the detector card (500 MeV at ATLAS) that is within the detector volume of the tracker results in a track, whereby some track reconstruction efficiency is applied that depends on the p_{T}. The four-vector components are undistorted. The simulation of the calorimeter response is simply given by the energy deposition of the particles within the active detector volume parameterized through a Gaussian smearing of the accumulated cell energy with a given variance defined by the detector configuration. The characteristic properties of the calorimeter like the varying resolution and granularity in different segments of the detector are specified in the corresponding detector card too.

Once the events are passed through the detector simulation, the analysis proceeds as



(a) n_{trk} *(b)* $n_{\mathrm{SD}}(k_t \geq 400\,\mathrm{GeV})$

Plot 11.23: Triple-differential cross section with respect to the track multiplicity (*left*) as well as $n_{\mathrm{SD}}(k_t > 400\,\mathrm{MeV})$ (*right*) for multijet and Drell-Yan production.

previously described in Chapter 9. The (multijet) event selection is applied and higher-level physical objects are reconstructed. Jets are build from calorimeter cells (which are available as a standard container in the output of the *Delphes*-based simulation of the ATLAS detector) using the anti-k_t clustering algorithm with a radius parameter of $R = 0.4$. If the jets fulfill the

criteria defined by the event selection (at least two jets with $p_{\mathrm{T}} > 250\,\mathrm{GeV}$, $p_{\mathrm{T}}^{\mathrm{lead}}/p_{\mathrm{T}}^{\mathrm{sublead}} < 1.5$ ect. see Table 9.3), tracks are matched to the calorimeter-based jets according to a ΔR-criterion; the "associated" tracks are then used to calculate the jet substructure observables of interest as it was done before. The same spectra of the triple-differential cross section are reconstructed for all multiplicities introduced in this chapter for both multijet and Drell-Yan production.

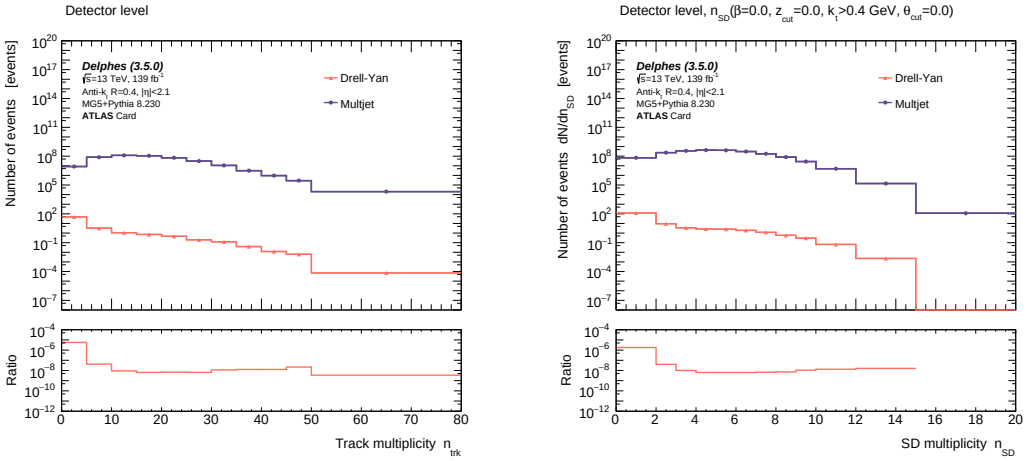
MadGraph5 then provides an estimation of the cross section $\sigma_{\mathrm{tot}}^{\mathrm{gen}}$ (including its statistical uncertainty). The luminosity (see Chapter 3.1) is not directly specified as it is a accelerator-dependent quantity. However, given the number of generated events (specified by the user) as well as the cross section approximation at LO provided by MadGraph5, the luminosity is simply given by $\mathcal{L}^{\mathrm{gen}} = N^{\mathrm{gen}}/\sigma_{\mathrm{tot}}^{\mathrm{gen}}$. Hence, weighting each event by a factor of $1/\mathcal{L}_{\mathrm{multijet/DY}}^{\mathrm{gen}}$ results in the correct cross section and therefore the right ratio between the production rate of multijet and Drell-Yan events. Finally, an additional factor of $\mathcal{L}_{\mathrm{Run2}}^{\mathrm{LHC}} (= 139\,\mathrm{fb}^{1})$ is applied to the event weight to account for the conditions during Run 2 of data-taking. This does not affect the result as the relative distance between the spectra of signal and background remains unchanged. However, it allows to compare the event yield in case of multijet events with the detector-level results of the previous section. The full event weight then reads

$$w_{\mathrm{X}} = \frac{\mathcal{L}_{\mathrm{Run2}}^{\mathrm{LHC}}}{\mathcal{L}_{\mathrm{X}}^{\mathrm{gen}}} = \mathcal{L}_{\mathrm{Run2}}^{\mathrm{LHC}} \frac{\sigma_{\mathrm{tot,X}}^{\mathrm{gen}}}{N_{\mathrm{X}}^{\mathrm{gen}}}, \tag{11.22}$$

whereby $\sigma_{\mathrm{tot}}^{\mathrm{gen}} \approx 4.1 \cdot 10^{4}\,p\mathrm{b}$ for multijet and $\sigma_{\mathrm{tot}}^{\mathrm{gen}} \approx 1.7 \cdot 10^{-2}\,p\mathrm{b}$. Because the calculation of the matrix element is at LO for both processes (multijet and Drell-Yan) and the parton shower simulation corresponds to Pythia's "Simple Shower", there are no non-trivial event weights at generator or hadron level due to complicated matrix-element-to-parton-shower matching etc. In this simple analysis, the application of the event weight according to Equation 11.22 therefore simply reduces to a scaling of the histograms, whereby the scale factor is different in case of multijet and Drell-Yan production.

Finally, after all those preparations, the triple-differential cross section for the track and SD multiplicities is reconstructed at detector level for the multijet- and Drell-Yan-based samples. The resulting spectra are shown in Figure 11.23. The individual spectra in Figure 11.23 show the *full* cross section measurement, i.e., in each transverse momentum, pseudorapidity and multiplicity bin (the two main structures correspond to the spectra in central and forward region, respectively). The ratio in each figure gives the ratio of the event yield in Drell-Yan with respect to the number of events measured in the multijet sample.

Inspecting the distributions shown in Figure 11.23 and comparing the respective spectra for signal and (reducible) background reveals a significant difference not only in the observed event yield of the two processes, but also regarding the shape of the multiplicity distributions. For reasons of clarity, Figure 11.24 shows the distributions of the track multiplicity (11.24a) as well as the SD multiplicity for all emissions with $k_t > 400\,\mathrm{MeV}$ (11.24b) marginalized over transverse momentum and pseudorapidity to obtain a single-differential distribution (als statements also hold for the other configuration of the SD multiplicity). Figure 11.24 provides a nice illustration of the distinctive features between signal and background. Starting with the shape of the distributions, the first characteristics that catches the eye is the shift towards smaller multiplicities (smaller number of tracks and emissions) in case of the background-related spectra due to Drell-Yan. This is hardly surprising, recalling that many jets reconstructed in case of the reducible background are actually leptons, being falsely identified and reconstructed as hadronic jets. Due to the characteristics of the electromagnetic shower, a smaller number of constituents is therefore expected; also, in case of hadronically decaying taus, the particle multiplicities is generally smaller. This implies that lower multiplicity bins in this measurement are generally more sensitive to contributions from background. A closer inspection of the ratio between signal and background, however, shows that this concern is unfounded as even in the phase space regions at risk (small multiplicities), the contribution to the cross section is below 1 ppm. Taking a look at the ratio for all phase space bins in Figure 11.23 shows that for some regions in phase space the contribution may reach 0.1 ‰, which is still completely negligible given the degree of precision of this measurement, which is (at its best) in the order of $\sim \mathcal{O}(1\,\%)$.



(a) n_{trk} *(b)* $n_{\mathrm{SD}}(k_t \geq 400\,\mathrm{GeV})$

Plot 11.24: Single-differential cross section with respect to the track multiplicity (*left*) and the SD multiplicity (*right*) for multijet events and for Drell-Yan.

11.5.2 Track Multiplicity n_{trk}^{500}

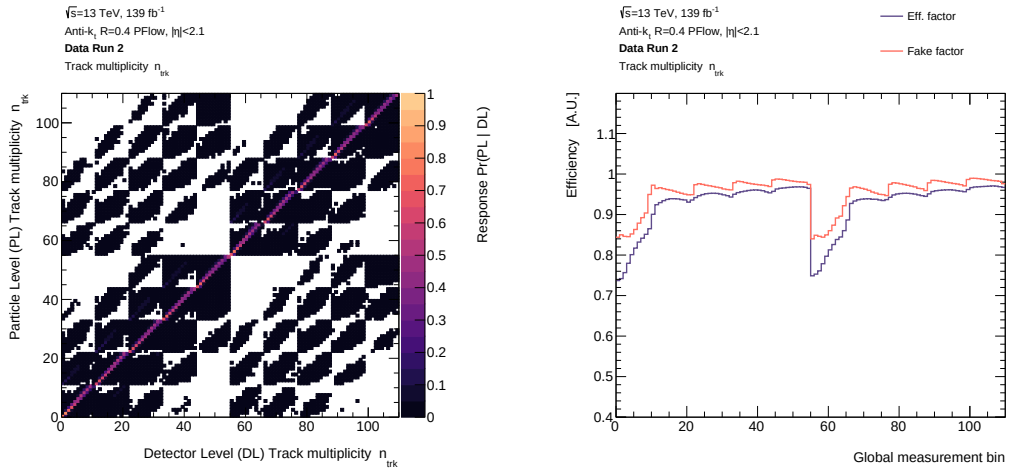
Turning the attention to the unfolding. Like it was the case for the previous section that introduced the measured spectra for all counting-based observables, the first jet substructure variable to consider for the unfolding-related results is the track multiplicity. The unfolding to particle level of the track multiplicity spectra will be dealt with in a little more detail compared to the SD multiplicity to introduce all relevant figures of merit as well as to provide some background information concerning the individual unfolding steps. Most of the insights gained in this first section, however, can be directly translated to the corresponding results for the SD multiplicity. Therefore, in case of n_{SD}, most results are presented without further comment, referring to this section.

As can be seen in Equation 11.20, the unfolding step involves three main incidents: *firstly*, the measured spectra of the multi-differential cross section at detector level that is supposed to be transformed particle level (see Section 11.4); *secondly*, the response/unfolding matrix that describes the correlation of the observable of interest at detector and particle level; *thirdly*, the efficiency and fake factors that account for efficiency losses due to migrations in and out of the phase space volume, respectively, over which the event rate is measured (see Figure 5.1).

The spectra for all jet substructure observables that are measured in this analysis have been introduced in Section 11.4 in detail.

The full response matrix of the *nominal*(=Pythia8) Monte Carlo as well as the efficiency and fake factors are shown in Figure 11.25. Those figures require some more explanation.

Figure 11.25a shows the full (properly normalized) response matrix of the measurement that is used to unfold the detector-level spectra to particle level. As already described in previous chapters, the response matrix is constructed by correlating the reconstructed value of the jet substructure observable at detector level to its counterpart at particle level, provided that both jets pass the event selection (otherwise, they contribute to the aforementioned efficiency or fake factors). The matrix obtained therefrom describes the fraction of jets within a certain phase space bin at particle level with respect to some phase space bin at detector level and therefore is a measure of the degree of migrations between different bins. Properly normalized (as done in Figure 11.25a), the response matrix can be interpreted as the *probability* of a substructure observable $\mathcal{O}^{\mathrm{PL}}_{i,\mathrm{JSS}}$ at truth level in the ith (global) bin being reconstructed as $\mathcal{O}^{\mathrm{DL}}_{j,\mathrm{JSS}}$ in bin number j at detector level (migrations taking place if $i \neq j$). Ideally, the amount of migrations between regions in phase space should be as small as possible because uncertainties are proportional to the fraction of the off-diagonal elements in the response matrix; furthermore, model-related biases are reduced. The latter point is due to the fact that



(a) Response matrix

(b) Efficiency and fake factors

Plot 11.25: Full response matrix (*left*) and efficiency and fake factors (*right*).

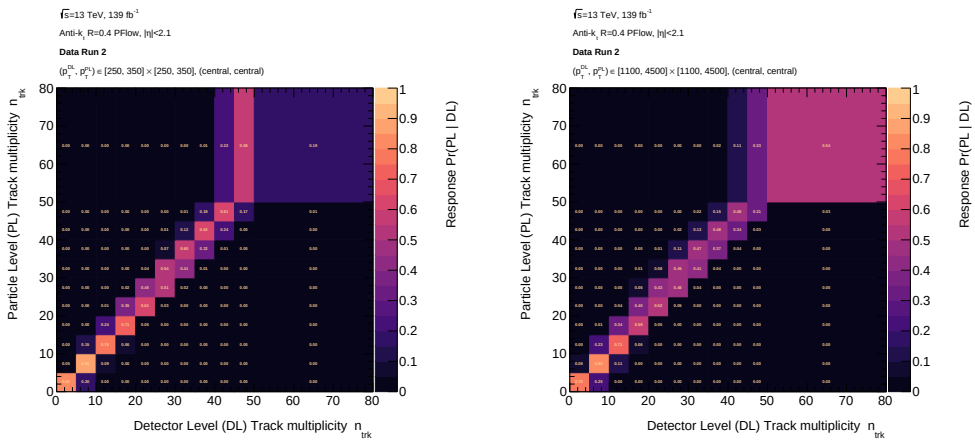
the response matrix is derived from simulation using the nominal Monte Carlo sample (PYTHIA in this case) because truth-level information must be available in order to build the response matrix in the first place. The model-dependent (using a different Monte Carlo event generator would result in a different response matrix) unfolding matrix, however, is used to unfold the measured spectra *in data* to particle level, which may result in bias in the unfolded cross section towards the nominal MC. This problem is addressed by the data-driven non-closure uncertainty (see Chapter 10.8). It is therefore desirable to obtain a response matrix *as diagonal as possible*.

The *diagonality*, i.e., the fraction of elements on the diagonal, can be adjusted mainly by the objects used to calculate the jet substructure observables as well as the binning, i.e., the granularity of the phase space discretization used for the measurement. The first aspect is the main reason for using tracks associated to the reconstructed PFlow jets for the calculation of the substructure observable instead of, for instance, calorimeter cells of PFlow objects directly. The resolution is *much worse* is case of calorimeter cells because many particles may easily contribute to the very same cell due to the comparatively coarse granularity of the hadronic calorimeter of the ATLAS detector. This results in a significant deterioration of the resolution of the reconstructed observable and therefore gives rise to considerable migrations between neighboring bins. Using tracks instead of, e.g., calorimeter cells, however, *significantly* improves the resolution of the measurement due to the excellent position resolution in the tracking system of the ATLAS detector and the possibility to resolve individual particles. Hence, using tracks for the reconstruction of the JSS observable maintains sufficiently strong correlations between particle and detector level, which is a basic requirement for the unfolding step to work in the first place. Another lever to control the degree of migrations is by means of the binning. This can easily be understood by recalling that migrations (mainly) occur due to distortions caused by the measuring apparatus, i.e., mismeasurement of the jet's transverse momentum due to lower activation thresholds in the calorimeter cells, dead material, contributions from pile-up, variations in the number of tracks etc. Of course, the sensitivity to those effects strongly depends on the granularity (the number of bins) at which the cross section is measured over the phase space volume. Hence, increasing the "coarseness" of the binning reduces the probability of jets migrating into different regions of phase space. However, reducing the number of bins results in a smaller *sampling* of the underlying target distribution; consequently, the information regarding the shape of the distribution is reduced as well. This aspect (regarding the interplay between phase space granularity and information concerning the distribution's shape) makes the optimization of the binning a very delicate matter because if the number of bins is chosen too small (low granularity), the discrimination power of the JSS observable

between quarks and gluons might be significantly reduced – otherwise, if the number of bins is chosen too large (i.e., high granularity), there will be a considerable amount of migrations between phase space regions, resulting in (possibly) very large systematic uncertainties and biases.

The matrix in Figure 11.25a shows the response for the *full* measurement, i.e., in all phase space bins for which the event rate has been determined. Its structure is similar to the one-dimensional spectra shown, for instance, in Figure 11.7[14]. As can be seen, there are migrations between several p_{T} regions. However, a significant fraction of jets is still contained on the diagonal of the matrix, indicating that the transverse momentum of most jets is correctly reconstructed given the (non-equidistant) width of the respective p_{T} bin (not visible in this representation). On closer inspection, an asymmetry in case of the off-diagonal elements can be observed (better seen in the single-differential response matrix in Figure 11.26). This asymmetry suggests that the probability of *underestimating* the transverse momentum of a jet at detector level is larger compared to overestimating it. This immediately makes sense given the fact that the multijet cross section is steeply falling with respect to the jet's p_{T}, which results in more jets migrating into lower p_{T} bins on average.

To make the non-equidistant binning explicit, it is reasonable to consider some examples



(a) $p_{\mathrm{T}} \in [250, 350)$ *(b)* $p_{\mathrm{T}} \in [1100, 4500)$

Plot 11.26: Individual response matrices for different p_{T} (in GeV) regimes in central region[15].

of the response matrix for certain pseudorapidity and p_{T}, similarly to the detector level distributions of the cross section in the previous section. The result is shown in Figure 11.26. In this representation, the non-equidistant binning of the multiplicity (see Table 11.2) is explicitly visible. In total, Figure 11.25a is made up of $\left(N_{p_{\mathrm{T}}}^{\mathrm{bins}} \cdot N_{\eta}^{\mathrm{bins}}\right)^2$ distributions like the ones shown

[14]Regarding the structure of the full response matrix, the matrix can be grouped into three categories, corresponding to the *three* observables with respect to the *triple*-differential cross section is measured. On the outer level, the matrix can be divided into four blocks (▪), each one featuring a similar structure with varying density. Those four groups correspond to the two pseudorapidity regions (central and forward). The "diagonal" structures ($[0, N_{p_{\mathrm{T}}\mathcal{O}_{\mathrm{JSS}}}^{\mathrm{bins}}]^2$ and $[N_{p_{\mathrm{T}}\mathcal{O}_{\mathrm{JSS}}}^{\mathrm{bins}}, 2N_{p_{\mathrm{T}}\mathcal{O}_{\mathrm{JSS}}}^{\mathrm{bins}}]^2$ with $N_{p_{\mathrm{T}}\mathcal{O}_{\mathrm{JSS}}}^{\mathrm{bins}} := N_{p_{\mathrm{T}}}^{\mathrm{bins}} \cdot N_{\mathcal{O}_{\mathrm{JSS}}}^{\mathrm{bins}}$) describe central (forward) jets at particle level being correctly reconstructed as central (forward) jets at detector level. As expect, the degree of migrations between the different pseudorapidity categories is very small because there are only two bins; furthermore, the angular components of the jets' four momentum are measured with high precision compared to the scalar entries. Therefore, the field $[0, N_{p_{\mathrm{T}}\mathcal{O}_{\mathrm{JSS}}}^{\mathrm{bins}}] \times [N_{p_{\mathrm{T}}\mathcal{O}_{\mathrm{JSS}}}^{\mathrm{bins}}, 2N_{p_{\mathrm{T}}\mathcal{O}_{\mathrm{JSS}}}^{\mathrm{bins}}]$, which corresponds to forward jets at particle level being (falsely) reconstructed as central at detector level is of low density. The next organizational level (⟋) corresponds to the individual transverse momentum bins and shows the migrations between those. Finally, the "innermost" structure (⟋) shows each individual multiplicity bin by means of the global bin index.

[15]The labels of the subfigures are an abbreviation for $(p_{\mathrm{T}}^{\mathrm{PL}}, p_{\mathrm{T}}^{\mathrm{DL}}) \in [p_{\mathrm{T}}^{\mathrm{min}}, p_{\mathrm{T}}^{\mathrm{max}}) \times [p_{\mathrm{T}}^{\mathrm{min}}, p_{\mathrm{T}}^{\mathrm{max}})$ because they refer to the diagonal elements of the full response matrix in Figure 11.25a

in Figure 11.26 [16]. Furthermore, this representation allows to study the evolution of the response matrix as a function of the transverse momentum.

As can be seen, the response matrix exhibits the desired features, that is – first of all –, a significant fraction of elements being positioned on the diagonal[17]; furthermore, the binning of the track multiplicity is comparatively fine. As the transverse momentum increases, the number of constituents (and thus the tracks) inside of the jets increases too, as can be seen by, for instance, comparing Figure in 11.8a with Figure 11.8b. This effect is also reflected in Figure 11.26b that shows the response matrix for jets with $p_{\mathrm{T}} \in [1100\,\mathrm{GeV}, 4500\,\mathrm{GeV})$. In this p_{T} regime, all track multiplicity bins are now populated. The matrix is still sufficiently diagonal, however, compared to the lower-p_{T} one in Figure 11.26a, the resolution slightly deteriorates, which can be seen by the modest increase of off-diagonal elements. This is expected as the resolution of track reconstruction decreases with increasing momentum of the charged particle's trajectory. Nonetheless, the response matrix (now going back to the full measurement in Figure 11.25a) is of good quality given the fact that the measurement is three-dimensional, i.e., the jets are able to migrate in three "directions".

As can be seen in Equation 11.20, the response matrix is only one ingredient that enters the unfolding. The measured number of events must also be corrected for acceptance and efficiency effects to account for jets that mistakenly migrate into or outta the (fiducial) phase space volume. The former ones are described by the *fake factor*, the latter ones are given by the *(selection) efficiency factor*. Both numbers are determined in simulation using the nominal sample in each phase space bin of the measurement. The result for the track multiplicity is shown in Figure 11.25b.

Starting with a description of the curve progression, both, the fake and efficiency factors, are shown *versus* the global measurement bin similar to Figure 11.7b, including the central and forward region. The repeating structures once, again, correspond to the different multiplicities for the various transverse momentum bins for central ($I \in [0, 55(= N_{p_{\mathrm{T}}}^{\mathrm{bins}} \cdot N_{n_{\mathrm{trk}}}^{\mathrm{bins}})$) and forward ($I \in [55, 110]$) jets. All values are bounded from above by one as expected for an efficiency. Furthermore, the efficiency increases with p_{T}, whereby a considerable efficiency drop in the very first p_{T} bin can be observed. This behavior is anticipated: as already mentioned, due to the steeply falling spectra, jets are more likely to migrate into regions of smaller transverse momentum. For jets in the very fist p_{T} bin, however, there is no lower-p_{T} bin to migrate into. As a consequence, all jets whose truth transverse momentum is within the p_{T} range defined by the very first momentum bin, but with a reconstructed p_{T} below the lower momentum cut, do not enter the measurement of the cross section *at all*. This would, of course, result in a systematic underestimation of the cross section, which must be accounted for by a (lower) efficiency in the corresponding bin. With increasing p_{T}, the jets are less-and-less likely to fall below the lower momentum threshold, which is why the efficiency increases with the transverse momentum of the jets. As described in Chapter 10, the efficiency and fake factors also play an important role in the calculation of the model-related systematic uncertainties. This is done by comparing the efficiency and fake factors between different Monte Carlo generators (the histograms in Figure 11.25a as well as Figure 11.25b are produced for *all* generators listed in Table 8.1) as they needed to calculate the systematic uncertainties explained in Chapter 10.

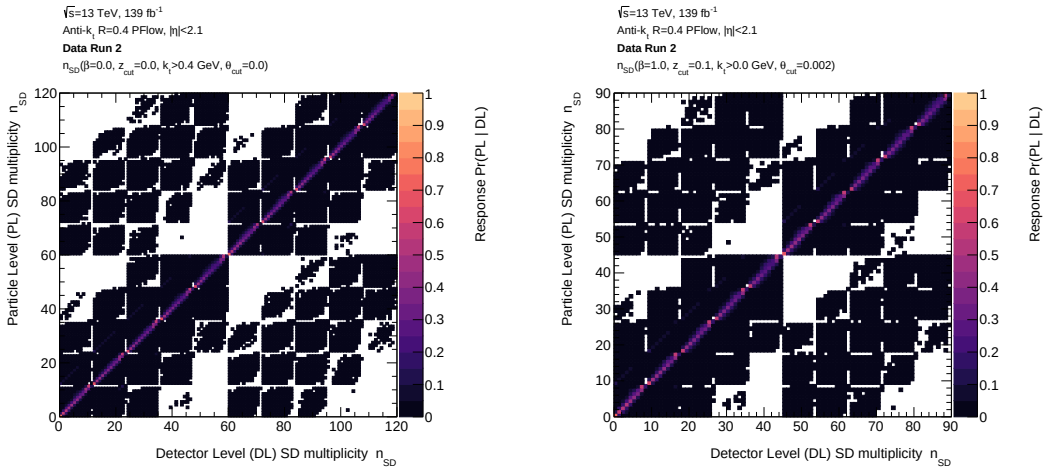
[16] The two distributions in Figure 11.26 show examples of the response matrices on the *diagonal* with respect to the transverse momentum as well as the relative pseudorapidity. For instance, Figure 11.26a shows the response matrix (single-differential with respect to n_{trk}) for particle- and detector-level jets being correctly identified as central (i.e. no migrations between the pseudorapidity categories) as well as with a reconstructed transverse momentum with $p_{\mathrm{T}} \in [250\,\mathrm{GeV}, 350\,\mathrm{GeV})$ (no migration between p_{T} bins). Of course, a corresponding response matrix exist for, e.g., both jets (at particle- and detector-level) being identified as central, but with migrations between p_{T} bins, for instance, $[250\,\mathrm{GeV}, 350\,\mathrm{GeV}) \times [350\,\mathrm{GeV}, 500\,\mathrm{GeV})$, i.e., the jet's p_{T} being underestimated at detector level, giving rise to migrations between adjacent bins.

[17] One may argue whether the fraction on the diagonal is significant or not. However, one should keep in mind that the measurement is three-dimensional, i.e., the jets may migrate in three "directions".

11.5.3 SD Multiplicity n_{SD}

Following the structure of Section 11.3, the next observable in line is the SD multiplicity with the two configuration of the (I)SD algorithm measured in this analysis. Due to the great similarity regarding the (rather) technical aspects of the unfolding between the SD multiplicity and the track multiplicity (along with the insights gained therefrom), most outcome are shown without providing elaborate explanations, but with reference to the corresponding figure in case of n_{trk}. Nonetheless, the SD and track multiplicity differ in several aspects and therefore require suited explanations if appropriate.

Once again, the first quantities to examine are the principal elements that enter the unfolding step, i.e., the response matrix as well as the efficiency and fake factors. The full response matrices are shown in Figure 11.27 for both configurations of the SD multiplicity. Similarly, the corresponding efficiency and fake factors are shown in Figure 11.28 for each



(a) $n_{\mathrm{SD}}(k_t \geq 400\,\mathrm{MeV})$ *(b)* $n_{\mathrm{SD}}(z_{\mathrm{cut}} = 0.1, \beta = 1, \theta_{\mathrm{cut}} = 0.002)$

Plot 11.27: Full response matrix for $n_{\mathrm{SD}}(k_t \geq 400\,\mathrm{MeV})$ (*left*) and $n_{\mathrm{SD}}(z_{\mathrm{cut}} = 0.1, \beta = 1, \theta_{\mathrm{cut}} = 0.002)$ (*right*).

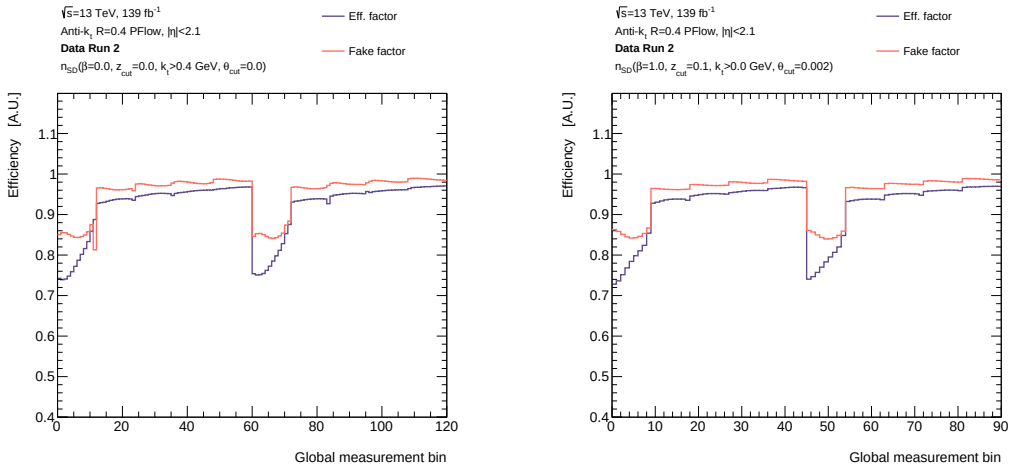
phase space bin of the measurement.

Starting with the response matrices in Figure 11.27, the first thing to recognize is the great resemblance with the corresponding Figure 11.25a for the track multiplicity in the global measurement bin representation. All statements previously provided regarding the full response matrix in case of the track multiplicity remain valid and can be converted to both configuration of the SD multiplicity n_{SD} one to one[18]. Compared to Figure 11.25a, the response matrices for the SD multiplicities look slightly less sparse; this is a consequence of the different binning used in case of n_{SD}. The difference between SD and track multiplicity are visible after transforming the global bin representation to the one in which the non-equidistant binning is explicit.

To continue a good and old tradition, Figure 11.29 shows the example of two response matrix with non-equidistant binning for the first configuration of the SD multiplicity with a lower k_t-cut being applied. Regarding the efficiency and fake factors in Figure 11.28, the same features as in the corresponding Figure 11.25b for the track multiplicity can be observed.

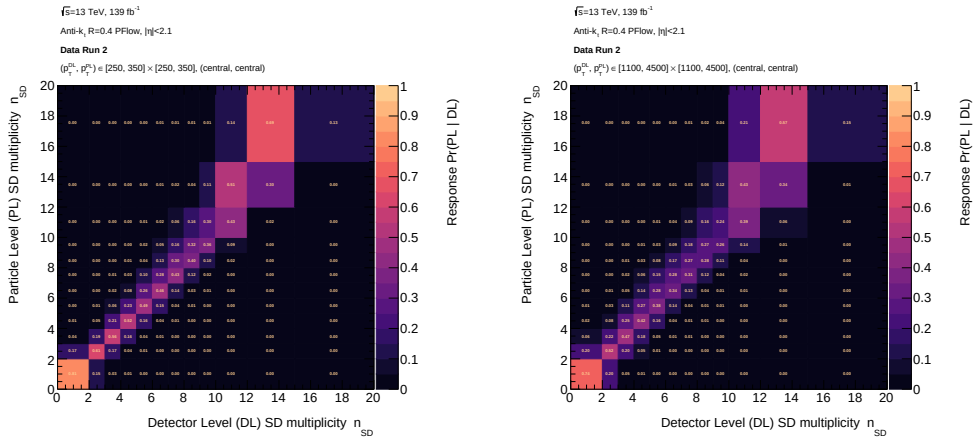
The following Figure 11.30 shows the same information but for the second configuration of the SD multiplicity $n_{\mathrm{SD}}(z_{\mathrm{cut}} = 0.1, \beta = 1, \theta_{\mathrm{cut}} = 0.002)$. The binning can be verified by comparing the axes' range with, e.g., the one in Figure 11.14 that show the reconstructed SD multiplicity at detector level for the very same p_{T} bins. Closer inspecting the distributions in Figure 11.29 and Figure 11.30 show the already familiar evolution of n_{SD} with increasing

[18] In fact, the response matrices only differ at the innermost structure on the level of the individual jet substructure observable bin because the double-differential multijet cross section is identical for all JSS observables studied in this thesis.



(a) $n_{\mathrm{SD}}(k_t \geq 400\,\mathrm{MeV})$ *(b)* $n_{\mathrm{SD}}(z_{\mathrm{cut}} = 0.1, \beta = 1, \theta_{\mathrm{cut}} = 0.002)$

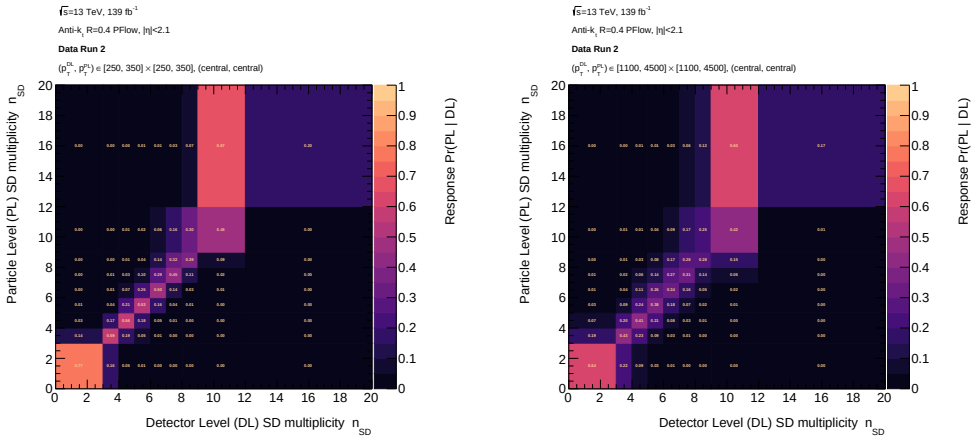
Plot 11.28: Full efficiency and fake factors for $n_{\mathrm{SD}}(k_t \geq 400\,\mathrm{MeV})$ (*left*) and $n_{\mathrm{SD}}(z_{\mathrm{cut}} = 0.1, \beta = 1, \theta_{\mathrm{cut}} = 0.002)$ (*right*).



(a) $p_{\mathrm{T}} \in [250, 350)$ *(b)* $p_{\mathrm{T}} \in [1100, 4500)$

Plot 11.29: Individual response matrices for different p_{T} regimes in central region for $n_{\mathrm{SD}}(k_t \geq 400\,\mathrm{MeV})$.

transverse momentum: the (average) number of emissions increases like the number of tracks (charged constituents) increase too. In case of low-p_{T} jets – like the situation shown in Figure 11.29a/11.30a –, the highest multiplicity bins are sparsely populated. Generally, the response matrices show the desired feature, i.e., most probability "mass" being on the diagonal. However, migrations can be observed, predominantly towards lower multiplicities as expected. The fraction of entries beyond/off the diagonal increases for higher-momentum jets as can nicely be seen by directly comparing, e.g., Figure 11.29a and Figure 11.29b. This is a consequence of the already mentioned decreasing resolution of the tracks with increasing p_{T}.



(a) $p_{\mathrm{T}} \in [250, 350)$ *(b)* $p_{\mathrm{T}} \in [1100, 4500)$

Plot 11.30: Individual response matrices for different p_{T} regimes in central region for $n_{\mathrm{SD}}(z_{\mathrm{cut}} = 0.1, \beta = 1, \theta_{\mathrm{cut}} = 0.002)$.

11.6 Particle-Level Distributions

In Section 11.4, the triple-differential multijet cross section with respect to the transverse momentum, the relative pseudorapidity and one of the counting-based observables has been measured at detector level for real *pp*-collision data recorded with the ATLAS detector between 2015 and 2018 (inclusive) and simulation for several Monte Carlo generators. The measured multijet spectra have been interpreted and several plausibility checks have been performed; furthermore, the data-to-MC agreement was qualitatively evaluated. Special attention was payed to the ratio between the double-differential multijet cross section $\mathrm{d}N/\mathrm{d}p_{\mathrm{T}}\mathrm{d}n_{\mathrm{count}}$ in central and forward region because both differ with respect to their composition/fraction of gluon- and quark-initialized jets, respectively. This information is used to extract the underlying quark and gluon distributions as well as their fractions according to the operational definition of quark and gluon jets (see Definition 2) using the "jet topics" approach.

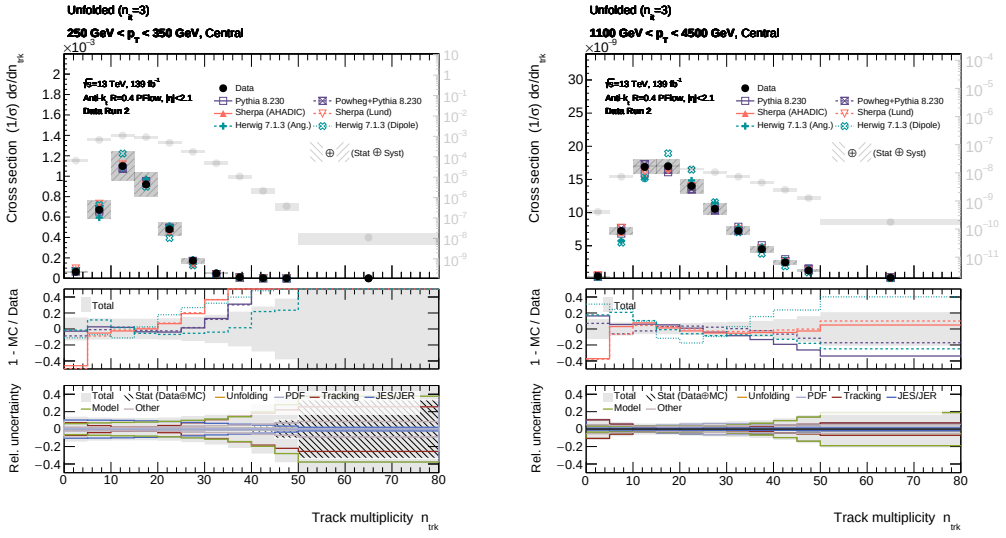
In the subsequent Section 11.5, the technical aspects of the unfolding have been discussed. The unfolding step intends to correct the measured spectra for detector and acceptance effects to obtain the corresponding distributions at particle level, which allows a direct comparison with predictions from theory (i.e., the output of some Monte Carlo event generator or analytic calculations if the observable/measurement is IRC-safe) without the need for detector folding. The *un*folding involves the inversion of the response matrix to transform the particle-level distribution to hadron level, as well as the correction for acceptance effects to account for efficiency losses and fakes due to migration in and out the active phase space volume over which the measurement is performed. In this context, the full response matrix was presented for all counting-based observables along with the efficiency and fake factors in each phase space bin. Besides the rather technical aspects of the unfolding, it was motivated that the contribution from the most dominant "reducible" source of background (Drell-Yan) is completely negligible for all regions in phase space considered in this analysis.

Once the detector-level spectra have been unfolded, the measured inclusive triple-differential multijet cross section can be compared to the truth-level prediction from Monte Carlo. Additionally, with the response matrices as well as efficiency and fake factors being available for all generators in Table 8.1 and all systematic variations (e.g., pile-up *re*weighting, JES/JER, modeling etc.), the statistical and systematic uncertainties can be calculated as described in detail in Chapter 10.

This section presents the *unfolded* triple-differential multijet cross section for *all* counting-based observables including all statistical uncertainties as well as all sources of systematic uncertainties considered in this measurement.

11.6.1 Track Multiplicity n_{trk}^{500}[19]

The first jet substructure observable to consider is the track multiplicity. The triple-differential multijet cross section with respect to n_{trk} at particle level is obtained by *simultaneously* unfolding the corresponding multijet spectra at detector level in Figure 11.7 and using the response matrix in Figure 11.25a as well as applying the correct efficiency and fake factors shown in Figure 11.25b. The result of the unfolded cross section for the *full* measurement is shown in Figure 11.39. *However*, before studying the relative event rate in all phase space bins (which basically summarized the result of the entire measurement), it is instructive to take a look at the single-differential distributions in some selected regions in transverse momentum beforehand, with the intention to gain some intuition for the uncertainties. This way, the cross section and its uncertainty can be studied with respect to the track multiplicity in a representation with the non-equidistant binning being explicitly visible. The result of the unfolded triple-differential cross section along with its uncertainties is shown for *central* jets in Figure 11.31 and forward jets in Figure 11.32 in two increasing-p_{T} bins in transverse momentum – $p_{\text{T}} \in [250\,\text{GeV}, 350\,\text{GeV})$ (11.31a/11.32a) and $p_{\text{T}} \in [1100\,\text{GeV}, 4500\,\text{GeV})$ (11.31b/11.32b).



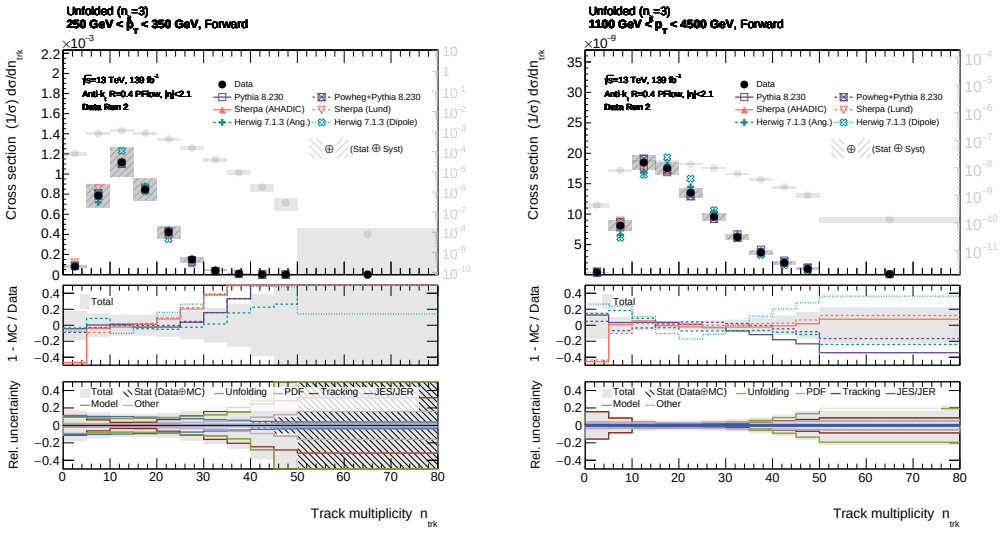
(a) $p_{\text{T}} \in [250, 350)$ *(b)* $p_{\text{T}} \in [1100, 4500)$

Plot 11.31: Individual *unfolded* single-differential multijet cross sections in bins of the track multiplicity n_{trk} for different p_{T} (in GeV) regimes in *central* region.

Figure 11.31 and Figure 11.32 show the unfolded multijet cross section as a function of the track multiplicity including all sources of statistical and systematic uncertainties as well as the ratio between all Monte Carlo samples and the unfolded spectra obtained from data. As it was the case for the detector-level spectra, the average number of tracks inside of a jet increases with transverse momentum and the width of the distribution broadens. Furthermore, the central distributions are shifted towards larger multiplicities compared to the forward ones, reflecting the different composition of quark- and gluon-initialized jets in both categories.

Concerning the systematic and statistical uncertainties, the measurement is dominated by the modeling-related uncertainties, the uncertainty due to the calibration of the jet energy scale (comprehensively described in Chapter 4.4.3) as well as the uncertainty related to the tracking. As explained in Chapter 10, the modeling-associated component of the total systematic uncertainty includes a component related to the precision of the underlying matrix element (ME), the parton shower (PS) and the hadronization (Had) model. Each variation of

[19]The corresponding numerical values of the relative triple-differential multijet cross section with respect to this observable are provided in Table B.2 in the Appendix of this report.



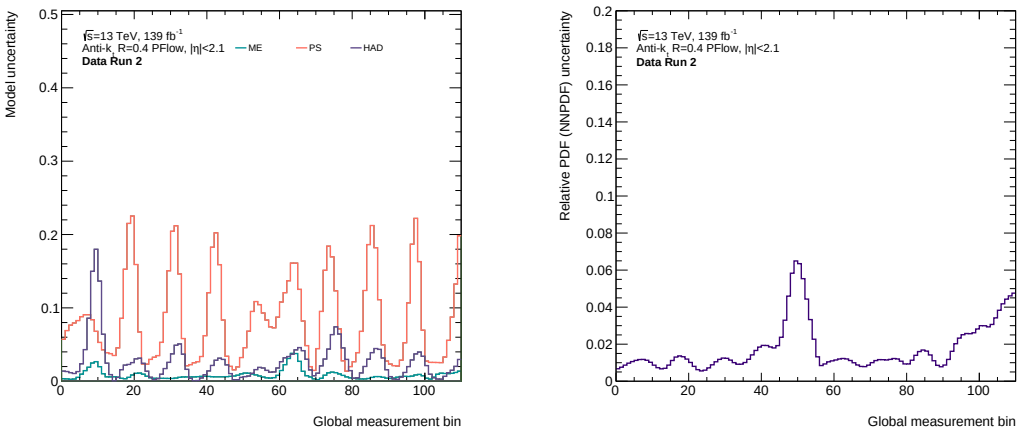
(a) $p_{\mathrm{T}} \in [250, 350)$ *(b)* $p_{\mathrm{T}} \in [1100, 4500)$

Plot 11.32: Individual *unfolded* single-differential multijet cross sections in bins of the track multiplicity n_{trk} for different p_{T} (in GeV) regimes in *forward* region.

the model in either of the different simulation steps translates into a variation of the triple-differential multijet cross section with respect to the one obtained using the nominal sample. This variation is associated with an uncertainty, reflecting – with a touch of over-simplification – the limited understanding of the underlying mechanism at work. The statistical uncertainty is only relevant in the higher multiplicity bins, but it reduces considerably for higher-momnetum jets. Furthermore, the statistical uncertainty in the highest multiplicity bin is larger in case of forward jets (compare Figure 11.31a and Figure 11.32a) because the spectrum is quark enriched, resulting in smaller number of particles on average.

Following the "evolution" of the uncertainty related to the parton shower for the two bins of transverse momentum in Figure 11.31 and Figure 11.32, one immediately notices its reduction with increasing transverse momenta of the jets. This behavior is expected, recalling the underlying premises and (fundamental) assumptions that enter the parton shower approximation. The basic idea common to *all* parton shower models was explained in Chapter 2.4. In a nutshell: the parton shower simulation is based on the observation of the multi-final state cross section being factorable into a cross section of smaller final state multiplicity times a factor that accounts for the emission of *collinear* radiation. In the limit of collinear radiation, most practical approaches to parton shower coincide; in particular, all definitions of the ordering variable for the emissions become equivalent in the collinear limit. Differences between the various models arise because the parton shower approach is also used *beyond* the collinear approximation defined by the matching and merging scheme. Consequently, the PS-related uncertainty reduces if the radiation angle becomes smaller (more collimated), which is the case for higher-p_{T} jets.

Because the modeling-related uncertainty is among the dominant sources of systematic uncertainties that limits the precision of this measurement, Figure 11.33a shows the uncertainty associated with the limited precision of the matrix element, the simulation of the parton shower as well as the hadronization step as a function of the global measurement bin for the entire measurement. Most of the other uncertainties show a somewhat "weaker" dependence on the transverse momentum. For instance, the hadronization-related uncertainty is in the order of just a few percent for almost all regions in phase space. The uncertainty associated with the limited precision of the matrix element calculation is almost negligible as can be seen in Figure 11.33a. Only the PDF-related uncertainty (besides the PS and the JES) show a noticeable dependence on the jet's p_{T} as can be seen in Figure 11.33b. Figure 11.33b nicely illustrates the



(a) Model uncertainties *(b)* PDF uncertainty

Plot 11.33: Model (matrix element, parton shower and hadronization) uncertainty (*left*) and PDF uncertainty (*right*).

uncertainty associated with the choice of the PDF in each phase space bin of the measurement. For most regions of phase space, the uncertainty is in the order of 1 % except for the very last bins that corresponds to very high-p_{T} jets. The eye-catching progression of the PDF-related uncertainty in Figure 11.33b for the individual single-differential distributions with respect to n_{trk} is explained by the different contributions of quarks and gluons: higher multiplicity are generally gluon-enriched, which results in larger systematic variations compared to smaller number of tracks associated with quarks. This result is expected but still interesting as it opens a potential window for the reduction of the gluon-related uncertainties of the PDF. This aspect will be further developed in the proposal of "future research suggestions" in the end of this report.

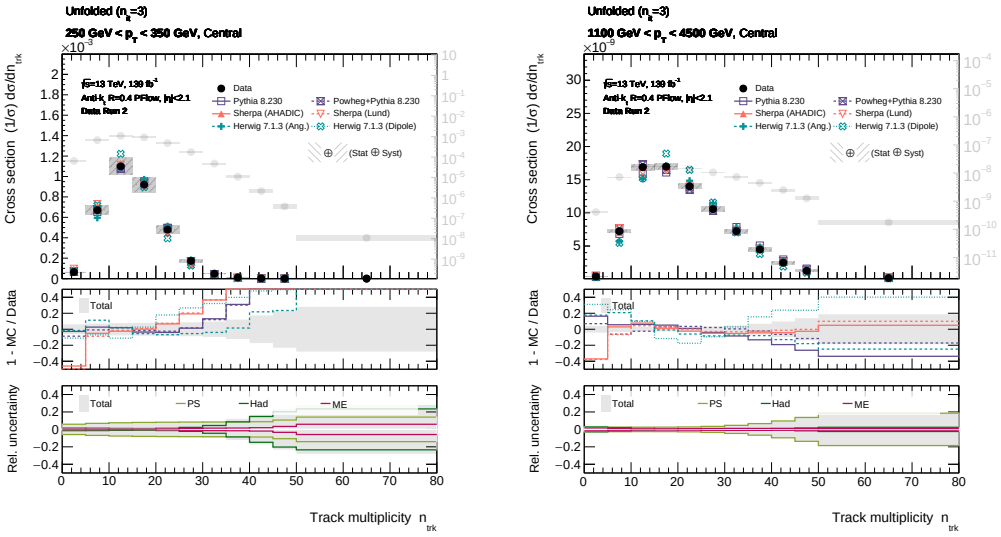
The following figures focus on individual uncertainties shown in Figure 11.31 for better visualization. The fist one shows once again the individual components of the model-related uncertainty.

In Figure 11.31 and Figure 11.32, several systematic uncertainties of lower contribution to the total uncertainty of the measurement have been summarized to one group whose individual components are shown in Figure 11.35. The uncertainty includes the data-driven non closure (not to be confused with the uncertainty due to the unfolding) (see Chapter 10.9) that quantifies the stability of the measurement in terms of different sub campaigns; the uncertainty due to various pile-up conditions (see Chapter 10.6); as well as variations of the cross section due to disabled tile modules (see Chapter 10.7).

Another important component of the total uncertainty is the uncertainty related to tracks. Uncertainties in the tracking must be taken into account because the measurement of the jet substructure observable is based on tracks that have been matched to PFlow jets. For a quick reminder of the tracking-related uncertainties, see Chapter 10.3. The total tracking-related uncertainty as well as its individual components (which have already been shown in Section 11.4 and Figure 11.11). The final results for the tracking uncertainties are summarized in Figure 11.36 for two bins of transverse momentum.

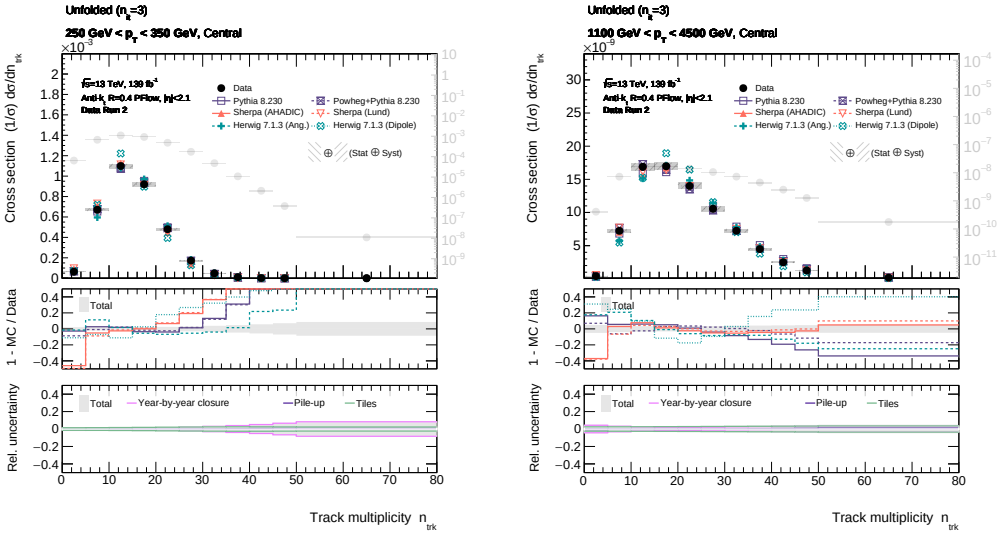
As previously mentioned, the precision of the measurement is – for the most part – determined by the modeling, JES and tracking-related uncertainties. The JES uncertainty itself consists of several nuisance parameters as has been explained in great detail in Chapter 4.4.3. Those components of the JES uncertainty account for different physical effects like, e.g., the uncertainty related to the detector response for different jet flavors.

The total uncertainty (including all sources of systematic and statistical uncertainties) of the triple-differential cross section measurement with respect to the track multiplicity is shown



(a) $p_{\mathrm{T}} \in [250, 350)$ *(b)* $p_{\mathrm{T}} \in [1100, 4500)$

Plot 11.34: Model-related uncertainty for two individual single-differential multijet cross sections in bins of the track multiplicity n_{trk} for different p_{T} regimes (in GeV).



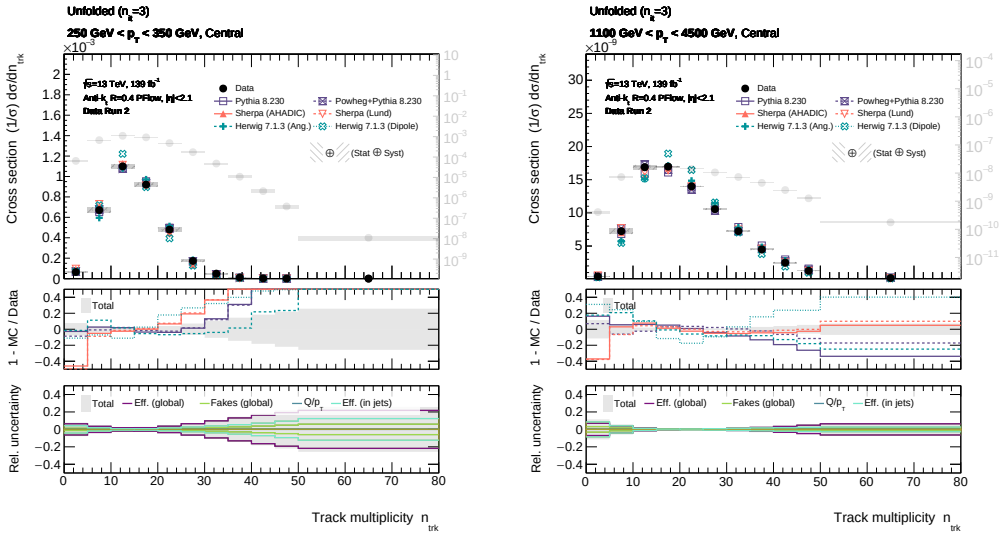
(a) $p_{\mathrm{T}} \in [250, 350)$ *(b)* $p_{\mathrm{T}} \in [1100, 4500)$

Plot 11.35: Other systematic uncertainties with lower contribution to the total uncertainty for two individual single-differential multijet cross sections in bins of the track multiplicity n_{trk} for different p_{T} regimes (in GeV).

in Figure 11.37.

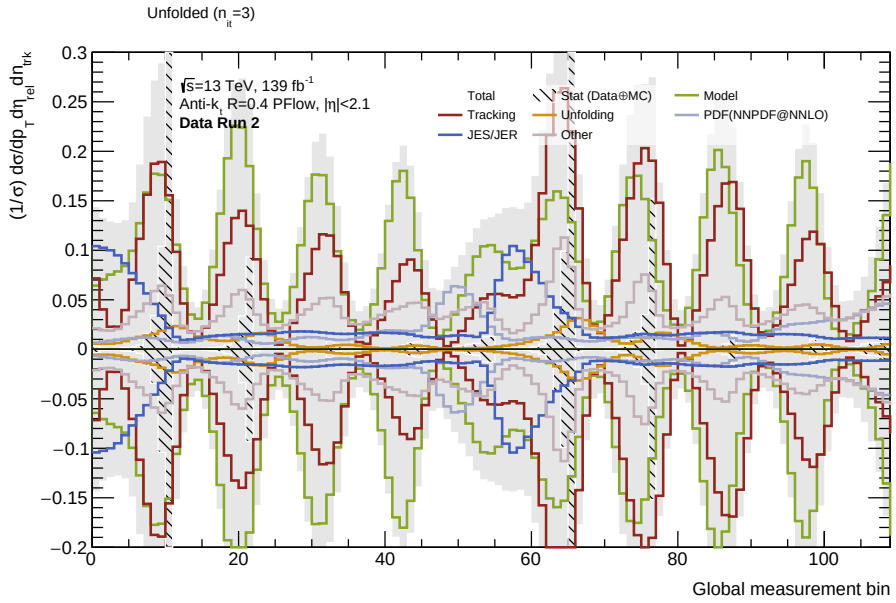
In light of the extraction of the underlying quark/gluon distributions and fractions using the procedure described in Chapter 7, Figure 11.38 shows the ratio of the unfolded spectra in central and forward region including all sources of systematic and statistical uncertainties for two bins in transverse momentum.

The uncertainties of the cross section ratio in Figure 11.38 are calculated assuming the systematic uncertainties of the central and forward category being fully correlated; the statistical uncertainties are assumed to be uncorrelated. The ratios show the expected behavior



(a) $p_{\mathrm{T}} \in [250, 350)$ *(b)* $p_{\mathrm{T}} \in [1100, 4500)$

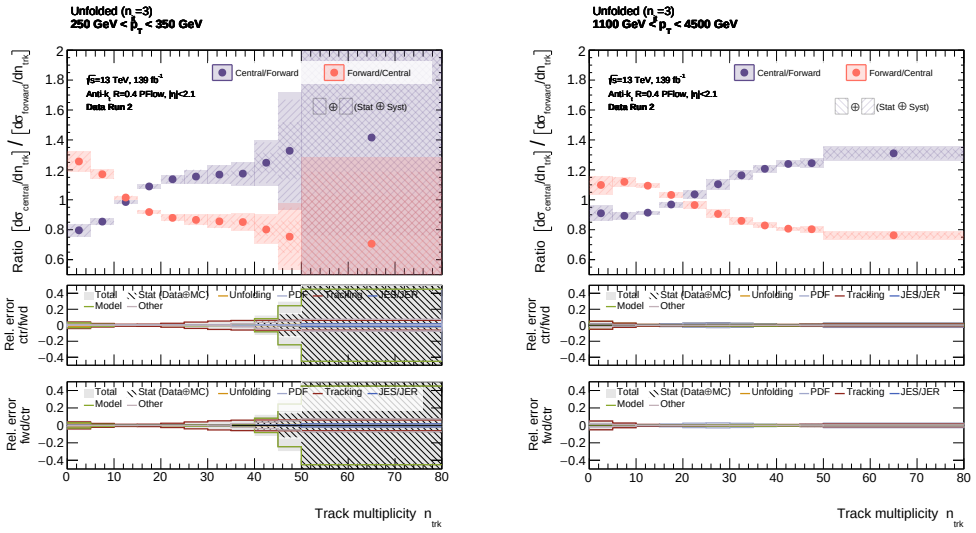
Plot 11.36: Tracking-related uncertainty for two individual single-differential multijet cross sections in bins of the track multiplicity n_{trk} for different p_{T} regimes (in GeV).



Plot 11.37: All sources of systematic and statistical uncertainties of the triple-differential multijet cross section measurement as a function of the track multiplicity.

already seen in the corresponding histograms obtained from the detector-level distributions, i.e., they (mostly) monotonically increase for higher track multiplicities, indicating that the corresponding phase space region is getting purer in gluon-like jets.

The two single-differential distributions in Figure 11.31 only show a section of the full triple-differential cross section that is shown in Figure 11.39 as a function of the global (phase space) measurement bin. Figure 11.39 shows the unfolded triple-differential multijet cross section with respect to the track multiplicity for all phase space bins, including the ratio between *all* simulated spectra at truth level and data, as well as the relative uncertainty in each



(a) $p_{\mathrm{T}} \in [250, 350)$ *(b)* $p_{\mathrm{T}} \in [1100, 4500)$

Plot 11.38: Individual ratios of the single-differential multijet cross section in central and forward region with respect to the track multiplicity for different transverse momentum (in GeV) regimes.

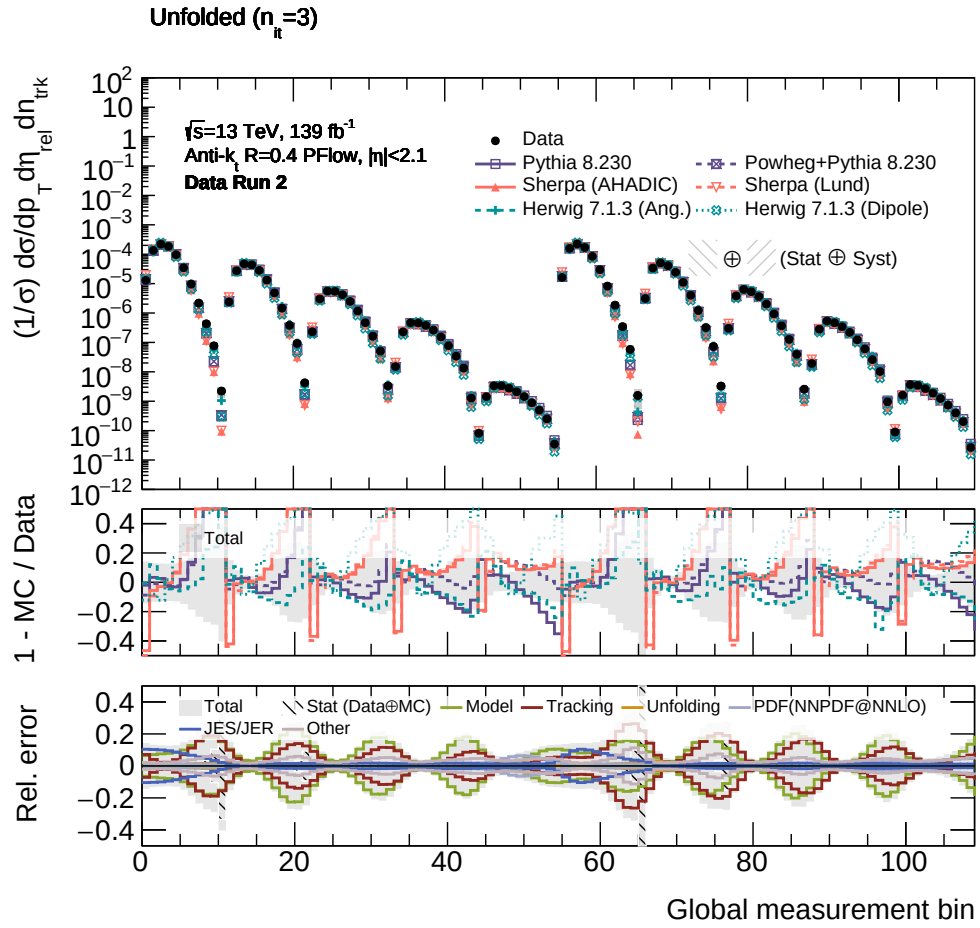
bin broken down to its individual components and the total uncertainty on the measurement.

The lowest sub-figure in Figure 11.39 shows all sources of systematic and statistical uncertainties (including the total uncertainty on the cross section measurement) for all phase space bins; this representation provides a nice overview of the degree of precision archived in this measurement (already shown in Figure 11.37). In summary, it can be concluded that the degree of uncertainty of the measurement is between 10 % and 20 %, which is expected for the measurement of a jet substructure observable.

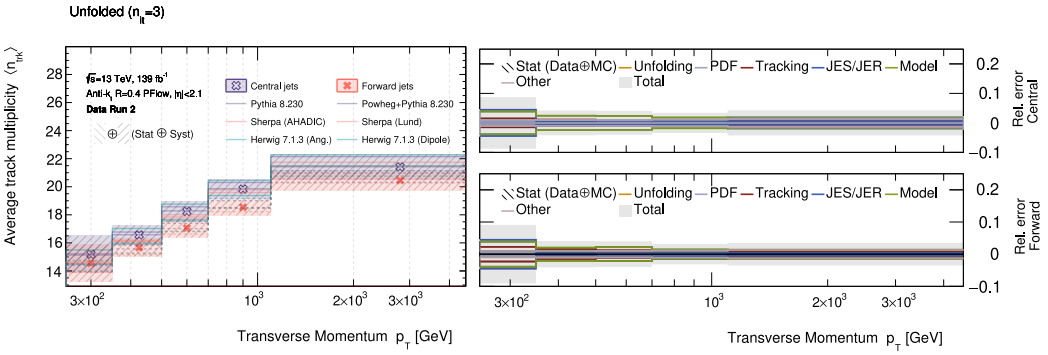
The last plot to consider in the context of the unfolded result for the track multiplicity is the *average* number of tracks in the jet in central and forward region, including *all* sources of uncertainties as a function of the jet's transverse momentum p_{T}. The result is shown in Figure 11.40. The average number of tracks is calculated from the single-differential cross section in central and forward region for each p_{T} bin according to

$$\langle n_{\mathrm{trk}} \rangle_{\mathrm{central/forward}} = \sum_k k \frac{1}{\sigma} \frac{\mathrm{d}\sigma}{\mathrm{d}n_{\mathrm{trk}}}\Big|_{\mathrm{central/forward}}, \tag{11.23}$$

whereby the cross section in Equation 11.23 is given by, e.g., the distributions shown in Figure 11.23. As expected, the average track multiplicity increases for increasing transverse momentum of the jets. Furthermore, the average number of tracks is larger in case of central jets because the central region is enriched by gluon-initialized jets. The difference between the mean value in central and forward region increases for the first part with growing p_{T}, but starts to decrease from $p_{\mathrm{T}} > 1\,\mathrm{TeV}$ onward, which is explained by the fact that the two mixtures approach each other for $p_{\mathrm{T}} > 1\,\mathrm{TeV}$ in terms of quark/gluon jets. This curve progression is (for the most part) determined by the fraction of quark- and gluon-initialized jets in central and forward region as well as the resolution of the reconstructed jet substructure observable. The exact fraction of quark and gluon jets is unknown at this point. However, the simulation-based fraction of gluon jets in Figure 9.1 (which should not be too far of reality) already suggests that the fraction of quark and gluon jets approach each other with increasing p_{T} of the jets. This implies that the two samples of jets in central and forward region provide less information about one specific jet flavor.



Plot 11.39: Full unfolded measurement of the triple-differential multijet cross section as a function of the track multiplicity at particle level including all sources of systematic and statistical uncertainties.

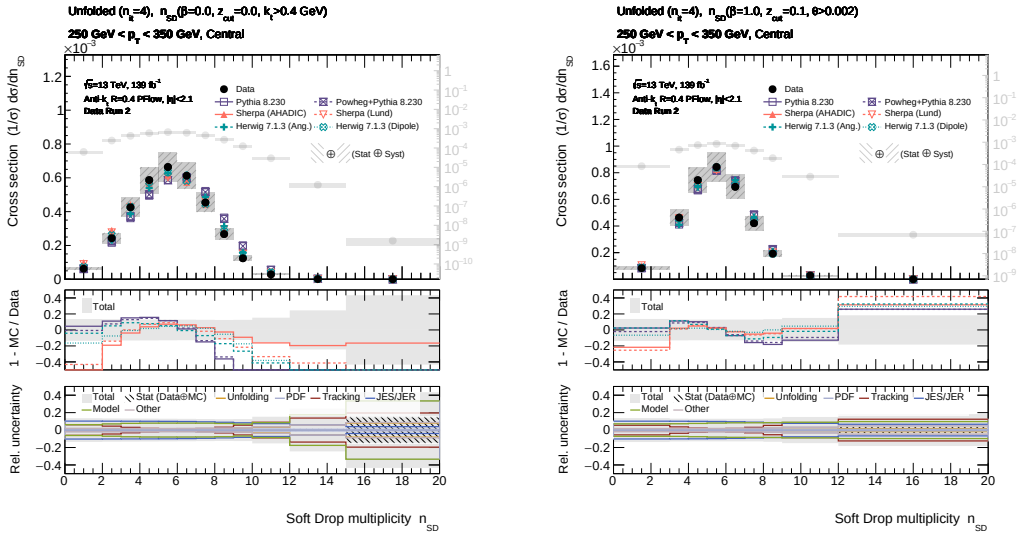


Plot 11.40: Average number of tracks as a function of the jet's transverse momentum

11.6.2 SD Multiplicity n_{SD}[20]

The next jet substructure variable in the family of counting-based observables is the SD multiplicity n_{SD} in its two different configurations introduced in the previous sections. The corresponding detector-level distributions and the unfolding-relevant information was already provided in Section 11.4.2 and Section 11.5.3, respectively.

Compared to the previous section that showed the unfolded results for the track multiplicity,



(a) $n_{\mathrm{SD}}(k_t \geq 400\mathrm{MeV})$ *(b)* $n_{\mathrm{SD}}(z_{\mathrm{cut}} = 0.1, \beta = 1, \theta_{\mathrm{cut}} = 0.002)$

Plot 11.41: Individual single-differential multijet cross sections in bins of the SD multiplicity $n_{\mathrm{SD}}(k_t \geq 400\mathrm{MeV})$ (*left*) and $n_{\mathrm{SD}}(z_{\mathrm{cut}} = 0.1, \beta = 1, \theta_{\mathrm{cut}} = 0.002)$ (*right*) in *central* region for jets with $250\,\mathrm{GeV} \leq p_{\mathrm{T}} < 350\,\mathrm{GeV}$.

the discussion of the outcomes in case of the SD multiplicity will be highly condensed because there is a significant overlap regarding the interpretation of the acquired insights for all counting-based observables. Therefore, most results will be stated without the need for any further detailed explanations, referring to the corresponding results in case of the track multiplicity in Section 11.6.1.

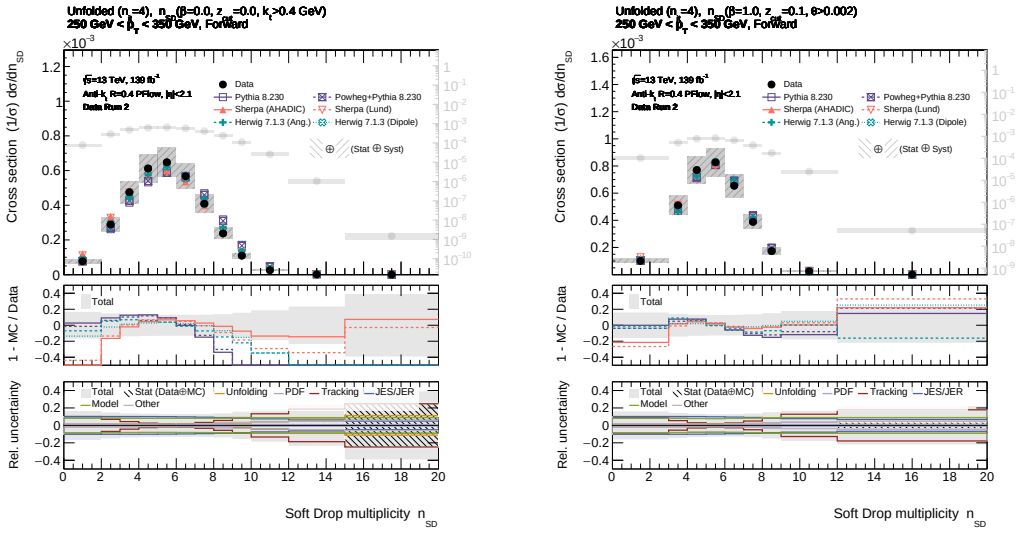
Mimicking the structure of the previous section, the first two pairs of distributions show the unfolded cross section as a function of the SD multiplicity $n_{\mathrm{SD}}(k_t > 400\,\mathrm{MeV})$ (*left*) and $n_{\mathrm{SD}}(z_{\mathrm{cut}} = 0.1, \beta = 1, \theta_{\mathrm{cut}} = 0.002)$ (*right*) for jets with a transverse momentum of $250\,\mathrm{GeV} \leq p_{\mathrm{T}} < 350\,\mathrm{GeV}$ for *central* jets in Figure 11.44a and *forward* jets in Figure 11.42. The following pair of figures in Figure 11.43 and Figure 11.44 show the corresponding information for jets with a transverse momnetum of $1100\,\mathrm{GeV} \leq p_{\mathrm{T}} < 4500\,\mathrm{GeV}$.

Figure 11.43 shows the same information but for jets falling in the interval $1100\,\mathrm{GeV} \leq p_{\mathrm{T}} < 4500\,\mathrm{GeV}$.

Like it was the case for the track multiplicity, the spectra corresponding to central and forward jets are very similar with a smaller number of emissions on average and slightly increased uncertainty due to the larger material budget (mostly driven by increased tracking-related uncertainties) in case of the latter one. The uncertainty is mostly dominated by the parton-shower, JES and track-related uncertainty.

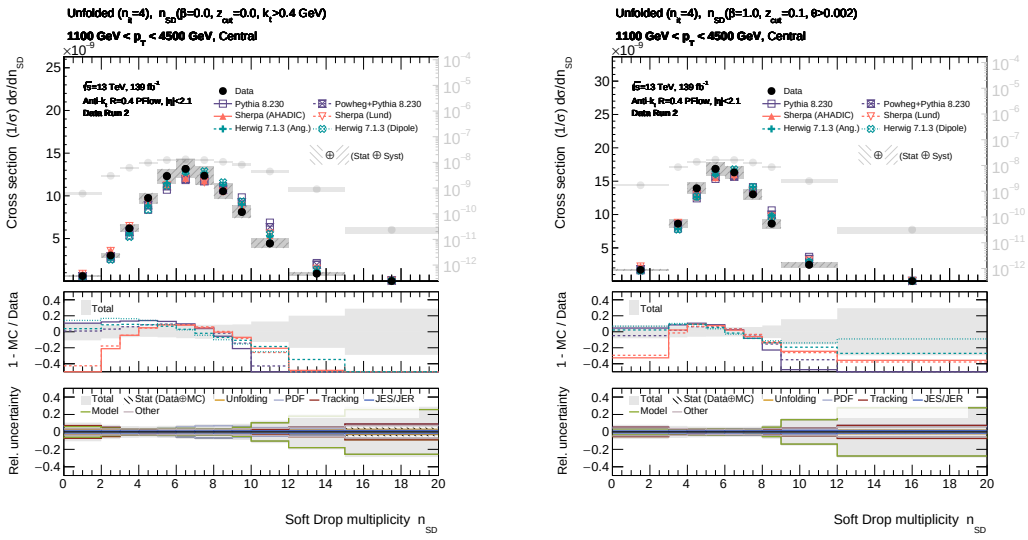
Figure 11.45 is supposed to highlight the calculation of the modeling-related uncertainty. As explained in Chapter 10.3, the modeling-related component of the uncertainty is composed of variations due to differences in the response matrix as well as the efficiency and fake

[20] The corresponding numerical values of the relative triple-differential multijet cross section with respect to this observable are provided in Table B.3 and Table B.4 in the Appendix of this report.



(a) $n_{\mathrm{SD}}(k_t \geq 400\mathrm{MeV})$ *(b)* $n_{\mathrm{SD}}(z_{\mathrm{cut}} = 0.1, \beta = 1, \theta_{\mathrm{cut}} = 0.002)$

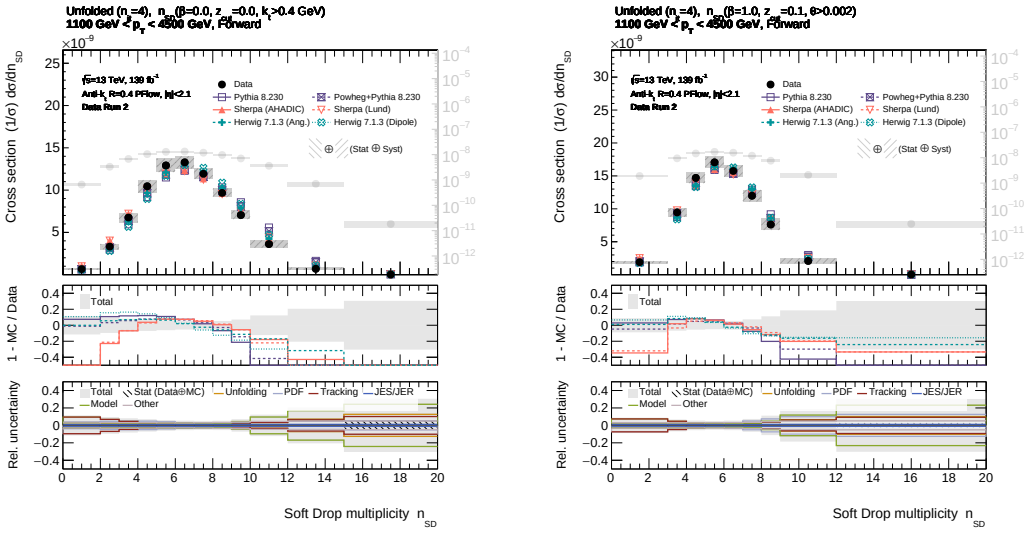
Plot 11.42: Individual single-differential multijet cross sections in bins of the SD multiplicity $n_{\mathrm{SD}}(k_t \geq 400\mathrm{MeV})$ (*left*) and $n_{\mathrm{SD}}(z_{\mathrm{cut}} = 0.1, \beta = 1, \theta_{\mathrm{cut}} = 0.002)$ (*right*) in *forward* region for jets with $250\,\mathrm{GeV} \leq p_{\mathrm{T}} < 350\,\mathrm{GeV}$.



(a) $n_{\mathrm{SD}}(k_t \geq 400\mathrm{MeV})$ *(b)* $n_{\mathrm{SD}}(z_{\mathrm{cut}} = 0.1, \beta = 1, \theta_{\mathrm{cut}} = 0.002)$

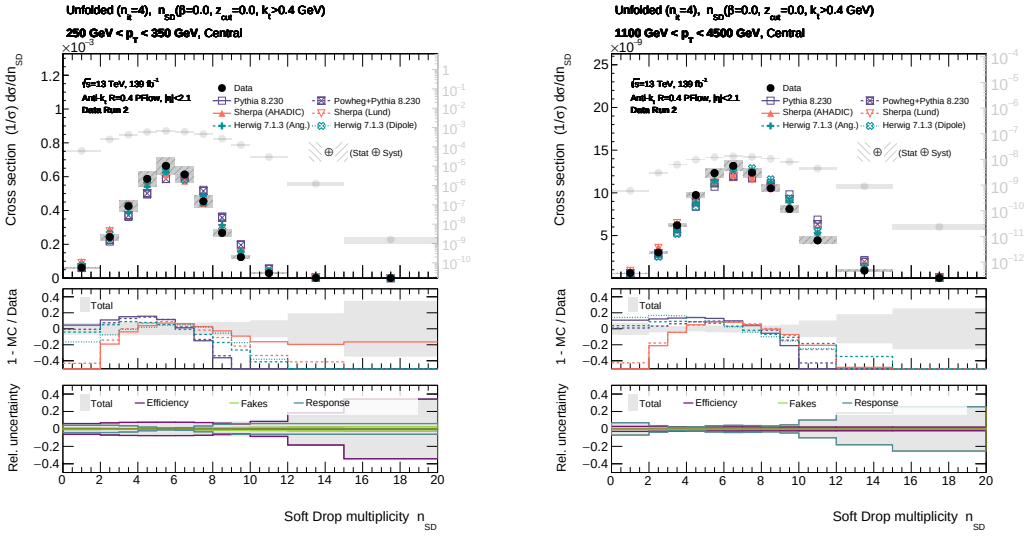
Plot 11.43: Individual single-differential multijet cross sections in bins of the SD multiplicity $n_{\mathrm{SD}}(k_t \geq 400\mathrm{MeV})$ (*left*) and $n_{\mathrm{SD}}(z_{\mathrm{cut}} = 0.1, \beta = 1, \theta_{\mathrm{cut}} = 0.002)$ (*right*) in *central* region for jets with $1100\,\mathrm{GeV} \leq p_{\mathrm{T}} < 4500\,\mathrm{GeV}$.

factors of two competing models concerning the description of the same physical process. In case of the PS-related uncertainty (which is shown in Figure 11.45 for $n_{\mathrm{SD}}(k_t \geq 400\mathrm{MeV})$ and two p_{T} bins), both HERWIG samples are compared. Figure 11.45 shows the variation of the PS-related uncertainty due to the response matrix as well as the efficiency and fake factors. As can be seen, the dominant contribution to the final PS-related uncertainty is due to variations in the response matrix between the two parton shower models, which is expected. Furthermore, the contribution due to the response matrix increases for larger multiplicities



(a) $n_{\mathrm{SD}}(k_t \geq 400\mathrm{MeV})$ *(b)* $n_{\mathrm{SD}}(z_{\mathrm{cut}} = 0.1, \beta = 1, \theta_{\mathrm{cut}} = 0.002)$

Plot 11.44: Individual single-differential multijet cross sections in bins of the SD multiplicity $n_{\mathrm{SD}}(k_t \geq 400\mathrm{MeV})$ (*left*) and $n_{\mathrm{SD}}(z_{\mathrm{cut}} = 0.1, \beta = 1, \theta_{\mathrm{cut}} = 0.002)$ (*right*) in *forward* region for jets with $1100\,\mathrm{GeV} \leq p_{\mathrm{T}} < 4500\,\mathrm{GeV}$.



(a) $250\,\mathrm{GeV} \leq p_{\mathrm{T}} < 350\,\mathrm{GeV}$ *(b)* $1100\,\mathrm{GeV} \leq p_{\mathrm{T}} < 4500\,\mathrm{GeV}$

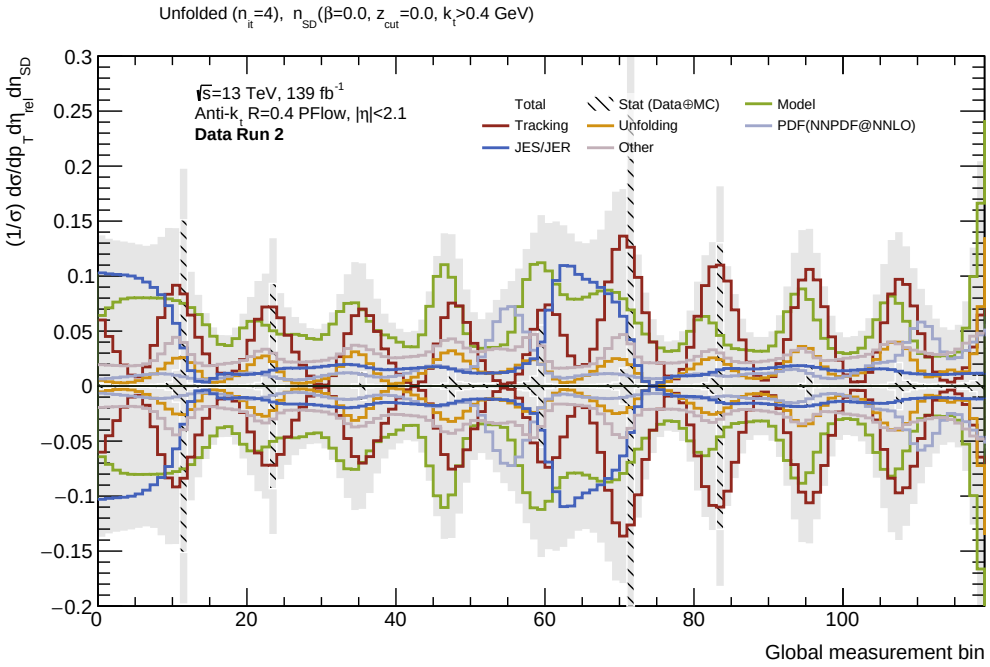
Plot 11.45: Different contributions to the PS-related uncertainty (efficiency and fake factors, response matrix) in bins of the SD multiplicity $n_{\mathrm{SD}}(k_t \geq 400\mathrm{MeV})$ for central jets with $250\,\mathrm{GeV} \leq p_{\mathrm{T}} < 350\,\mathrm{GeV}$ (*left*) and $1100\,\mathrm{GeV} \leq p_{\mathrm{T}} < 4500\,\mathrm{GeV}$ (*right*).

(this statement holds true for all counting-based observables in this analysis) because jets with large multiplicities usually contain a significant amount of soft(er) contributions, which are more sensitive to modeling-related effects. The final model-related uncertainty is given by an envelope of contributions from the response matrix and the efficiency and fake factors (see Chapter 10.2).

All sources of systematic and statistical uncertainties (including the total uncertainty) of the triple-differential cross section with respect to n_{SD} in each phase space bin are shown in

Figure 11.46 for $n_{\mathrm{SD}}(k_t \geq 400\mathrm{MeV})$ (the results for $n_{\mathrm{SD}}(z_{\mathrm{cut}} = 0.1, \beta = 1, \theta_{\mathrm{cut}} = 0.002)$ are very similar). The uncertainty is very close to the corresponding one of the track multiplicity shown in Figure 11.37, which is expected as both observables are highly correlated.

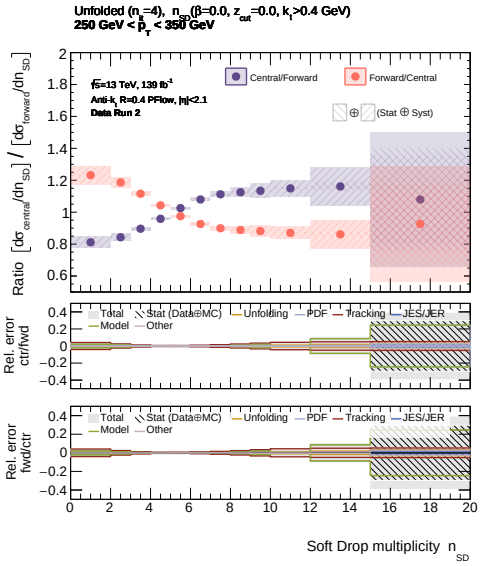
Due to its importance in the context of the extraction of the underlying quark/gluon



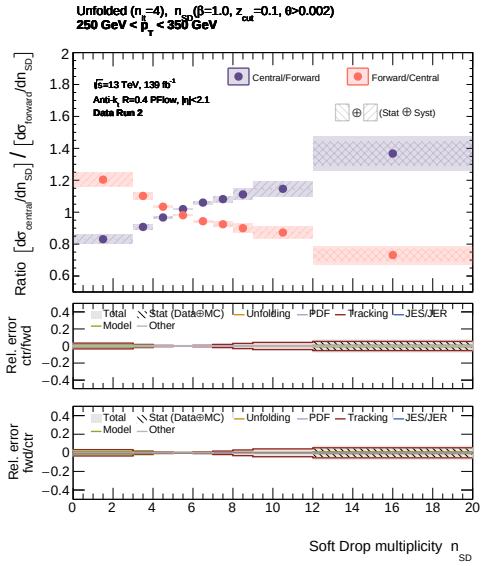
Plot 11.46: All sources of systematic and statistical uncertainties of the triple-differential multijet cross section measurement as a function of the SD multiplicity $n_{\mathrm{SD}}(k_t \geq 400\mathrm{MeV})$.

distributions and fractions in the chapter after next, the ratio between the unfolded triple-differential multijet cross section in *central* and *forward* region is shown by means of two examples in Figure 11.47 and Figure 11.48 for two bins in transverse momentum and both configurations of the SD multiplicity. In general, the cross section ratios show a similar distribution like the ones for the track multiplicity, i.e., an approximately *monotonically* increasing curve progression with the number of "valid" SD emissions.

Last but not least, the average number of emissions is derived once again from the triple-differential cross section; the result is shown in Figure 11.49 for the average number of SD emission in the jet for both configurations of the SD multiplicity.

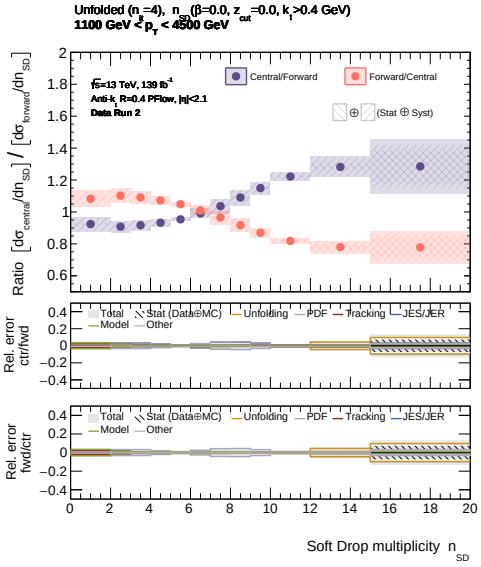


(a) $n_{\mathrm{SD}}(k_t \geq 400\mathrm{MeV})$

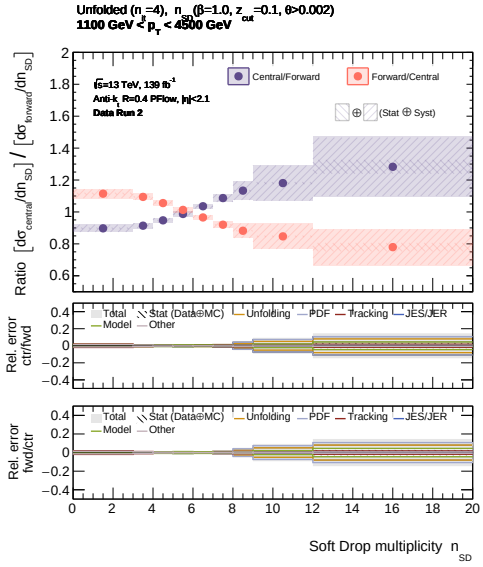


(b) $n_{\mathrm{SD}}(z_{\mathrm{cut}} = 0.1, \beta = 1, \theta_{\mathrm{cut}} = 0.002)$

Plot 11.47: Ratio between the unfolded single-differential distribution in central and forward region for $n_{\mathrm{SD}}(k_t \geq 400\mathrm{MeV})$ (*left*) and $n_{\mathrm{SD}}(z_{\mathrm{cut}} = 0.1, \beta = 1, \theta_{\mathrm{cut}} = 0.002)$ (*right*) for jets with $250\,\mathrm{GeV} \leq p_{\mathrm{T}} < 350\,\mathrm{GeV}$.

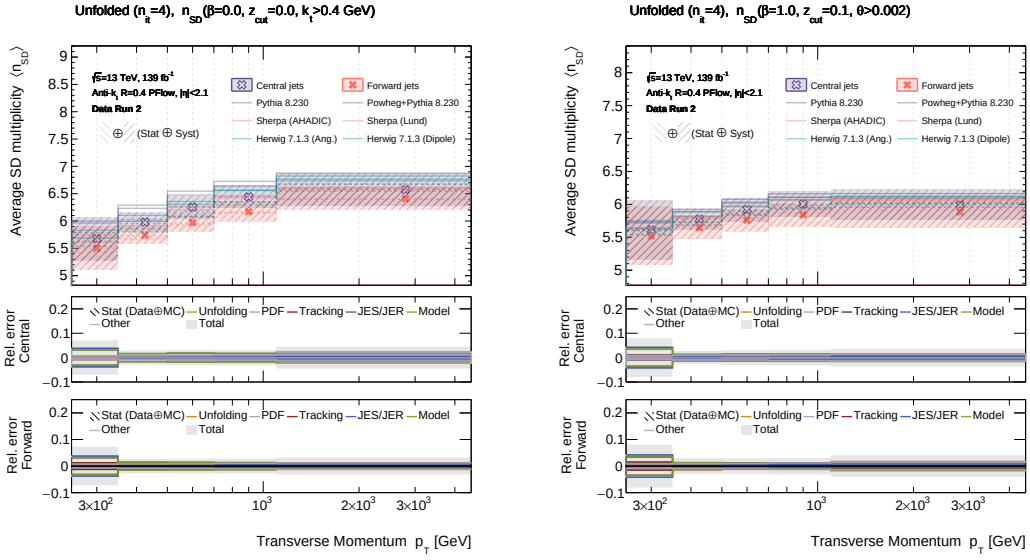


(a) $n_{\mathrm{SD}}(k_t \geq 400\mathrm{MeV})$



(b) $n_{\mathrm{SD}}(z_{\mathrm{cut}} = 0.1, \beta = 1, \theta_{\mathrm{cut}} = 0.002)$

Plot 11.48: Ratio between the unfolded single-differential distribution in central and forward region for $n_{\mathrm{SD}}(k_t \geq 400\mathrm{MeV})$ (*left*) and $n_{\mathrm{SD}}(z_{\mathrm{cut}} = 0.1, \beta = 1, \theta_{\mathrm{cut}} = 0.002)$ (*right*) for jets with $1100\,\mathrm{GeV} \leq p_{\mathrm{T}} < 4500\,\mathrm{GeV}$.



(a) $n_{\mathrm{SD}}(k_t \geq 400\mathrm{MeV})$ *(b)* $n_{\mathrm{SD}}(z_{\mathrm{cut}} = 0.1, \beta = 1, \theta_{\mathrm{cut}} = 0.002)$

Plot 11.49: Average number of emissions as a function of the jet's transverse momentum for $n_{\mathrm{SD}}(k_t \geq 400\mathrm{MeV})$ (*left*) and $n_{\mathrm{SD}}(z_{\mathrm{cut}} = 0.1, \beta = 1, \theta_{\mathrm{cut}} = 0.002)$ (*right*).

12. Measurement of the Triple-Differential Multijet Cross Section $\frac{1}{\sigma}\frac{\mathrm{d}^3\sigma}{\mathrm{d}p_\mathrm{T}\,\mathrm{d}\eta_\mathrm{rel}\,\mathrm{d}z_\mathrm{NN}}$

The previous chapter presented the measurement of the triple-differential multijet cross section as a function of the transverse momentum p_T, the relative orientation between the two leading-p_T jets in the event η_rel, as well as a *counting-based* jet substructure observable n_count, which provides sufficient sensitivity to quark- and gluon-initialized showers to result in mutual irreducible (underlying) probability distributions. In the context of this measurement, two groups of counting-based observables have been studied. The first observable is given by the number of tracks (with $p_\mathrm{T} > 500\,\mathrm{MeV}$) that have been matched to the jet, which is a proxy of the multiplicity of *charged* constituents in the jet. The second class of counting-based observables is given by the number of (valid) Soft Drop emissions in the jet that is obtained by iteratively declustering the angular-ordered history of the jets and counting all emissions that pass the criterion of the respective configuration of the ISD algorithm. All measurements of the differential cross section have been unfolded to particle level and systematic as well as statistical uncertainties have been derived.

The motivation for measuring the relative rate of multijet events as a function of a counting-based jet substructure observable is given by the fact they are "simple" in terms of interpretability, yet provide a decent sensitivity to the quark- or gluon-like nature of the hadronic shower that is not limited by Casimir scaling (see Chapter 7.3). The latter point is important for the observables to result in probability distributions that are mutual irreducible, which is a necessary condition for the extraction of the underlying quark/gluon fractions by means of the "jet topic" approach explained in Chapter 7.

The triple-differential multijet cross section was not only measured as a function of a counting-based substructure variable, but also with respect to a "score" predicted by a neural network based on the "Deep Sets" architecture (see Chapter 6.3). The motivation for using the "Deep Sets" paradigm is given by the fact that it allows to implement functions over sets that are inherently invariant under the permutation of the elements (proper sets) and therefore naturally account for symmetries generally associated with jets. In contrast to the counting-based observables, the inference process in case of a neural network is much less transparent; however, the resulting score usually is significantly more powerful in terms of quark-gluon discrimination. The latter point will allow to extract more quark- and gluon-related information from the two mixtures of jets (central and forward samples), which should help withdrawing the underlying probability distributions and quark/gluon fractions. Furthermore, the NN-related results are expected to be less sensitive to tracking-related uncertainties.

This chapter presents the results for the above-mentioned measurement of the multijet triple-differential cross section as a function of the NN-score as predicted by *two* neural networks that are based on the "Deep Set" architecture. The organization of this chapter mostly mimic the structure of the previous one in case of the counting-based jet substructure observables. In the first section of this chapter (12.1), the actual NN-based models are introduced. The performance in terms of training stability as well as classification between quarks and gluons highly depends on the format of the data that is provided to the model. The second section (12.3) therefore introduces the preprocessing strategy of the data. The next section (12.2) briefly introduces the network architecture of both baseline models. Once the configuration of both models has been described, the next section (12.4) presents some results related to

the training of the neural networks. The objective of this section is to provide some intuition for the NN-based models and to convince the reader that the training process as well as the inference of the NN-score is well under control and stable. After the networks have been trained, the triple-differential multijet cross section is measured as a function of the score predicted by both models (12.5). Like it was done in case of the counting-based observables, the subsequent section (12.6) deals with some rather technical aspects of the unfolding step. In the final section of this chapter (12.7), the unfolded results of the cross section including all sources of statistical and systematic uncertainties are presented.

12.1 Neural-Network-Based Observables utilizing Energy Flow Networks

The application of machine-learning methods in high-energy particle physics goes back several decades. In fact, both disciplines have greatly influenced each other over the course of time. Especially the utilization of (mostly "deep") neural networks (see Chapter 6) as binary or multiclass classifiers for, e.g., flavor tagging purposes have become a standard method in many analyses for a long time already. But, there are plenty of other use cases like, for instance, signal-background discrimination, *un*binned unfolding, anomaly tagging based on variational auto encoders, generative adversarial networks used in simulation, graph neural networks for track reconstruction ... – the list could be continued forever with no end in sight.

One key to success when working with machine-learning-based models is to chose the appropriate representation of the data along with the correct model or architecture, i.e., a network architecture particularly suited to analyze *this* representation of the data. In this context, "representation" simply refers to the set of transformations that are applied to the *raw* data to obtain a different format suited for the intended task. In the scope of jet-related physics, the raw data is usually given by a cloud of "points"(=particles) distributed over an η–ϕ-grid, each one being associated with a scalar quantity that is commonly identified with the energy or the transverse momentum of the respective constituent. Formally, a jet $J \in \mathbb{R}^{(3,1)}$ can be expressed as a *set* of particles $\{j\}_{j\in J}$, with $N = |J| \in \mathbb{N}(:= n_{\text{const}})$ denoting the *cardinality* of the set, henceforth referred to as *constituent multiplicity*. This is a highly *non*-trivial type of data as the number of constituents inside a jet, N, is subject to statistical fluctuations on a jet-to-jet basis (*variable-length input*); furthermore, there is no universal, physically well motivated ordering scheme regarding the hierarchy of the particles inside of a jet (*unordered*). The latter point may come as a surprise as in many of today's applications, the constituents inside the jets are ordered based on the k_t or the transverse momentum p_{T} (decreasing order) or the angular separation with respect to the center of the jet. Given a particle jet, both orderings (as others) are arbitrary[1] due to quantum mechanical indestiunushibility since all arrangements result in the very same (global) observable and thus being equally valid.

Regarding the different representations of the data appropriate for jets, they can – loosely speaking – be subdivided into three categories: *Firstly*, the particle content of the jet is mapped to a *single* number (commonly referred to as a jet substructure variable) by means of a function τ with $\tau : \wp\left(\mathbb{R}^{(3,1)}\right) \to \mathbb{R}$, whereby $\wp\left(\mathbb{R}^{(3,1)}\right)$ denotes the set of *all* subsets (the power set) of the four-dimensional Minkowsky space $\mathbb{R}^{(3,1)}$. As can be shown, given the available degrees of freedom, a set of those variables $\{\tau_i\}_{i=1}^{M}$ of *constant length* $M \in \mathbb{N}$ span a basis and can be used as an input to train a neural network. This method is widely used in many heavy-flavor tagging applications as for those topologies the substructure variables τ_i already exhibit considerable discrimination power (an example would be the jet mass window in case of top tagging, Energy Correlation Functions [221], N-subjettiness (ratios) [222, 223]). *Secondly*, the representation of the jet in terms of an "image-like" description has gained increasing popularity over the recent years. This representation of the activity inside of the jet essentially emerge simultaneously with the first appearance of convolutional neural networks, which is an architecture particularly well suited for image-based data with translational invariance. Portraying the jet as an image seems

[1] Being arbitrary does not imply that a particular ordering scheme may not be appropriate in one application or another (see, e.g., SD multiplicity).

natural due to the close correspondence with the distributed spray of energy in the calorimeter system of the detector, which is usually associated with a jet. Furthermore, the dimensionality of the data is fixed by the discretization or pixelation of the image (such as the calorimeter granularity). The image concept, however, is not restricted to the intuitive representation of the activity directly measured in the calorimeter of the detector, but can be used for other observables as well. A visually appealing example of an alternative application is the Lund Jet Plane [224], which was used for several tagging applications already. Image-based taggers have proven to work well in the context of jet tagging; however, they are limited by the granularity of the pixel grid. One major disadvantage of using image-based inputs in the scope of high-energy particle physics is due to the circumstance that on an event-on-event or jet-on-jet basis, only a small fraction of the relevant detector components/pixels are active, i.e., with an actual pixel activation above zero. This generally results in the "pixelized" representation of the data being very sparse, with only a tiny proportion of the cells/voxels exhibit a non-vanishing energy deposition. It has been known for a long time already that neural networks trained on sparse data show a comparatively "poor" performance. One reason is that the effective size of the data set reduces significantly if a large fraction of the data's support is sparse. If – additionally – the cell's activity spans several orders of magnitude (like it is the case for the energy and the transverse momentum), the network might not be able to correctly learn correlations between the different pixels (especially in the tails of the distributions or only properly learns correlations between the first highest-p_{T} constituents). In general, this drawback of discretized data provides a motivating argument for the last category of data representations. *Thirdly*, the entire sequence of numbers associated with each constituent inside of the jet is used as input (like it was the case for τ). In the context of machine learning, this representation requires a dedicated architecture that is able to process variable-length data. As mentioned in Chapter 6.1, RNNs have been invented exactly for this use case. However, it was shown in several studies (especially to be mentioned in this context is the excellent work by Vinyals *et al.* [225]) that the output of the function f learned by the RNN *significantly* depends on the (partially) arbitrarily imposed ordering of the elements, i.e., $f(x_1, ..., x_M) \neq f(x_{\pi(1)}, ..., x_{\pi(M)})$, whereby π denotes a permutation. Hence, invariance with respect to the permutation of the elements is *not* respected – which is the main motivation for neural networks based on the Deep Set architecture.

In Chapter 6.3, the Deep Set neural networks have been conceptually introduced as an machine learning architecture that allows to learn permutation invariant functions f that are defined over variable-length (i.e., proper) sets. The main idea behind Deep Sets is the insight that a permutation invariant function $f : \mathfrak{X} \to \mathcal{Y}$ defined over a set $\mathfrak{X}$ can be constructed using *two* continuous functions $\rho : \mathbb{R}^\ell \to \mathcal{Y}$ and $\phi : \mathfrak{X} \to \mathbb{R}^\ell$ by the following relation $f(X) = \rho\left(\sum_{x \in X} \phi(x)\right)$, whereby in practical applications the functions ρ and ϕ are approximated by neural networks and $X \subset 2^{\mathfrak{X}}$ being a subset of the function's domain. Adopting the notation from Patrick T. Komiske, Eric M. Metodiev, and Jesse Thaler in Ref. [27], the *Deep Sets Theorem* can be summarized as follows:

Theorem (27, Thm. "Deep Sets Theorem", p. 6/7)**.** *Let $\mathfrak{X} \subset \mathbb{R}^d$ be compact, $X \subset 2^{\mathfrak{X}}$ be the space of sets with bounded cardinality of elements in $\mathfrak{X}$, and $Y \subset \mathbb{R}$ be a bounded interval. Consider a continuous function $f : X \to Y$ that is invariant under permutations of its inputs $f(x_1, ..., x_M) = f(x_{\pi(1)}, ..., x_{\pi(M)})$ for all $x_i \in \mathfrak{X}$ and $\pi \in S_M$. Then there exists a sufficiently large integer ℓ and continuous functions $\Phi : \mathfrak{X} \to \mathbb{R}^\ell$, $F : \mathbb{R}^\ell \to Y$ such that the following holds to an arbitrarily good approximation*

$$f(\{x_1, ..., x_M\}) = F\left(\sum_{i=1}^{M} \Phi(x_i)\right) [,]$$

whereby the validity of the theorem above relies on the proof provided in Ref. [3], appendix A, p. 12.

Translating the Theorem 12.1 to jet-physics is straightforward as the natural representation of jets in terms of sets of particles already corresponds to the functional input required by f. (Once again guided by the notation from Ref. [27]) given a jet with M constituents each one associated with some feature vector $p_i \in \mathbb{R}^d$ (like, for instance, the four-momentum, the charge, some ID from particle identification etc.); according to the Deep Sets Theorem 12.1 as well as

the UA Theorem 6.1, each continuous observable given by

$$z_{\text{PFN}}(\{p_1, ..., p_M\}) = F\left(\sum_{i=1}^{M} \Phi(p_i)\right) \in [0, 1], \tag{12.1}$$

can be approximated arbitrarily well. Within the context of this thesis, the observable z_{PFN} (according to Equation 12.1) is henceforth (following the nomenclature of the original publication in Ref. [226] by Patrick T. Komiske, Eric M. Metodiev, Jesse Thaler) referred to as *Particle Flow Network (PFN) score*.

In Equation 12.1, the function $\Phi : \mathbb{R}^d \to \mathbb{R}^\ell$ denotes a *map* that is applied *per-particle* (i.e., for each constituent $\{p_i\}_{i=1}^M$ individually) that maps each constituent of the jet onto an ℓ-dimensional *latent space*. The linear sum of the individual latent space embedding in Equation 12.1 results in a latent space representation of the *entire* jet. Finally, the continuous function $F : \mathbb{R}^\ell \to Y(= \mathbb{R})$ maps the total representation to the observable.

Interestingly, the family of observables approximated by the decomposition in Equation 12.1 can be restricted to IRC-safe ones (see Chapter 1.3.2) by the following simple modification

$$z_{\text{EFN}}(\{p_1, ..., p_M\}) = F\left(\sum_{i=1}^{M} z_i \Phi(\hat{p}_i)\right) \in [0, 1], \tag{12.2}$$

whereby z_i corresponds to some energy-related (constituent's energy or p_{T}) factor and $\hat{p}_i$ gives the angular components (η, ϕ) of the ith particle in the jet. The factorized version in Equation 12.2 is henceforth referred (again following the notation in Ref. [226]) to as *Energy Flow Network (EFN) score*. The IRC-safety of Expression 12.2 is directly apparent from the fact the relation is linear in energy and continuous in space. Formally, in order for the observable z_{EFN} to be IRC-safe, the following two conditions must hold

$$F\left(\sum_{i=1}^{M} z_i \Phi(\hat{p}_i)\right) = \lim_{\substack{\epsilon \to 0 \\ \lambda \to 0}} F\left(\underbrace{\Phi(\epsilon, \hat{p}_0)}_{\text{IR-safe}} + \overbrace{\lambda z_1 \Phi(\hat{p}_1)}^{\text{C-safe}} + (1-\lambda) z_1 \Phi(\hat{p}_1) + \sum_{i=2}^{M} z_i \Phi(\hat{p}_i)\right),$$

both of which are trivially fulfilled given the decomposition according to Equation 12.2.

To shed some light and to gain some intuition on the energy and particle flow network, it is instructive to take a look at some special cases of the two functions F and Φ in the case of some well-known observables. For instance, given $F(x^\mu) = \sqrt{x^\mu x_\mu}$ and $\Phi = p^\mu$, results in the jet mass $m^2 = p^\mu p_\mu$. Similarly, if $F(x) = x$ and $\Phi = 1$, the observable represented by the decomposition is the constituent multiplicity $M = \sum_{i=1}^{M} 1$ and hence returns the limit in the sum in Equation 12.1 and 12.2 (for more examples in this regard, the reader is refereed to Table 1 in the original publication in Ref. [27]).

12.2 Data Preprocessing and Training Data[2,3]

This section introduces the separate preprocessing transformations that are applied to the input data used to train the machine-learning models presented in this report. The objective of the individual preprocessing steps is to exploit as *much* symmetry information encoded in the input data as possible and to prepare the data accordingly in order to avoid the model focusing on learning those symmetries, which do not contain additional information that can be used in the classification task. This procedure makes the learning task of the neural network much more efficient since it does not have to learn redundant symmetries encoded in the data, but may focus on the main features of the data. Furthermore, taking symmetries into account results in faster convergence and allows for a smaller data set to be used. It is also beneficial

[3] Essential parts of the succeeding text passages have been adopted (almost) literally from my Master's thesis "*Towards a Data-Driven Simulation of QCD Radiation with Generative Models utilizing Machine Learning Methods*", 2019, Universität Heidelberg, Christof Sauer [49].

in terms of interpreting the prediction of a model and its inputs. Hence, it is worth to dwell on this subject for a moment as it establishes a basis for the entire NN-based measurement of the differential cross section.

The input to the aforementioned EFN and PFN are the constituents (or some object derived from it) associated to a *single* jet; in this analysis, this jet corresponds to either the leading or subleading jet in the event. The reconstructed jet (i.e., the global phenomenological object) is associated with a collection of constituent particles P_j and a four-vector $P_J = \sum_{j\in J} P_j$ given by

$$P_J = E_J \left(1, \frac{\cos\phi_J}{\cosh\eta_J}, \frac{\sin\phi_J}{\cosh\eta_J}, \frac{\sinh\eta_J}{\cosh\eta_J}\right) \tag{12.3}$$

$$= \sum_{j\in J} E_j \left(1, \frac{\cos\phi_j}{\cosh\eta_j}, \frac{\sin\phi_j}{\cosh\eta_j}, \frac{\sinh\eta_j}{\cosh\eta_j}\right), \tag{12.4}$$

whereby the summation is taken over all constituent particles $\{j\}_{j\in J}$ of the jet J that have been clustered by the respective reconstruction algorithm. The "nature" of the constituents depend on the object that have been used to cluster the jet, e.g., topo clusters, tracks only, PFlow objects (like in this analysis), TCCs, UFOs etc. (see Chapter 4.4.1). In this analysis, the constituents are given by *tracks* associated/matched (see Chapter 9.2.1) to PFlow jets, i.e., only the charged components in Equation 12.3 are considered. The constituents of the jet could, in principle, directly (meaning: *un*-preprocessed) be used to train the neural network; however, they are preprocessed by a consecutive sequence of transformations beforehand due to the reasons mentioned above.

In the next paragraphs, the following preprocessing steps are discussed: Lorentz transformation (boost) of the leading/subleading-p_T jet to the point of origin, performing a PCA, a parity transformation and a scaling of the scalar components of the constituents' four-momentum.

Step 1: Lorentz Boost and Rotation

The reconstructed jet according to Equation 12.3 (as well as its individual constituents in Equation 12.4) is associated with some position in the detector, i.e., in the discretized η–ϕ grid. To prepare the training set for the neural network, it is advisable to "standardize" the data to a certain extent. Therefore, the four-vector of the reconstructed jet P_J according to Equation 12.3 is transformed via a Lorentz transformation $\Lambda = (\Lambda^\alpha_\beta)$ (strictly speaking, a Lorentz boost) followed by a simple rotation R in Euclidean space such that the spatial position of its barycenter$(= (\eta_J, \phi_J))$ is located at $(0, 0)$ in the η–ϕ grid. To this end, the four-momentum of the jet's constituent particles must be transformed accordingly. With this in mind, the transformation is given by

$$P'_J = R\Lambda\, P_J \tag{12.5}$$

$$= \begin{bmatrix} \cosh\eta^\mathrm{b} & 0 & 0 & -\sinh\eta^\mathrm{b} \\ 0 & \cos\phi^\mathrm{rot} & -\sin\phi^\mathrm{rot} & 0 \\ 0 & \sin\phi^\mathrm{rot} & \cos\phi^\mathrm{rot} & 0 \\ -\sinh\eta^\mathrm{b} & 0 & 0 & \cosh\eta^\mathrm{b} \end{bmatrix} \cdot \begin{bmatrix} E_J \\ p_{Jx} \\ p_{Jy} \\ p_{Jz} \end{bmatrix} \tag{12.6}$$

$$\overset{(12.3)}{=} \sum_{j\in J} (R\Lambda\, P_j) = \sum_{j\in J} P'_j, \tag{12.7}$$

whereby $\gamma^\mathrm{b} = \tanh\eta^\mathrm{b}$ and $\gamma^\mathrm{b}\beta^\mathrm{b} = \sinh\eta^\mathrm{b}$ have been used. It remains to determine the boost parameter η^b and the rotation angle ϕ^rot to fully define the transformation above. From the matrix multiplication in Equation 12.7, one gets the following condition on the transformed energy E'_J and the z component p'_{Jz} of the jet after the transformation

$$p_{Jz}\cosh\eta^\mathrm{b} - E_J\sinh\eta^\mathrm{b} = p'_{Jz} \overset{!}{=} 0. \tag{12.8}$$

Equation 12.8 defines the boost parameter $\tanh\eta^\mathrm{b} = \beta^\mathrm{b} = p_{Jz}/E_J = \tanh\eta_J$. Similarly, the defining condition for the rotation angle is given by

$$p_{Jx}\cos\phi^\mathrm{rot} + p_{Jy}\sin\phi^\mathrm{rot} = 0; \tag{12.9}$$

hence, the vector in the transverse plane must be rotated by the following amount $\phi^{\text{rot}} = -\arctan\left(p_{Jy}/p_{Jx}\right)$.

The transformation according to Equation 12.7 (defined by η^{b}, ϕ^{rot}) is applied to *all* constituents that make up the jet. Hence – by construction – the reconstructed object, i.e., the jet will be centered in the η–ϕ plane.

Step 2: PCA-Based Rotation

After the fist preprocessing step, the result is a region $\{(x, y)\mathbb{R}^2 \,|\, x^2 + y^2 < R^2\}$ in the η–ϕ grid close to the detector's center that contains the transformed constituents of the jet. The second preprocessing step serves the purpose of taking advantage of the underlying rotational symmetry of the jet by conducting a *Principal Component Analysis* (PCA) that finds the *principal* components of the data – the jet in this case – on an *jet-on-jet* basis.

The PCA is a technique that is frequently used in data science; its objective is to find the *axis of maximum variance* in a distribution of data points by solving an eigenvalue problem. Often, a PCA is used to reduce the dimensionality of the data by projecting it on the principal axis; here, however, the PCA defines a "special axis" that is used to rotate each jet on an jet-on-jet basis such that the Principal Component (PC) (axis) of this particular particle cloud points at (arbitrarily chosen) 12 o'clock. The PCA then provides the angle $\alpha_{\text{rot}}^{\text{PC}}$ by which the particle cloud must be rotated *counterclockwise*.

The eigenvalue problem that must be solved in context of the PCA is a very simple one. As it has been described above, after the first preprocessing step the jet is given by a set of constituents distributed around the point of origin $(0, 0)$ within a radius R, which corresponds to the radius parameter that was used to reconstruct the anti-k_t jet. Each particle with index $i \in \mathbb{N}$ inside of the jet has an associated position (η_i, ϕ_i) as well as an associated energy and momentum value E_i and p_{T}. The mean position/barycenter in the jet is given by the weighed sum

$$\mu_{\eta(\phi)} = \sum_{i=1}^{N} p_{\text{T}i}\eta_i(\phi_i) / \sum_{i=1}^{N} p_{\text{T}i} i, \tag{12.10}$$

which (after the first transformation step) will be close to zero. The next step is to compute the entries of the correlation matrix Σ_{ij}, which is simply given by $\Sigma_{ij} = \mathbb{E}[p_{\text{T}i}p_{\text{T}j}] - \mu_i\mu_j$ with $i, j \in \{\eta, \phi\}$. To get the fist principal axis, one needs to compute the eigenvalues λ^{PC} and corresponding Eigenvectors x^{PC} of the covariance matrix $\det|\Sigma - \lambda^{\text{PC}}\mathbb{1}^{2\times 2}| = 0$; so, the characteristic polynomial is given by

$$\det\left|\begin{bmatrix}\Sigma_{\eta\eta} - \lambda^{\text{PC}} & \Sigma_{\eta\phi} \\ \Sigma_{\phi\eta} & \Sigma_{\phi\phi} - \lambda^{\text{PC}}\end{bmatrix}\right| = \left(\lambda^{\text{PC}}\right)^2 - 2\lambda^{\text{PC}}\Sigma_{\eta\phi} + \Sigma_{\eta\eta}\Sigma_{\phi\phi} - \Sigma_{\eta\phi}^2 = 0, \tag{12.11}$$

whereby the symmetry of the covariance matrix $\Sigma_{\eta\phi} = \Sigma_{\phi\eta}$ was used. The Polynomial 12.11 can easy be solved for the eigenvalues $\lambda_{1,2}^{\text{PC}} = \Sigma_{\eta\phi} \pm \sqrt{2\Sigma_{\eta\phi}^2 - \Sigma_{\eta\eta}\Sigma_{\phi\phi}}$. On that basis, the kernel of Σ is computed; the elements of the kernel $x^{\text{PC}} \in \ker(\Sigma)$ are then satisfying the following conditions

$$(\Sigma_{\eta\eta}x_1^{\text{PC}} - \lambda^{\text{PC}}) + \Sigma_{\eta\phi}x_2^{\text{PC}} = 0, \tag{12.12}$$

$$(\Sigma_{\phi\phi}x_2^{\text{PC}} - \lambda^{\text{PC}}) + \Sigma_{\phi\eta}x_1^{\text{PC}} = 0. \tag{12.13}$$

Based the solution of the system of linear Equations 12.12 and 12.13, the rotation angle for the jet is given by

$$\alpha_{\text{rot}}^{\text{PC}} = \frac{\pi}{2} - \arctan\left(\frac{x_2^{\text{PC}}}{x_1^{\text{PC}}}\right). \tag{12.14}$$

The determined PC axis $x_{\text{PC}} = (x_1^{\text{PC}}x_2^{\text{PC}})^{\text{T}}$ is only defined up to a sign as can easily be seen by substituting $x_{1,2}^{\text{PC}} \to -x_{1,2}^{\text{PC}}$ in Equation 12.13, which is why one is free to let it point at 12 o'clock. For this choice, the larger amount of energy will always be contained on the *northern* hemisphere of the particle cloud after the transformation has been applied.

Step 3: Party Transformation

After the PCA-based rotation of the jet's constituents has been applied, there is one trivial symmetry left: spatial inversion or parity transformation (this can already be seen in the previous paragraph since the basis given by the PCA is only defined up to a sign). To avoid this underlying symmetry of the data to be learned by the neural network, the jet is transformed such that the *largest fraction* of energy is *always* placed on the *eastern* hemisphere of the rotated η–ϕ grid. The energy fraction on the western $E_{\triangleleft}$ and the eastern $E_{\triangleright}$ half of the jet (with $E^{\text{jet}} = E_{\triangleleft} + E_{\triangleright}$) is given by

$$E_{\triangleleft(\triangleright)} = \sum_{\substack{-R \leq \eta < 0 \\ (0 \leq \eta \leq R)}} \sum_{0 \leq \phi \leq R} E_{\eta\phi}, \tag{12.15}$$

with $E_{\eta\phi}$ denoting the energy of the constituent at position (η, ϕ). Based on Equation 12.15, the "parity operation" is then defined as

$$P = \begin{cases} +1 & \text{if } E_{\triangleright} \geq E_{\triangleleft} \quad \text{(do nothing)} \\ -1 & \text{otherwise} \quad \text{(flip/mirror)} \end{cases}, \tag{12.16}$$

with the parity transformation rules $\eta_i \to P\eta_i$ and $\phi_i \to P\phi_i$.

Step 4: Normalization

All previous steps transform the angular components of the constituents, leaving the energy and the transverse momentum unchanged. The scalar components, however, cover a large range of possible values that span several orders of magnitude. In order to improve training stability and convergence, the range of numbers should be $\sim \mathcal{O}(1)$. This is accomplished by a simple scaling of the scalar components with respect to the total p_{T}[4] of the jet. The transformed scalar entries of the constituents' four-momentum is then given by

$$p_{\text{T}j} \to \frac{p_{\text{T}j}}{\sum_{j \in J} E_j}, \tag{12.17}$$

$$E_j \to \frac{E_j}{\sum_{j \in J} E_j}, \tag{12.18}$$

whereby j runs over *all* constituents in the jet J. The transformations according to Equation 12.17 and Equation 12.18 are both linear and hence structure-preserving, i.e., the distribution over possible values $p_{\text{T}j}, E_j$ remain unchanged – just the support of the distributions are mapped to different values. Therefore, the range of values still span several orders of magnitude after this transformation. This can (but does not necessarily have to) be a problem because the neural network has to generalize over a vast range of values. There are various approaches to deal with this issue.

One ways to deal with this problem is to map the scalar components to its logarithm according to $p_{\text{T}j} \to \log p_{\text{T}j}$ and $E_j \to \log E_j$. This has been done; however, no improvement regarding the performance (neither in convergence nor in classification power) was observed.

Another possibility is to train several neural networks f_i with $i \in \mathbb{N}^{\leq N_{\text{models}}}$ on *exclusive* domains(=regions of phase space) χ_i. This results in a *family* of models $\{f_i\}_{i=1}^{N_{\text{models}}}$ each one of which is applied to arguments falling into the corresponding domain $x \in \chi_i$ exclusively[5]. This method results in a *considerable* additional logistical effort as all models must be trained *and* evaluated separately; but, one is (usually) rewarded with an increased performance.

In the context of this measurement, one could, in principle, train a neural network for each p_{T} bin presented in Table 9.4. This, however, may cause problems if the trained model is applied to detector level objects with migrations between neighboring bins being present. In this case, the neural network might be applied to jets with a very different topology compared

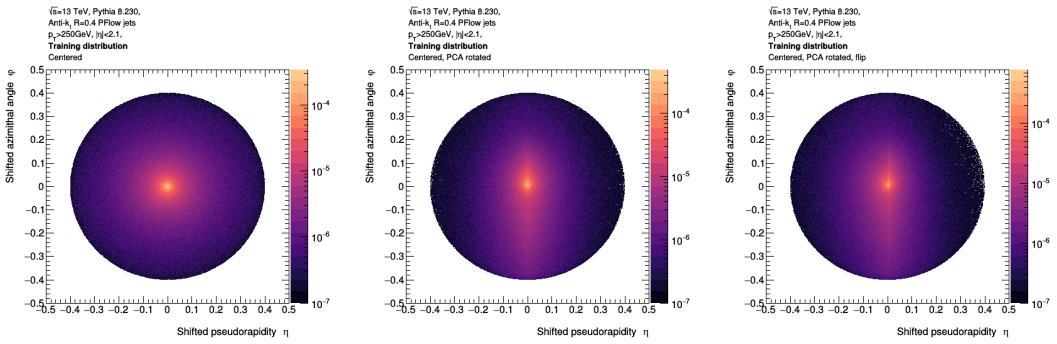
[4] Of course, the total energy of the jet would be an equivalent alternative; but, it case of hadron colliders, the transverse momentum represents a more sensible choice.

[5] It should be mentioned that the entire analysis has been repeated using detector-level jets to train the neural networks – the proof of the pudding is the eating. The result basically remains unchanged.

to the data the model was trained on, which may result in significant deterioration as machine learning models usually are not capable of extrapolating into "unseen" domains. Due to this limitation, two models have been trained for two exclusive regions of phase space: one neural network, $f_{[250,1000]}$, trained (and evaluated) on jets with $p_T \in [250\,\text{GeV}, 1000\,\text{GeV}]$, and a second model, $f_{(1000,4500]}$, for jets falling into the phase space region $p_\text{T} \in [1000\,\text{GeV}, 4500\,\text{GeV}]$. This separation indeed results in an improvement in discrimination power regarding quark- and gluon-initialized jets is in the order of O(1 %) for the higher-p_T bin (in case of the lower-p_T, no improvement was observed)[6]. Considering the comparatively small performance gain given the substantial extra effort and expense, this approach was discarded if favor of just *one neural network* trained on the entire domain $[250\,\text{GeV}, 4500\,\text{GeV}]$.

Full Preprocessing Chain

This paragraph concludes the preprocessing part by applying all aforementioned transformation steps consecutively to the data and paraphrase the final result. In order to visualize the preprocessed jets and its constituents, the jets are represented in form of an "image" of fine granularity, with the axis of abscissa and the axis of ordinates correspond to the pseudorapidity and the azimuthal angle of the (transformed) jets, respectively. The z-axis does not simply count the number of particles given an η–ϕ bin, but instead shows the occurrences of constituents within a certain $[\eta + \delta\eta, \phi + \delta\phi]^2$ interval weighted by the transverse momentum p_T of each individual particle. In more concrete terms, given a jet J and its associated particles/constituents $\{j\}_{j\in J}$, the image is populated by $\{(\eta_j, \phi_j, p_{\text{T}j}/p_{\text{T}J})\}_{\text{j}\in\text{J}}$, whereby $p_{\text{T}j}/p_{\text{T}J}$ corresponds to the p_T of constituent j normalized by the *total* transverse momentum of the jet $p_{\text{T}J}$. With this normalization applied, the final images shown in Figure 12.1 can be regarded as the average "appearance" of a jet given a large collection of jets and its constituents. This helps in visualizing the effect of the individual preprocessing steps, which are barely visible on a jet-on-jet basis. The final result of the preprocessing is shown in Figure 12.1 for the three main transformations



(a) Lorentz boost *(b)* PCA rotation *(c)* Parity flip

Plot 12.1: The three main preprocessing transformations applied to data in the order of their application: a Lorentz boost (*left*), PCA rotation (*right*), parity "flip" (*right*).

applied to the data, i.e., a Lorentz boost that transforms the jets into a fixed region in the detector (12.1a); followed by a rotation of the PC axis such that it points at 12 o'clock (12.1b); and, a party flip that exploits the "reflection" or parity symmetry of the jet (12.1c). The effects of the single transformations are clearly visible in the average representation.

The Lorentz boost transforms all jets to the point of origin of the detector as shown in Figure 12.1a. Without this step, all particles would be – more or less – "uniformly" (at least in ϕ as can be seen in Figure 9.9b even though not in η as obvious from Figure 9.9a) distributed over the acceptance range of $[-2.5, 2.5] \times [0, 2\pi)$ defined by the (fiducial) phase space (see

[6]Besides the above-mentioned additional logistical effort, using a family of neural networks that have been trained on exclusive ranges in transverse momentum is conceptually questionable because this basically corresponds to using different jet substructure observables.

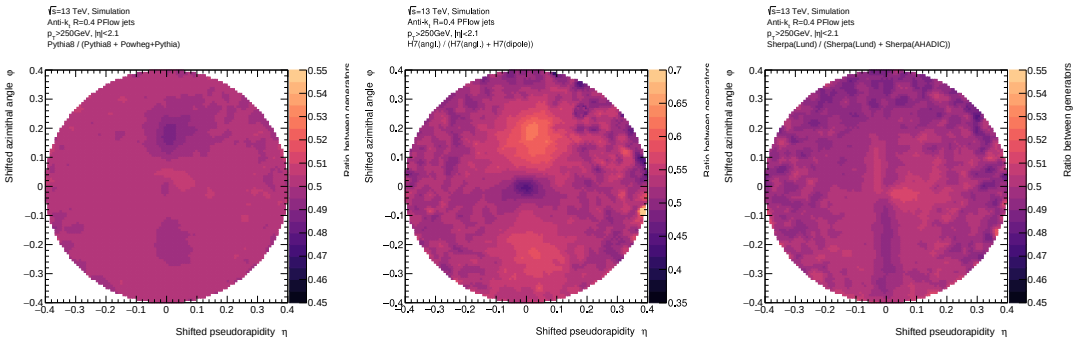
Chapter 9.3). Omitting this transformation, the neural network would be unable to identity the same jet (regarding the relative orientation and momentum of its constituents) at two different locations in the detector as being identical[7]. The standardized jets are now rotated based on the axis of maximum variance of the particle cloud. The result of this PCA transformation is shown in Figure 12.1b. As clearly visible, the largest variance is along the axis of abscissa due to the convention of letting the PCA pointing at 12 o'clock. The asymmetry between upper and lower have of the image is a result of the aforementioned ambiguity regarding the sign of the PC axis. In this convention, the sign is chosen such that it points to the hemisphere of largest energy. Last but not least, Figure 12.1c shows the parity transformation that breaks the symmetry with respect to "east" and "west" such that the largest amount of energy is always contained on the eastern side of the jet.

Jet Image Ratios[8]

The "images" according to Figure 12.1 show the structure of a large collection of jets for the respective preprocessing transformation or combination of those; hence, they can be considered as a representation of the average appearance of a jet given the respective transformation under consideration. As explained previously, after all preprocessing transformations have been applied to the jets, all trivial symmetries are exploited, which results in a representation that is perfectly suited to study the internal structure of jets. A particular interesting question is how the preprocessed and image-based representation allows to factorize different physical processes (e.g., matrix element (hard process), parton shower, hadronization etc.) such that they can be studied individually (similarly to the factorization of different processes in the Lund jet plane).

To further investigate this point, the *fully* preprocessed average jet image (as shown in Figure 12.1) is generated for *each* Monte Calo generator listed in Table 8.1 individually. The resulting two-dimensional distributions are hard to interpret; however, ratios between certain generators reveal interesting structures as shown in Figure 12.2.

Each Figure in 12.2 shows the ratios of a pair of two generators that differ with respect of



(a) Pythia8,Powheg+Pythia8 *(b)* Herwig7(angl./dipole) *(c)* Sherpa(Lund/AHADIC)

Plot 12.2: Ratios between fully preprocessed (Lorentz boost, PCA rotation, parity flip) jet images for different Monte Carlo event generators to factorize different physical effects. The three figures show the ratio between Pythia8 and Powheg+Pythia8 (*left*); both NLO Herwig7 samples with different parton shower models (*middle*); and both Sherpa samples that differ with respect to their hadronization model (*right*).

(at least) one aspect concerning the modeling of the event. For instance, Figure 12.2a shows the

[7]Of course, the neural network may implement such a "standardization" by itself; however, this requires significantly more training data as well as more training iterations for the model to archive this task. The anticipation of symmetries encoded in the data, relieves the neural network from the burden of learning those redundant information and hence results in a considerable improvement of training performance.

[8]The following section is not necessary for the general understanding of the analysis presented in this thesis and may be safely skipped by the reader

ratio of the fully preprocessed jet image for PYTHIA8 and POWHEG+PYTHIA8. Both samples use the very same parton shower model (PYTHIA's k_t-ordered shower), but differ with respect to the calculation of the underlying matrix element (LO *versus* NLO calculation). The ratio, therefore, allows to study the effect of different orders of the matrix element calculation on the internal structure of jets. As already stated, the resulting distribution(s) in Figure 12.2a for the comparison of the two PYTHIA8 samples with different orders of the matrix element calculation are mostly flat.

Going to Figure 12.2b that show the ratio between the two NLO HERWIG(7) generators that differ with regard to their parton shower model (angular-ordered *versus* dipole shower), a clear impact on the internal structure of the jets can be observed. This is expected because the internal structure of the jets is mostly formed in the parton shower step. The resulting distribution(s) are quite symmetric, which is expected due to the isotropic nature of the parton shower simulation. The noticeable shift towards positive values is a consequence of the parity transformation (3rd step in Section 12.2). However, both approaches to parton showers vary significantly concerning the amount of emissions/radiation generated close to the core of the jet. Generally, the dipole model seems to produce more radiation or harder emissions compared to the one based on an angular-ordered parton shower. Compared to the other modeling effects (matrix element and hadronization), the effect of different parton shower models is mostly pronounced, which is why it gives rise to one of the dominant sources of systematic uncertainties of the measurements presented in this report.

Last but not least, the third plot in Figure 12.2c shows the effect of different hadronization models on the internal arrangement of constituents and distribution of energy inside of the jet. The ratio is based on the two SHERPA samples used in this analysis (see Table 9.4). Compared to the parton shower or the matrix element, the SHERPA-based distribution sensitive to hadronization effects exhibit a richer structure that can be explained by the different number of soft and hard particles produced by the Lund- and the cluster-based hadronization model, respectively.

12.2.1 Training Data and Model Inference

The preprocessing steps introduced above are applied to the constituents of the jets *for each* jet individually. The question that remains to be answered is whether the model is supposed to be trained on data consisting of jets at *particle* or *detector level*. There are indeed many arguments which speak in favor of training the neural network on particle-level information as will be motivated in the following.

An important point to consider is the fact that the models are trained in a *supervised* fashion, which means that the labels indicating the associated class (signal(=gluon) or background=(light quarks)) must be available on a jet-on-jet bases, i.e., for each jet individually (which is conceptually problematic as explained in Chapter 7.1.1). Strictly speaking, this information is only available at *truth level*. Of course, a corresponding label can be assigned to the jets at detector level as well by matching it to a corresponding particle-level jet by means of some jet matching criterion (see Chapter 9.2.2). However, those truth-matching schemes are never fully efficient, with the corollary being that not all jets are matched correctly. This may introduce a dependency on the truth-matching scheme, which is meant to be avoided. Another argument giving support for preferring particle-level objects over detector-level ones is the circumstance that the latter ones include the full simulation of the ATLAS detector – an information that is supposed to be "marginalized" during the unfolding step. This simulation can never fully reflect the actual measurement apparatus in all gory details and hence may introduce a bias towards simulation-specific features present in detector-level jets that are absent in case of real collision data on which the model is evaluated.

Due to the arguments provided above, all models presented in this thesis have been trained on particle-level objects *exclusively*. The total training, validation and testing set consists of the very same jets introduced in Chapter 9.2.2, reconstructed from events generated with PYTHIA 8.230. The training data consists of leading *and* subleading as well as light quark and gluon jets in *equal parts* to avoid a bias of the model. In total, all models are trained on two million jets in total over the full range of transverse momentum. The implementation of the

neural networks as well as their training is done using `Keras` [227], `Tensorflow` [228] as well as the EnergyFlow package [229] for the EFN and PFN.

In order to use the trained neural network as an observable in the very same framework utilized to measure the track n_{trk} and the Soft Drop n_{SD} multiplicity described in Chapter 11, the PFN and the EFN *must* be ported to a format compatible with C++ for the purpose of doing inference. Doing a fully python-based analysis would be infeasible given the enormous amount of events in simulation and data as well as including all systematic variations and the reclustering of the jets involved. Loading the pre-trained models in C++ and – in particular – harmonize it with the ATLAS software environment was a non-trivial task. The conversion was accomplished by using the *Open Neural Network eXchange* (ONNX) framework, which is *"[t]he open standard for machine learning interoperability"* [230]. In this workflow, the pre-trained model is converted to a file format that can be loaded into a computational graph in C++ ready for inference[9].

An important point to mention is that even though the neural networks are trained on truth-level information, they are *also* applied to detector-level objects as part of doing inference to obtain the classification score of the trained model for, e.g., a reconstructed jet in real collision data. This is a crucial step and it was by no means clear *a priory* whether this application of the truth-level-trained model to detector-level objects will succeed! Towards all apprehensions, the models turned out to perform well for both particle and detector level which manifests itself in a clear correlation between the two and in the response matrix being sufficiently diagonal for the purpose of unfolding.

12.3 Network Architecture and Baseline Model

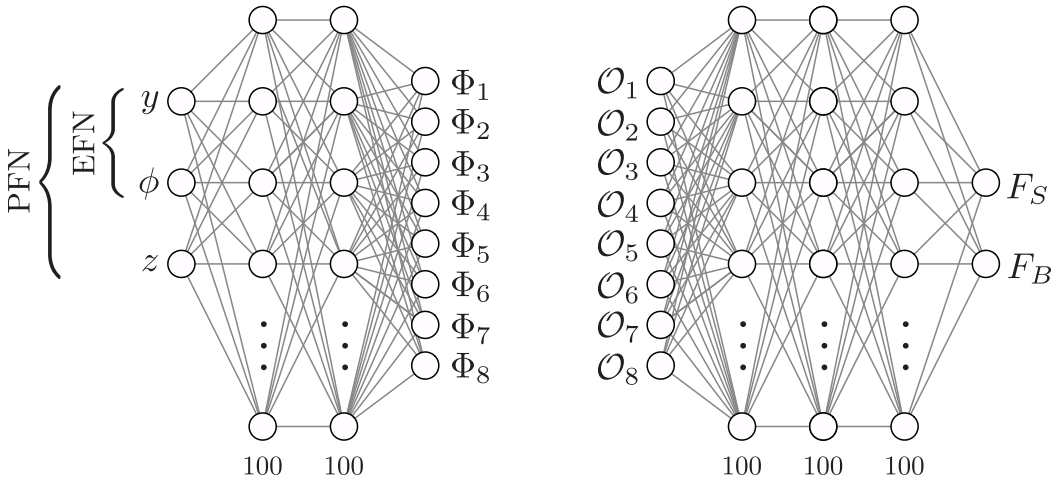
With the preprocessing of the data being concluded, this section quickly introduces the baseline model for the PFN and the EFN, respectively. According to Equation 12.2 and 12.1, Energy Flow networks (or rather Deep Set networks) consist of two dense feed-forward neural networks to approximate the functions F and Φ (or ρ, ϕ following the notation of the original publication). The permutation-invariant parameterization of the observable is imposed by the linear sum of the per-particle mappings. Regarding the optimization of the model parameter, the main focus is on the architecture of F and Φ as well as on the dimensionality $\ell \in \mathbb{N}$ of the latent space.

Concerning the two neural networks F and Φ, the main hyperparameter to optimize are the number of layers (i.e., the "dept" of the model) as well as the number of nodes for each layer. Both, EFN and PFN, turned out to be very robust with respect to different model configurations. For this reason, the Energy Flow networks used in this study are based on the baseline model presented in Ref. [27] (see Section 3.2).

The neural networks that are supposed to provide an approximation of Φ consists of *three* individual layers with the first two comprising 100 nodes and the number of nodes in the last layer corresponding to the dimension of the latent space ℓ. With this architecture, the approximated function is of the form $\Phi : \mathbb{R}^d \to \mathbb{R}^\ell$ with $(d = 3)$ in case of the PFN and $\Phi : \mathbb{R}^{d-1} \times \mathbb{R}^1 \to \mathbb{R}^\ell$ with $(d = 3)$ in case of the EFN, whereby $d \in \mathbb{N}$ denotes the number of features that are identical in case of the PFN and EFN. The dimensionality of the latent space ℓ is left unspecified at this point as the impact of this parameter will be further investigated for a vast set of dimensions later.

The architecture of the neural network intended to be an approximator of the function F closely follows the design of Φ ("mimicking paradigm"). The model consists of *three* layers with 100 nodes each. The output layer of F includes two nodes that provide a "softmaxed" signal, which gives the probability for the received input being compatible with a signal ($x \xrightarrow{F_{\text{S}}} 1$) or a background ($x \xrightarrow{F_{\text{B}}} 0$) event, whereby F_{S} and F_{B} denotes the output of the first and the

[9]The actual transition from python to C++ is somewhat more complicated and the interested reader is referred to the provided references to the respective repositories given at the relevant parts of this report. One "problem" was the python implementation of the EFN and the PFN in the EnergyFlow package [229] that utilizes a custom-designed masking layer without one-to-one correspondence in ONNX. To circumvent this problem, the source code in [229] had to be modified accordingly. This, however, does not affect the functionality of both models.



(a) Approximator of Φ *(b)* Approximator of F

Fig. 12.1: The deep neural networks corresponding to the *per*-particle mapping Φ (*left*) and the function F (*right*) with an 8-dimensional latent space for an EFN- and PFN-based architecture (adapted and modified from Ref. 27, Fig. 4, p. 14). The latent space dimension of '8' does *not* correspond to the parameter used for the measurement, but has been chosen for reasons of clarity to reduce the number of connections.

second node, respectively, as shown in Figure 12.1.

All dense layers described above use the ReLU activation function as well as He-uniform for initializing the weights of the models (see Chapter 6.4). Other activation functions and model primers have been studied with little to no impact in the final result.

Important to note: the model architecture illustrated in Figure 12.1 has two outputs: one of which corresponds to the probability of the received input being classified as "signal" $z_{\mathrm{S}}(= F_{\mathrm{S}}) \in [0, 1]$ (gluon-like jets) and a second one for the jet originating form "background" $z_{\mathrm{B}}(= F_{\mathrm{B}}) \in [0, 1]$ (jets initialized by a light quark) with the following normalization condition being enforced by the final softmax activation $z_{\mathrm{S}} + z_{\mathrm{B}} = 1$ (unitary). For the entire further course of the report at hand, the *Neural Network score* z_{NN} *always* – if not explicitly stated otherwise! – refers to the signal score, i.e., *the probability for the current jet being initialized by a gluon*, i.e., F_{S}.

12.4 Training and Model Optimization[10]

With the baseline architecture being settled, some remaining hyperparameters and their impact on the training as well as the performance of the final model must be studied. Those optimization studies should be conducted with great care to ensure that the models obtained during the hyperparameter scan are comparable with each other. To this end, *all* random number seeds have been fixed globally such that each model for a specific point in hyperparameter space can be reproduced deterministically. It also ensures that all models start from a similar initial position (they are usually not identical since some hyperparameters (like, for instance, the latent space dimension) change the number of trainable parameters and therefore the initial prior over the model too). Those optimization studies are quite comprehensive and usually consume a considerable amount of time; however, the insights gained (physics-wise) are rather moderate. Nonetheless, a small selection of results will be presented in this section for the

[10]The following section is not necessary for the general understanding of the analysis presented in this thesis and may be safely skipped by the reader – but I do not recommend it.

sake of completeness and to convince the reader that all models used in the measurements are trained properly and well understood.

Figures of Merit

In order to estimate and compare the performance of certain configurations of the neural network(s) in hyperparameter space (the space spanned by all adjustable/tunable parameters of the model), *Figures of Merit* (FoM) must be defined that either quantify the performance of the model concerning the discrimination power between signal and background, or, the conversion and training progress of the model(s) in general. An obvious choice for monitoring the training progress of the neural network is the *total loss* that quantifies the difference between the expected class and the corresponding prediction of the model at its current training state. For this analysis, the loss is given by the *categorical cross-entropy* (see Chapter 6.4.1) to properly account for the two outputs of the model shown in Figure 12.1. The loss should reduce with increasing number of training iterations for the model to converge to a (local) minimum. To estimate the classification performance, one could use, for instance, the overall classification accuracy, which is among the simplest evaluation metrics. A disadvantage of the accuracy is its significant dependence on the respective threshold value used to compute the metric; furthermore, it yields wrong results in case of data sets with highly imbalanced classes. Since the classification task of the model is binary (gluons and light quarks), the *Area Under the Curve* (AUC) represents a more appropriate choice because this metric is invariant with respect to different threshold values.

There are various ways of representing the foregoing metrics graphically. To keep as much information as possible, the loss and the AUC are evaluated batch-by-batch to estimate the variance of those metrics within the scope of a full training epoch. This variance may be large due statistical inter-batch variations (which is expected); therefore, the median Q_2 (second quartile) of the distribution is indicated separately along with the *InterQuartile Range* (IQR) that gives the difference between the 75 % and the 25 % percentiles of the data distribution as well as the related whiskers at $Q_2 \pm 1.5 \cdot \mathrm{IQR}$[11].

Using the training loss and the AUC (or the accuracy) as a figure of merit might seem redundant at first sight; however, it should be noted that both measures are fundamentally different and therefore not related with each other. The loss is a *metric* (in its proper sense, i.e., it fulfills the corresponding axioms) that measures a distance between the truth value and the ones predicted the model in its current state. Lower losses signal that the model improves with respect to the training data, but it does not necessarily imply higher AUCs (nor accuracy).

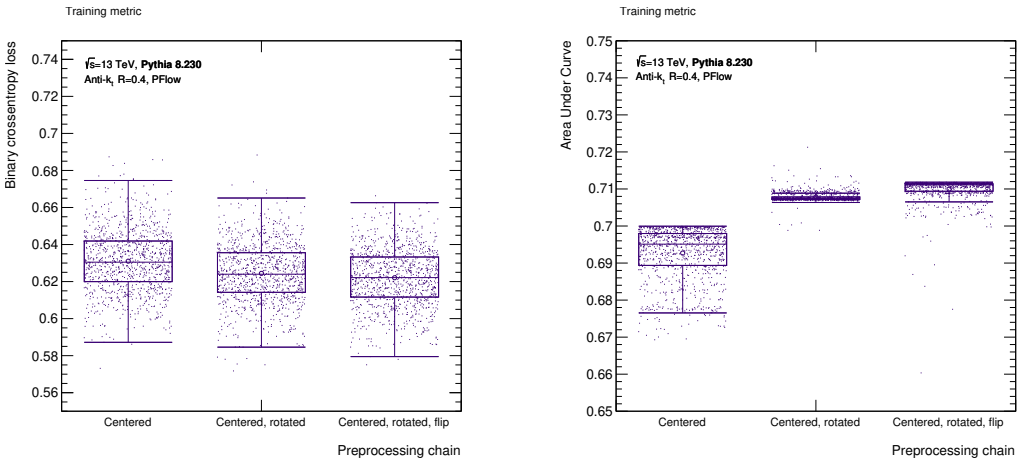
Preprocessing Chain

The first aspect to verify is if the tedious preprocessing strategy introduced in Section 12.2 actually comes along with an increasingly positive impact on the performance of the model with respect to the training stability as well as the discrimination between signal(=gluons) and background(=(light) quarks). Staring with the effect of the different preprocessing steps is appropriate since the insights gained at this point are (mostly) independent of the configuration regarding other hyperparameters such as the dimension of the latent space or the number of constituents in the jet to account for.

The data processing outlined in Section 12.2 can be divided into three main groups: standardization, PCA rotation and mirroring. The effect of the aforementioned data transformations is shown in Figure 12.3 for the training loss (12.3a) as well as the AUC (12.3b).

Staring with the evolution of the training loss for different preprocessing in Figure 12.3a, a comparatively weak, but a noticeable dependence can be observed. The first distribution corresponds the jets being transformed to the point of origin of the detector. This is the minimum necessary standardization that *must* be applied to the data in order to ensure stability of the model during training phase. Training the neural networks on completely

[11] If this sounds somewhat unfamiliar, think of the IQR as something similar as the (1σ) width of, e.g., a Gaussian but as a model-independent quantity.



(a) Loss *(b)* AUC

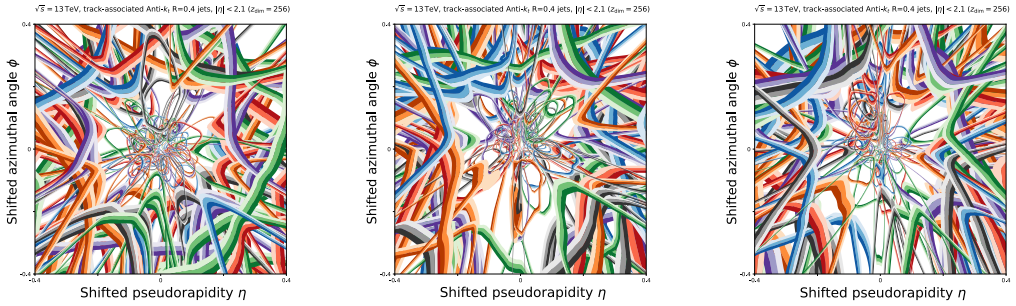
Plot 12.3: Effect of combining different preprocessing steps introduces in Section 12.2 according to the evolution of the training loss (*left*) and the AUC (*right*). Each distribution of points for a preprocessing method shows the full range of values taken by the respective metric for *all* training iterations.

*un*preprocessed jets instead, results in a significant performance loss with the model struggling to converge to a local minimum of the loss function. The subsequent distribution extends the centered jets by a rotation of the PC axis, which accounts for rotation invariance of the jets. Because symmetries in the data deteriorate training performance (the model has to infer those symmetry transformations during training without improving classification power), exploiting this symmetry (already at data processing) generally improves performance considerably. The same goes for including a parity transformation, which corresponds to the "mirror" symmetry of the jet image. Keep in mind that the improvements visible in Figure 12.3 are really *solely* due to the different preprocessing steps because *all* seeds of *all* random number generators have been fixed, i.e., all models start from the very same configuration. This means, in particular, that all models are identical (at the beginning of the training!); furthermore, all models receive the training points in exactly the same order. As can be seen from Figure 12.3a, the training loss takes a large range of different values during training. This variance with respect to the cost function is, of course, expected since the training data is divided into several batches, each one being composed of 500 individual data points(=jets) and its constituents. The composition of jets in one single batch of data is completely random; hence, there is a considerable statistical variation between consecutive training iterations for which the performance of the neural network may vary significantly – there might be low- and high-p_{T} jets present in the current batch. The noticeable point in the periphery (meaning: beyond the whiskers at $Q_2 \pm 1.5 \cdot \mathrm{IQR}$) of Figure 12.5a originate form the very first training iterations at which the model is basically untrained and the distributions over weights is random; at this stage, the model is basically randomly guessing classes for the input it receives. However, the accumulation of points in the denser region close to the median within the IQR indicates that the model converges with increasing number of training steps.

Figure 12.3b shows the effect of the different preprocessing combinations for the AUC, which estimates the classification performance of the binary classifier independent of a pre-defined threshold values that separates the two categories. In terms of classification performance, the impact of the different transformations is much more pronounced than for the training loss in Figure 12.5a. This result provides a strong motivation for using the full preprocessing chain outlined in Section 12.2. The significant improvement of the different transformations can be understood by recalling distinctive feature of gluon and (light) quark jets. In this picture, the Lorentz boost does not provide any support for discriminating between gluons and

quarks – it just prepares a standardized form of the data that can easily be analyzed by the machine-learning model. The rotation of the PC axis and the consecutive parity flip, on the other hand, provide a usefull representation in terms of the distribution of energy within the jet.

Given the significant improvement of classification performance, all jets used during training



(a) Lorentz boost *(b)* PCA rotation *(c)* Parity flip

Plot 12.4: Simultaneous representation of all filter boundaries $\mathbb{F}^{\ell}_{\Phi}$ for each component of Φ with $\ell = 256$. The illustrations above correspond to the three preprocessing steps introduced in Section 12.2: Lorentz boost (*left*), PCA rotation (*middle*) and parity transformation (*right*).

and inference are *fully* preprocessed by all steps introduced in Section 12.2, which includes a Lorentz boost, a PCA rotation as well as a parity transformation – in this particular order.

The effect of the different preprocessing strategies can also be studied by visualizing the latent space constructed by the neural network through investigating the learned function Φ. In case of EFNs, the map $\Phi : \mathbb{R}^2 \to \mathbb{R}^{\ell}$ is a function that solely depends on the angular components η, ϕ of the particle's four-momentum and hence can be directly visualized in an η–ϕ-plane. To visualize the structure of Φ, the function is evaluated on a dense, pixelized $(\eta, \phi) \in [-R, R]^2$ grid, whereby $R(= 0.4)$ denotes the radius of the jet used in the reconstruction algorithm. This results in a set of ℓ filters $\mathbb{F}^{\ell}_{\Phi} := \{(\eta, \phi) \in [-R, R]^2 \mid \Phi_{\ell}(\eta, \phi) \in \mathbb{R}\}$, with ℓ being the number of latent space dimensions, all of which together provide a visual representation of the latent space learned by the neural network approximating Φ. Each one of the ℓ filters $\mathbb{F}^{\ell}_{\Phi}$ defines a "surface" (or an active "area") that can be used to determine a "boundary" constructed from the contour of the full activity for each individual filter. Following the convention in Ref. [27], those boundaries $\mathbb{B}^{\ell}_{\Phi}$ (one for each component of Φ as shown in Figure 12.1) are given by all points of $\mathbb{F}^{\ell}_{\Phi}$ that fall in the range from 45 % to 55 % of the maximum value taken by Φ, i.e., $\mathbb{B}^{\ell}_{\Phi} := \{x \in \mathbb{F}^{\ell}_{\Phi} \mid 0.45 \cdot \sup(\mathbb{F}^{\ell}_{\Phi}) \leq x \leq 0.55 \cdot \sup(\mathbb{F}^{\ell}_{\Phi})\}$. Those boundaries can simultaneously be visualized to picture the entire latent space at once as shown in Figure 12.4 for the three preprocessing steps by overlapping all boundaries for all filters (for a more detailed introduction, the interested reader is referred to the original publication [27] that introduced this visualization method).

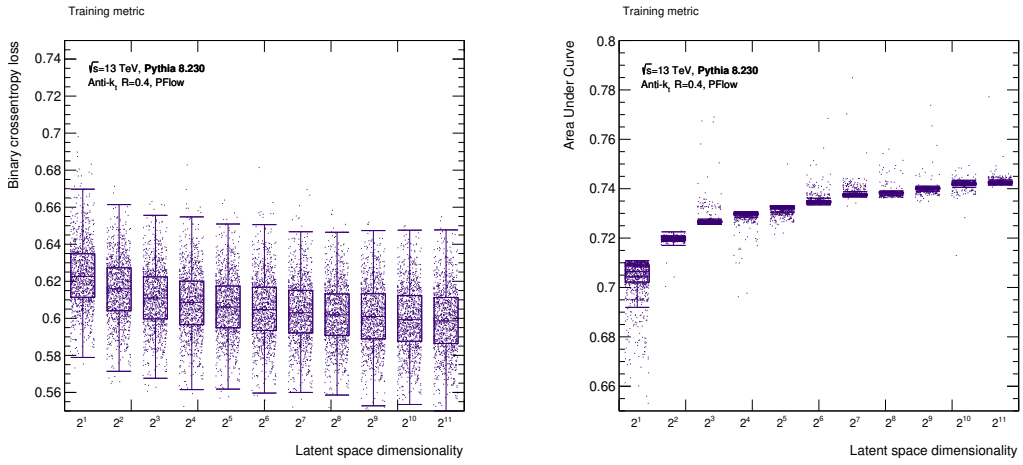
All three figures in Figure 12.4 show an interesting structure that can be explained by the respective preprocessing method. The first Figure 12.4a corresponds to the jets being only transformed by a Lorentz boost such that they are centered in the detector. Accordingly, the resulting filter boundaries are very symmetric with respect to the origin of the image. Furthermore, the density of filters significantly increase towards the center of jet, which is in agreement with the corresponding Figure 12.1 where most activity is in the core of the jet; hence, the neural network identifies most of the relevant structure being present in this region of the jet. The adjoining Figure 12.4b includes the already mentioned rotation of the PC axis. This rotation "breaks" the symmetry along the horizontal because the PC axis points at north (compare the counterpart Figure 12.1b). The effect of this preprocessing step is clearly recognizable by the arrangement of filters through a strong asymmetry between the lower ($\Delta\phi \in [-R, 0)$) and the upper ($\Delta\phi \in [0, R]$) part of the image. The same is true for the full preprocessing chain that includes the parity transformation in Figure 12.4c. This

step additionally introduces an asymmetry between the left ($\Delta\eta \in [-R, 0)$) and the right part ($\Delta\eta \in [0, R]$) of the image, which, again, is clearly recognized by the neural network.

Latent Space Dimension

In the previous Section 12.3 – which introduced the baseline model for the EFN as well as PFN –, the dimensionality of the latent space ℓ was purposely left unspecified. The size of the latent space, however, is among the most important model parameters as its purpose is to embed the constituents of the jets into a latent description of the jet given by the linear sum in Equation 12.2 and Equation 12.1. The dimension of the latent space can be thought of as a "bottleneck" for the information extracted from the particles. Accordingly, larger latent spaces allow the model to learn more features of the input data and hence usually result in better discrimination between signal and background – at the cost of significantly increased training parameters and (most likely) increased model dependency. The first step (regarding the optimization of hyperparameters) is therefore to study the effect of various latent space dimensions and to settle on a value that will be used for the entire further course of this analysis.

Figure 12.5 shows the impact of different latent space dimensions in increasing powers of



(a) Loss *(b)* AUC

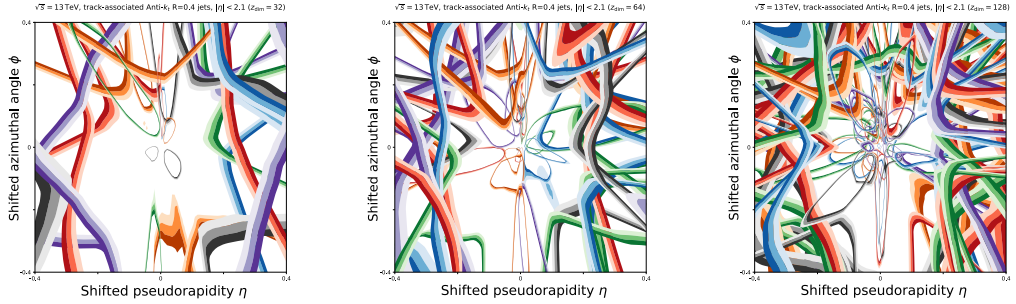
Plot 12.5: Effect of different latent space dimensions ℓ according to the evolution of the training loss (*left*) and the AUC (*right*).

two $\ell \in \{2^n\}_{n=1}^{8}$ for the training loss (12.5a) and the AUC (12.5b). For each dimension of the latent space, all values of the respective metric taken during training are shown (marginalized over *all* training iterations). Studying the evolution of the median with increasing latent space dimensions, a clear tendency towards smaller values of the loss can be observed. This basically indicates that the model improves in its classification task with increasing number of latent space dimensions.

Figure 12.5b shows the same information, but with the cross entropy loss being replaced by the AUC. The first feature to notice by comparing Figure 12.5b with Figure 12.5a is the considerably smaller variance of values as well as the clear improvement of classification power for larger latent space dimensions. In case of just two latent space dimensions, the variance is large compared to the other configurations, which indicates that the model requires considerable more training iterations to converge to a comparatively small AUC. With increasing ℓ, the model is able extract more features from data, which improves the discrimination between gluons and light quarks. A further small improvement of the AUC can be achieved by going beyond a latent space dimension of 2^8; however, the number of trainable model parameters grows significantly and therewith the time required to train the model and – most importantly – to do inference. In general, the trained model shows an overall good discrimination performance for gluons and light quarks with an AUC of approximately 74 % for a latent space size of 2^8,

whereby the AUC presented here is inclusive in p_{T}, i.e., the classification performance is not shown for jets within exclusive p_{T} regions.

Based on the results presented in this paragraph, the dimension of the latent space for the



(a) $\ell = 32$ *(b)* $\ell = 64$ *(c)* $\ell = 128$

Plot 12.6: Simultaneous representation of all filter boundaries $\mathbb{F}_{\Phi}^{\ell}$ for a latent space dimension of $\ell = 32$ (*left*), $\ell = 64$ (*middle*) and $\ell = 128$ (*right*).

EFN as well as the PFN was chosen to be 2^8, which is a good compromise between classification performance as well as computational time. Given this latent space configuration, the EFN consists of a total of $82,358$ trainable weights. This number is slightly larger in case of the PFN with $82,458$ weights due to the different arrangements of the input in the first layer of the model.

Like it was done in the previous paragraph for the different preprocessing steps, it is instructive to visualize the latent space for some dimensions ℓ. As a reminder: for a latent space with dimension ℓ, a set of filter boundaries $\mathbb{F}_{\Phi}^{\ell}$ can be constructed and simultaneously visualized. The result is shown in Figure 12.6 for three selected examples $\ell \in \{32, 64, 128\}$.

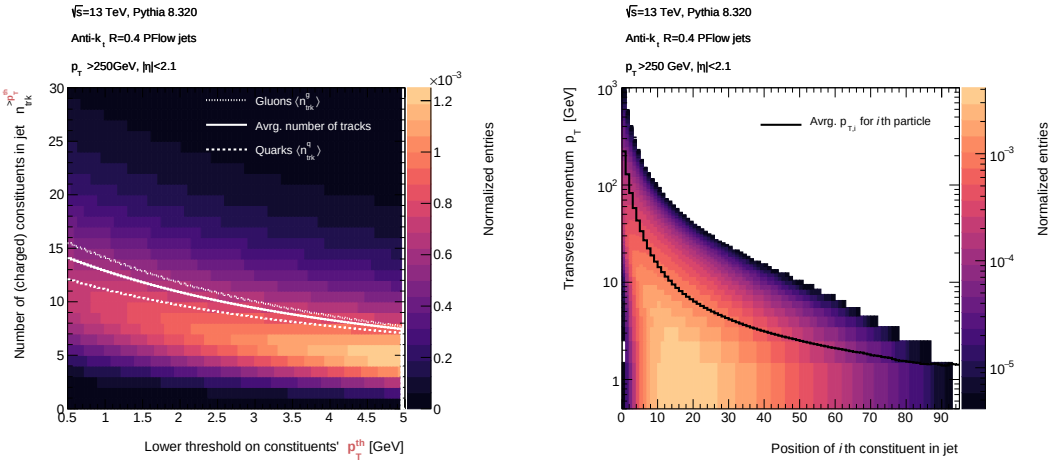
As can be seen in Figure 12.6, the neural network gains flexibility for increasing latent space dimensions and is able to extract more features from the internal structure of the jet. The corresponding arrangement of filters for a latent space dimension of $\ell = 256$ used in this measurement has already been shown in Figure 12.4c. The number of "boundaries" in Figure 12.6 corresponds to the dimension of the respective latent space.

Number of Constituents

The next item on the list of parameters to optimize is the number of constituents n_{trk}[12] (tracks in this case) inside the jet that are supposed to be used when computing the neural network score. In principle, all particles clustered to a jet should be used; however, because the neural network is trained on data derived from Monte Carlo, this number should be chosen carefully to avoid a bias of the model towards the respective generator that was used to produce the training data, which will result in an increased data-driven non-closure uncertainty (see Chapter 10.8). This is due to the fact of significant deviations between different families of Monte Carlo generators regarding the modeling of "soft" constituents inside of the jets. Reducing the impact of those particles associated with low p_{T} is the objective of the upper cut on the number of constituents $n_{\mathrm{trk,cut}}^{\mathrm{up}}$.

For this cut to be meaningful, an ordering scheme must be applied to the constituents of the jets. The most obvious choice is to sort all particles $\{j\}_{j \in J}$ associated/matched to a jet J in *decreasing* order of transverse momentum $p_{\mathrm{T},i} > p_{\mathrm{T},i+1}$ for all $i \in \mathbb{N}^{\leq |J|}$ such that the first constituent *always* corresponds to the hardest one. As explained in Chapter 6.3, "Deep Sets" are permutation invariant functions over sets; hence, the just arbitrary imposed ordering scheme does *not* affect the neural network score *at all* (cutting on the number of constituents, of course, does). With this ordering in place, an upper cut of $n_{\mathrm{trk,cut}}^{\mathrm{up}} \leq |J|$ can be applied to

[12] An alternative notation that is used throughout this report is n_{ch}, which denotes the number of charged constituents.



(a) Lower p_T *(b)* Constituent's p_T

Plot 12.7: Constituent/track multiplicity as a function of a lower p_T cut (*left*) and constituent p_T as a function of its position in the jet after sorting (*right*).

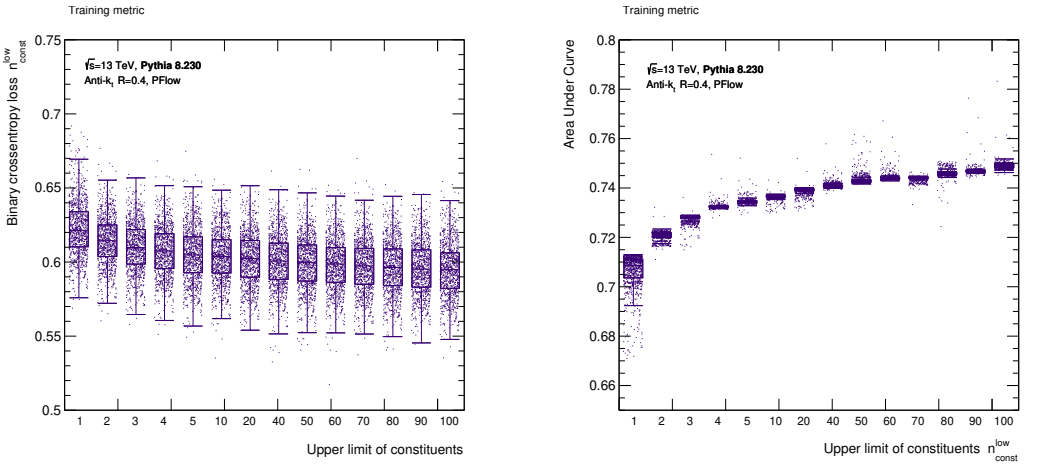
all jets, with all constituents $i > n^{\mathrm{up}}_{\mathrm{trk,cut}}$ being discarded, whereby i denotes the position of the ith constituent inside of the jet (sorted according to the p_T). Of course, if the number of particles in the jet is below $n^{\mathrm{up}}_{\mathrm{trk,cut}}$, the jet remains unchanged.

The result of the aforementioned ordering of the jet's constituents is shown in Figure 12.7b that shows the distribution of the transverse momentum for each constituent as a function of its position in the jet *after* sorting.

A very important point to consider is that the jet is fully preprocessed *before* any cut on the number of constituents is applied. This detail will prove important in just a moment; so, keep it in mind.

The effect of several upper cuts on the number of particles is shown in Figure 12.8 for the loss (12.8a) and the AUC (12.8b) in Figure 12.8.

As one can see from the loss in Figure 12.8a, the training performance in terms of the



(a) Loss *(b)* AUC

Plot 12.8: Effect of cutting on the number of constituents inside the jet according to the evolution of the training loss (*left*) and the AUC (*right*).

loss is only weekly affected by this cut. Nonetheless, there is a clear tendency towards lower losses for more constituents being kept inside the jet, implying that the model improves in its

classification task. This, of course, is expected as more information for the classification task is available. More interesting insights arise when taking a look at the AUC in Figure 12.8b. As expected, the classification performance clearly improves for keeping more constituents inside of the jets as the particle multiplicity is among the strongest discriminators for quark- and gluon-initialized jets. Astoundingly, even for just *one* (i.e., the hardest) constituent being used during evaluation of the model, the classification power is *surprisingly* high (as indicated by an AUC of roughly 71 % over all jets). This is not due to a mistake but a consequence of the preprocessing of the jet (see Section 12.2). As emphasized previously, the jet is fully preprocessed *prior* to the cut on the number of constituents! This has interesting consequences: if the jet is transformed to the point of origin of the detector, the internal distribution of the constituents, i.e., their relative orientation remains unchanged. Since gluon jets (on average) contain more emissions than quark-initialized ones, the total energy of the jet is distributed over a larger number of particles, which results in broader jets due to energy-momentum conservation. This implies that the average radial distance $\langle \Delta R \rangle$ of the hardest constituent ΔR with respect to $(0, 0)$ in the η–ϕ grid is larger for gluons than for quarks $\langle \Delta R \rangle_g > \langle \Delta R \rangle_q$. This indicates that in case of just one constituent being included, the neural network learns an observable that is strongly correlated to the "jet width" $\sum_{j \in J} p_{\mathrm{T},j} \Delta R(j, J)/p_{\mathrm{T},J}$, which is another discrimination variable commonly used in quark-gluon discrimination. Additionally, the constituents of a gluon jet are on average on a lower energy scale that the quark one's due to the already mentioned larger number of constituents. The discrimination power based on the energy or the p_{T}, however, is much weaker compared to the width or the constituent multiplicity of the jet and does unfortunately not lead to mutual irreducible probability distributions (see Chapter 7.3).

To strike the right balance between a potential bias of the trained model towards the baseline generator (which is used to train the neural network) and the discrimination power that is strongly correlated to the number of particles, the main focus was on the diagonality of the response matrix as well as the data-driven (non) closure. Those studies include numerous iterations of the *entire* analysis for different upper constituent cuts including the complete *re*-training of the models, reconstruction of the triple-differential multijet cross section, unfolding etc. Presenting and discussion all this information would go far beyond the scope of this report; therefore, the conclusion is just quoted at this occasion. After all studies involving the aforesaid steps, an upper limit of $n^{\mathrm{up}}_{\mathrm{trk,cut}} = 50$ on the number of (charged) constituents inside the jet turned out to be very appropriate. But, it should not be left unmentioned that reducing this number down to 30 particles almost does not affect the final result at all (as anticipated from Figure 12.8b).

Alternatively to directly applying a cut on the number of constituents inside the jet, it could be considered to introduce a lower threshold criterion on the transverse momentum p_{T} of the constituents above the already used 500 MeV below which all particles are discarded. Of course, this lower threshold $p^{\mathrm{th}}_{\mathrm{T}}$ is (anti)correlated with the constituent multiplicity as can be seen in the corresponding Figure 12.7a (increasing the cut on the constituents' momenta lowers the number of tracks in the jet). This figure also shows the average number of constituents for quark and gluon jets; as expected from QCD, gluon-like jets tend to contain larger number of particles than quark jets. Since the discrimination power of the classifier strongly depends on the discrepancy between the gluon- and quark-associated curve, the lower cut on the particles' p_{T} (as well as on the upper limit on the constituent multiplicity) should be chosen in view of keeping the distance between the two curves as large as possible.

EFN *versus* PFN

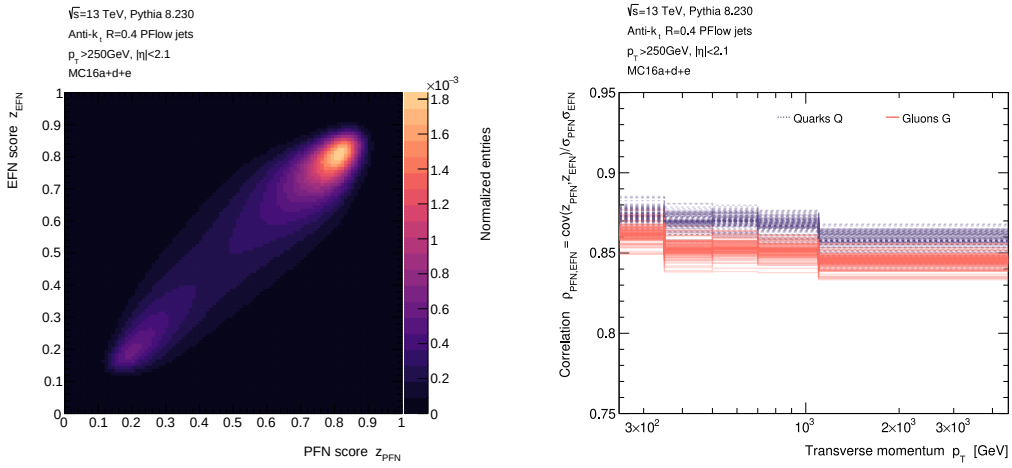
In Section 12.1, the EFN and PFN have been introduced, both follow the philosophy of the "Deep Sets" paradigm for the purpose of learning a function that is invariant with respect to permutations of the elements in the set over which the function is defined. Both networks are quite similar regarding their architecture. The PFN basically is an implementation of the model proposed in the pioneering publication in Ref. [3]; the EFN introduces a small modification by factorizing the positional and energy components as shown in Equation 12.2. Due to the close resemblance of energy- and particle-flow networks, the neural network score z_{EFN} and

z_{PFN} of the two models is expected to be highly correlated with each other. Investigating this correlation (among other things) is subject of this paragraph.

The direct correlation between z_{EFN} and z_{PFN} is shown in Figure 12.9a, which shows the neural network score for the very same jets in case of an EFN- and a PFN-based model inclusive in transverse momentum. The close connection between EFNs and PFNs is immediately evident: both scores are strongly correlated, whereby the degree of correlation depends on the transverse momentum of the jets. According to the strong correlation, both models are almost identical with respect to their discrimination power regarding gluon and (light) quarks. An argument in the favor of an EFN rather than the PFN is the already mentioned feature of the former one resulting in IRC-safe observables – in principle. This statement is true, however, it does not hold in case of this measurement since it (this measurement) is fundamentally *not* IRC-safe due to the use of "track-based/matched" jets. But, it certainly is an argument for future measurements if this analysis is extended to one that account for all constituents in the jet[13].

Marginalizing the joint distribution in Figure 12.9a along z_{EFN} and z_{PFN} results in the corresponding inclusive distributions of the two network scores. The outcome of this projection is shown in Figure 12.10a. The two distributions nicely illustrate the features of the neural network score, i.e., two distinct regions of an enhanced cross section each one of which corresponds to one category the model is supposed to discriminate. In accordance with the definition of the class labels in Chapter 9.2 and Equation 9.4, a neural network score closer to '1' indicates gluon-like jets while smaller values close to '0' are likely due to quark-initialized jets accordingly. The detailed interpretation of those distributions is subject to another section of this chapter; at this point, it should only be mentioned that all distributions look as expected and exhibit the desired properties.

Figure 12.10b shows the neural network score for EFN and PFN in different bins of



(a) z_{PFN} *versus* z_{EFN} *(b)* Inclusive PFN and EFN score

Plot 12.9: Correlation between the predicted network score of the PFN and the EFN each one evaluated for the very same jets (*left*) and the distributions for the EFN and PFN score inclusive in p_{T} for different seeds (*right*).

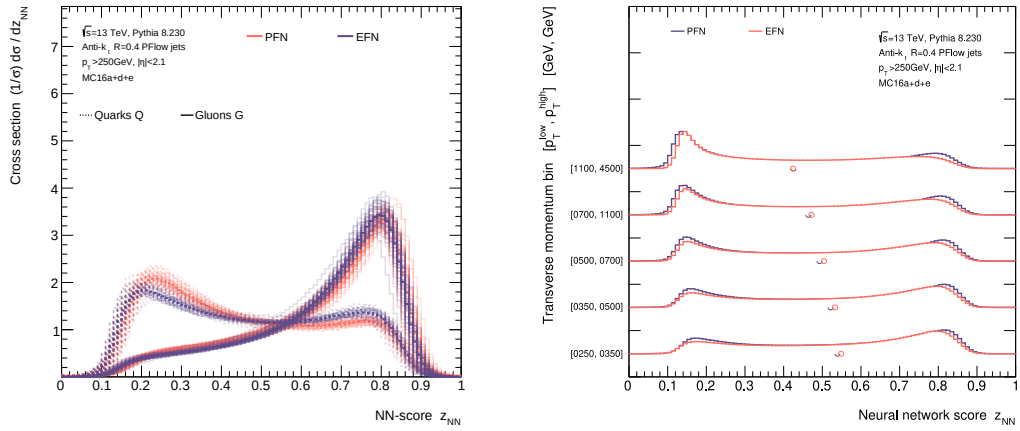
transverse momentum represented by an evolution graph with arbitrary normalization. The indicated circles correspond to the mean value of the distributions. As can clearly be seen, the distributions show a strong dependence on the p_{T} of the jets, which itself is highly correlated to the quark and gluon fraction of the samples. This important point (which is one motivation for using a neural network in the first place) will be dealt with in more detail in the upcoming sections; hence, it should only be noted (for the sake of brevity) that the distributions follow the

[13] IRC-safety in the context of the NN-score means that one could – in principle – compare analytic calculation of z_{NN} with a measurement of the probability distribution/differential cross section $P(z_{\text{NN}})$.

expectations from the theoretical calculations regarding the quark- and gluon-related component of the inclusive dijet cross section and its "evolution" with the transverse momentum of the jets. Overall, dijet samples are *gluon enriched*; however, the fraction of quark-initialized jets increases – which would result in a larger DDNC uncertainty (see Chapter 10.8) – towards higher transverse momentum as can be seen in Figure 12.10b through a shift of the distribution in direction of lower values of the neural network score (i.e., "quark-like" jets).

In summary, it can be said that the score given by the PFN and the EFN are highly correlated and therefore have a similar performance regarding the discrimination between quark and gluon jets. The PFN shows a slightly better discrimination power between the two categories because the architecture defined by Equation 12.2 is more general and therefore more flexible, with regard to the discrimination task.

The ROC curves in Figure 12.11 show the classification performance of both binary



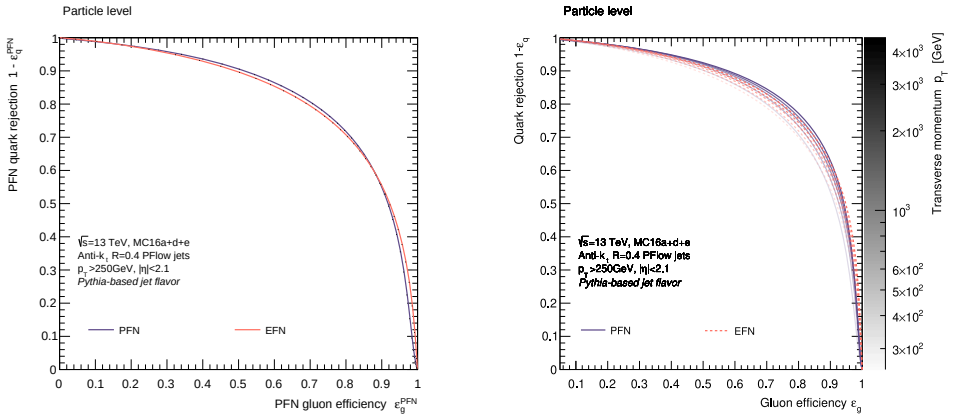
(a) Inclusive PFN and EFN score

(b) Evolution with p_{T}

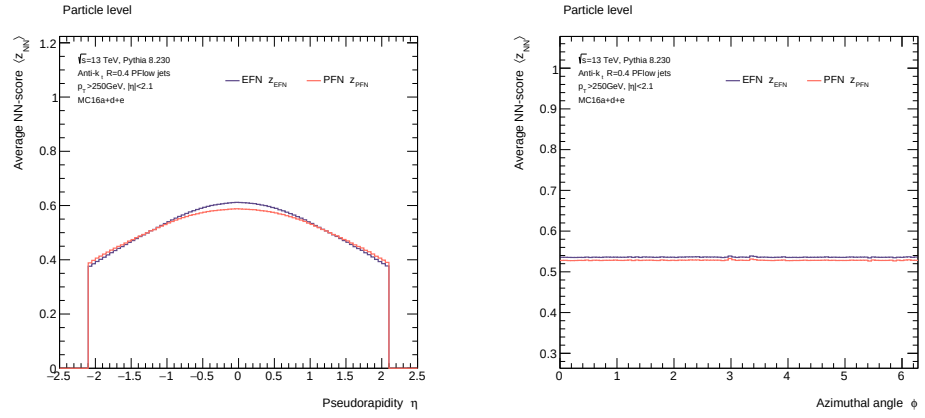
Plot 12.10: The distributions for the EFN and PFN score inclusive in p_{T} (*left*) and for the individual p_{T} bins used in this measurement (*right*). Each underlying model in the distribution above used a different seed for the random number generators for the initialization of the weights, which results in a distribution over models.

classifiers in case of the baseline generator (for all MC campaigns) inclusive in transverse momentum p_{T} as well as the discrimination power of the models for the individual p_{T} bins of the measurement. As can be seen and already expected from the distribution of the NN-score in Figure 12.10a, the PFN exhibits a higher discrimination power with respect to quark and gluon jets compared to the EFN. This follows from the already mentioned differences between the two architectures, which allows the PFN to extract more correlations between the individual constituents of the jet.

Turning the attention to a somewhat more technical aspect, Figure 12.12 shows the average of the NN-score as predicted by the PFN as well as the EFN as a function of the jet's pseudorapidity (12.12a) and the azimuthal angle (12.12b). Starting with the pseudorapidity in Figure 12.12a, a clear correlation with the (average) NN-score for both, PFN and EFN, can be observed. This is easily explained by the correlation between the quark/gluon fraction and the pseudorapidity of the jets in the detector (one motivation for changing the normalization scheme for the scalar components). As mentioned several times already, this property immediately follows from the underlying PDF of light quarks and the gluon, resulting in more gluon-initialized jets in the central region of the detector on average. This characteristic feature of the proton's PDF is one of the main motivations of measuring the triple-differential multijet cross section in two disjoint categories of rapidity, with the objective to obtain two mixtures of quark- and gluon-initialized jets. Turning back to Figure 12.12a, central jets ($\eta \approx 0$) tend to receive (on average) higher NN-scores by the neural networks, indicating that the fraction of



Plot 12.11: Classification performance of the PFN and EFN evaluated by means of the ROC curve for the baseline generator (PYTHIA 8.230) inclusive in transverse momentum (*left*) and for the individual p_{T} bins of the measurement (*right*).



(a) Pseudorapidity η *(b)* Azimuthal angle ϕ

Plot 12.12: The average NN-score for the PFN as well as the EFN as a function of the pseudorapidity (*left*) and the azimuthal angle (*right*).

gluon-initialized jets is larger compared to more forward jets.

On the contrary, no correlation between the fraction of quark and gluon jets and the azimuthal angle ϕ is expected. This is confirmed by Figure 12.12b, which shows once again the average NN-score of the PFN and EFN *versus* the azimuthal angle of the reconstructed jets.

Configuration Summary

After all model-specific parameters have been defined, the training parameters such as the optimizer and learning rate (see Chapter 6.4) must be determined. Generally, it can be said that the models used in this analysis are quite robust with respect to the choice of training-related hyperparameters. Therefore, based on a few superficial studies, the ADAM optimizer [231] at its default configuration with a learning "rate"/weight of $\alpha_l = 0.0005$ was chosen.

Furthermore, both models are trained for 200 epochs with a batch size of 500 data points(=jets), which results in 300000 iterations given that the full training data set consists of two Million jets. For the sake of clarity, all data from the previous sections regarding the configuration of the model or the training setup are summarized in detail in Table 12.1.

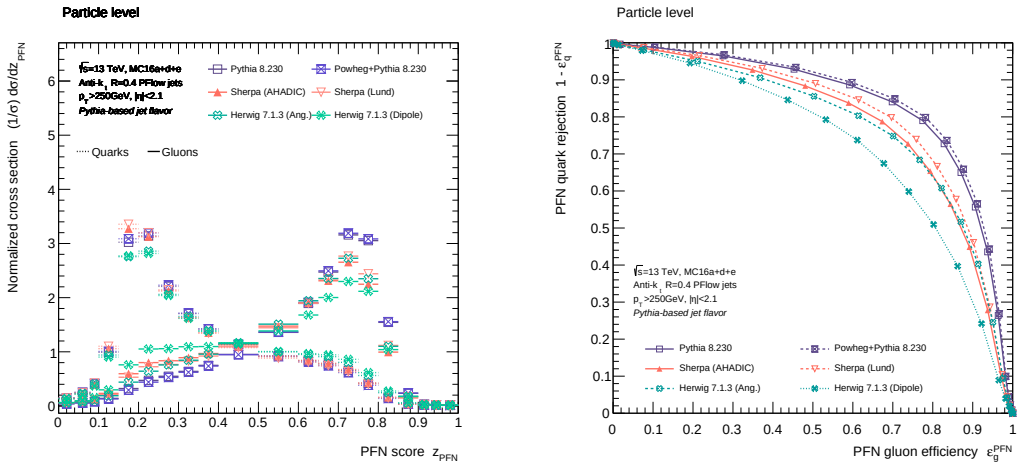
	EFN	*PFN*
	Model	
Latent space	256	
N parameters	$82,358$	$82,458$
N features	3	
$Layer_{Nodes}$	$F(1_{100}, 2_{100}, 3_{100})$, $\Phi(1_{100}, 2_{100})$	
	Input	
Type	Lead. and sublead. jet in fid. phase space at PL	
Features	$\left\{\left(p_{\mathrm{T},j}/E_J, \left(\eta'_j, \phi'_j\right)\right)\right\}_{j\in J}$	$\left\{\left(p_{\mathrm{T},j}/E_J, \eta'_j, \phi'_j\right)\right\}_{j\in J}$
n_{trk}	$\in [1, 50]$	
Preprocessing	Lorentz boost, PCA rotation, parity flip	
	Training	
Optimizer	$\mathrm{Adam}(\alpha_l = 0.5 \cdot 10^{-4}, \beta_1 = 0.9, \beta_2 = 0.999)$	
N training	2 Million events	
N epochs, Batch Size	200, 500	

Table 12.1: Summary table of the most important hyperparameters regarding the configuration of the baseline models (EFN and PFN), their input as well as the training of those.

Final Distribution and Discrimination Performance

This paragraph concludes this somewhat technical section by presenting the final distribution of the PFN's score (the corresponding score predicted by the EFN is very similar) for quark- and gluon-labeled jets, respectively, as well as the resulting classification performance. The histograms in Figure 12.13 show the distribution of the neural network score z_{PFN} associated with quark- and gluon-initialized jets inclusive in transverse momentum for all Monte Calo samples used in this measurement (see Table 8.1 for the total list of the generators used in this analysis along with their distinguishing features). At this point, it is important to highlight, once again, that the jet flavor definition (jet-by-jet label) introduced in Chapter 9.2.2 was first introduced and optimized for Pythia-related studies within ATLAS. Nonetheless, for all alternative generators in Figure 12.13, the very same definition was used in order to construct the corresponding distributions of the score for quark and gluon jet. One should therefore always keep in mind that the efficiency of this flavor definition was not studied in great detail for generators besides Pythia. Accordingly, caution should be exercised in interpreting the results – even though if on average one should expect to get the correct assignment[14]. After an initial rough inspection of the data, it can be stated that the general tendency of the distributions of the network score agrees between *all* Monte Carlo samples, which confirms the statement that the label definition on average results in the correct jet flavor being assigned. As mentioned previously, higher scores correspond to higher probability of the current jet to originate from a gluon. This feature is clearly visible in Figure 12.13 for all samples. Accordingly, the lower the score, the higher the probability of the jet being gluon initialized. As one can see, there is an excellent separation between the different categories with an non-negligible overlap between quark and gluon jets. The latter effect is unavoidable to a certain degree as quark- and gluon-initialized jets are not perfectly separable. It cannot be ruled out that

[14]This assumption is justified as the jet label definition is based on the hardest constituent in the jet, which is usually well modeled in Monte Carlo.



(a) Distributions *(b)* ROC curves

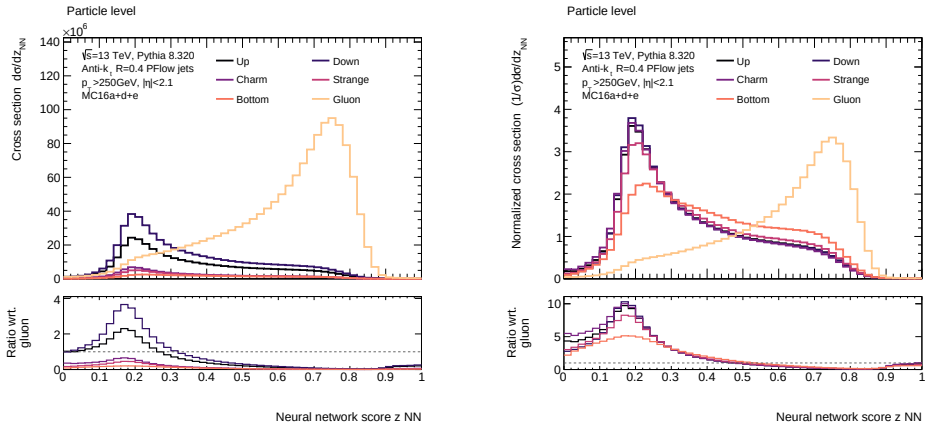
Plot 12.13: Quark and gluon (PYTHIA-based label) distributions of the PFN score z_{PFN} for all Monte Carlo samples used in this measurement (*left*) as well as the corresponding ROC curves derived therefrom (*right*).

further optimization of the model will result in an improved classification performance – one the contrary: it is quite likely. But, given the fact that (at LO) the difference between light quarks and gluons is basically determined by the color factors $C_{\mathrm{A/F}}$ (which further reduces discrimination power at higher orders as explained in Chapter 11.1), indistinguishable phase space configurations are expected.

Based on the distributions in Figure 12.13a as well as a predefined threshold cut separating the two categories, the ROC curves for each Monte Carlo generator can be computed to get the gluon efficiency ϵ_g^{MC} and the corresponding quark rejection $\epsilon_q^{\mathrm{rej,MC}} := 1-\epsilon_q^{\mathrm{MC}}$ for different thresholds $z_{\mathrm{th}} \in [0,1]$. In this case, the gluon efficiency is given by $\epsilon_g^{\mathrm{MC}}(z_{\mathrm{th}}) = \int_{z_{\mathrm{th}}}^{1} \mathrm{d}z_{\mathrm{PFN}}\, P_q^{\mathrm{MC}}(z\mathrm{PFN})$, with the quark rejection $\epsilon_q^{\mathrm{rej,MC}}(z_{\mathrm{th}}) = 1 - \int_0^{z_{\mathrm{th}}} \mathrm{d}z_{\mathrm{PFN}}\, P_q^{\mathrm{MC}}(z_{\mathrm{PFN}})$ being defined accordingly, whereby $P_f^{\mathrm{PFN,MC}}(z_{\mathrm{NN}})$ denote the *normalized* probability distribution of the PFN's score for jet flavor f (gluon (g) or light quark (q)) for one of the used Monte Carlo generator. The result is shown in Figure 12.13b.

Analyzing the ROC curves corresponding to the different Monte Carlo event generators in Figure 12.13b individually, one notes that PYTHIA as well as POWHEG+PYTHIA show the best classification performance. This is expected because the classifier model was trained on data generated from this Monte Carlo. Furthermore, the difference between the two PYTHIA-based samples is almost unnoticeable small, which is expected as the main discriminating features between quark- and gluon-initialized jets are generated during parton shower simulation. Accordingly, the remaining Monte Carlo generators perform worse compared to the baseline sample, but are in a similar range of numbers.

The last point to be discussed in the context of this section is the response of the neural networks for the individual jet flavors. This point is important because the basic assumption for the extraction of the underling quark/gluon distributions and fractions using "topic" modeling (as described in Chapter 7) is that the measured spectra are a mixture of contributions of quarks and gluons, i.e., only *two categories*. Since the final result of this analysis are the fractions of (light) quark and gluon jets in the different regions of transverse momentum, it is important to verify that the neural network (roughly) shows the same performance for all of the different quark flavors. Assumed the neural network would tend to systematically assign much higher scores for jets originating from beauty quarks, the finally measured "gluon fraction" would not correspond to (pure) gluons but to a mixture of gluons and beauty quarks. This does certainly *not* mean that there must not be any overlap between the different flavors at all – this is impossible because the phase spaces populated by (light) quarks and gluons are



(a) Non-normalized *(b)* Normalized

Plot 12.14: Inclusive single-differential cross section of the NN-score as predicted by the PFN (very similar results are obtained for the EFN) for all light quark flavors and the gluon non-normalized (*left*) and normalized to unity (*right*) for the validation set.

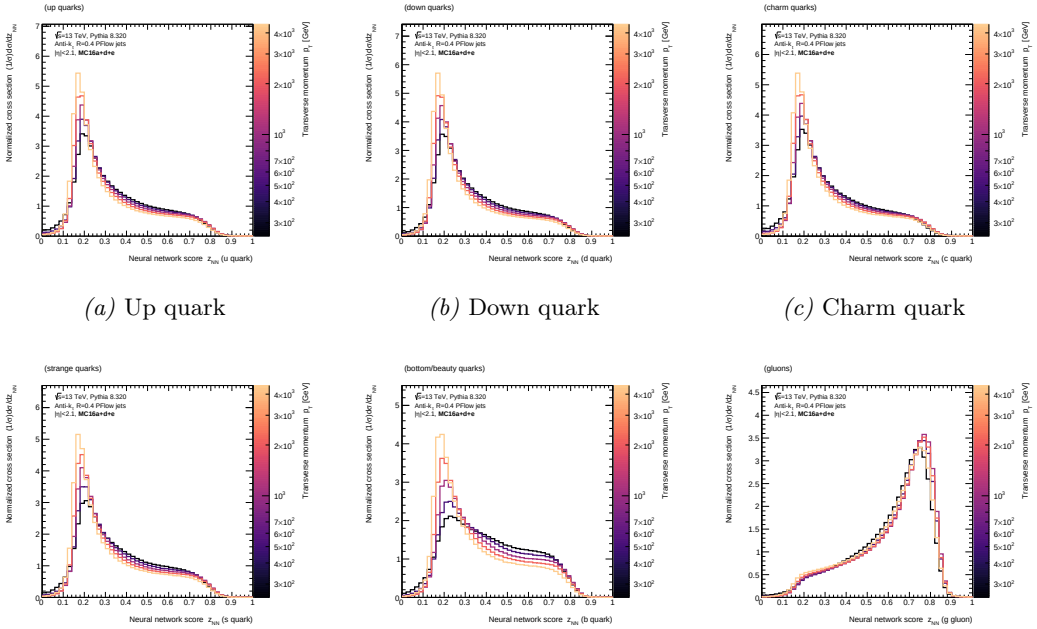
fundamentally not-separable on a jet-by-jet basis –; however, it does mean that the probability distributions with respect to the NN-score (or the response of the model) should be "similar" for all (light) quark flavors and different to the one corresponding to the gluon.

Figure 12.14 shows the NN-score as predicted by the PFN and EFN for the different quark-flavors (up, down, charm, strange and bottom/beauty) as well as the gluon. First of all, Figure 12.14a shows the total cross section (inclusive in p_{T}) with respect to the NN-score for all jet-flavors. The integral of the gluon-related distribution is dominant because it is a single category and multijet events are naturally enriched by gluon-initialized jets. Concerning the shape of the distributions, there is a significant separation between the quark- and gluon-like jets. As expected, jets originating from gluons tend to receive higher scores, while jets initialized by light quarks are more likely to get labeled by a lower NN-score. However, the representation in Figure 12.14a makes it very hard to compare the shapes of the individual quark flavors as their fraction to the total event yield differs by several orders of magnitude. To allow for a systematic comparison of the shape of the individual spectra for the different flavors, each distribution is normalized by its integral; the result is shown in Figure 12.14b. As can be seen, a clear difference between the different quark-related flavors can be observed. Especially, the distribution of the NN-score associated with the bottom quark differs with respect to the other, lighter quark types [15]. The response of the neural network is similar for the other quark flavors; only the charm quark shows some discrepancy, which is much less pronounced that the one related to the bottom quark though.

First of all, the result in Figure 12.14b is not worrisome. The majority of bottom-quark-initialized jets are still correctly classified as being quark-like jets. Furthermore, the overall fraction of bottom quarks in the jet samples is negligible compared to the other flavors; hence, the bias to the final quark/gluon fractions is very small and certainly covered by the overall precision of the measurement. And, lastly, the correlation of the jet substructure variable between the bottom quark and the gluon is much weaker for the counting-based observables introduced in Chapter 11; hence, a bias to the fractions obtained using the cross section with respect to the NN-score should be visible by comparing to the corresponding result derived using the counting-based observables.

Figure 12.15 shows the distribution of the NN-score of the various quark flavors (12.15a–

[15] It is not fully clear why the probability of misidentification is larger in case of b-quarks than for the other light quark flavors. Actually, one might "naïvely" expect the opposite behavior as jets initialized by b-quarks are (on average) more compressed (smaller width). This "problem" could be addressed by including b-tagging information into the classification decision.



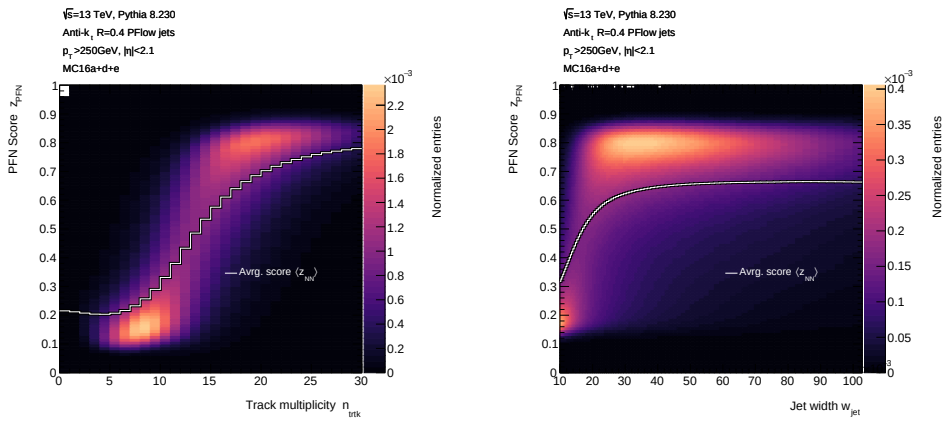
(a) Up quark *(b)* Down quark *(c)* Charm quark

(d) Strange quark *(e)* Bottom quark *(f)* Gluon

Plot 12.15: Inclusive (in η_{rel}) normalizes single-differential cross section with respect to the PFN-score for all relevant jet flavors parameterized as a function of transverse momentum of the jets.

12.15e) and the gluon (12.15f) for different bins of transverse momentum of the jets.

The last aspect studied in this section is the correlation of the score predicted by the neural network with the track multiplicity as well as the width of the jet, which will prove beneficial for the interpretation of the results at some point later. The correlation with respect to the PFN-score is shown in Figure 12.16 for the number of tracks assigned/matched to the jet (12.16a) as well as the width of the jet (12.16b). Of course, the NN-score is correlated to many



(a) NN-score *versus* n_{trk} *(b)* NN-score *versus* w_{jet}

Plot 12.16: Correlation between the NN-score as predicted by the PFN and the number of tracks (*left*) and the "width" of the jets (*right*).

more features; however, the track multiplicity as well as the jet width are highly correlated to the score (and weakly correlated with respect to each other, which reduces the amount of redundant information) and are, furthermore, among the most simplest ones. As one can

see, the NN-score is highly correlated with both jet substructure observables, which is – of course – expected since track multiplicity and jet width are both correlated to features related to quarks and gluons. In case of n_{trk}, jets with higher multiplicities are usually labeled with higher neural network scores, indicating that the jets are more likely to originate from gluons. For the jet width, narrower jets are more likely to receive lower NN-scores, which is expected because quark-initialized jets are smaller on average compared to gluons.

12.5 Detector-Level Distributions

With the two baseline models of the PFN and EFN being established, the newly defined observables are released into the rough wilderness and applied to real pp-collision data events recorded with the ATLAS detector, as well as events generated in Monte Carlo with the full detector simulation applied. At this point, it is worthwhile to highlight once again that both models have been trained on truth-level information solely (the reason for training the neural networks on particle instead of detector-level information was discussed in Section 12.2.1 in great detail). Up to this point, it was uncertain whether this choice would result in large discrepancies between the neural network score predicted by the models for the two different representations of the event, which would manifest itself in significant migrations resulting in a response matrix with a large fraction of non-diagonal elements. This concern, however, has turned out to be unfounded since inference works well at truth and reconstruction level as will be shown in this section.

The present section shows a selection of results of the triple-differential multijet cross section measurement with respect to the previously introduced NN-score *at detector level*. This includes the comparison of the shape of the respective JSS variable measured in data and simulation, as well as a closer look at the response matrix obtained for the nominal Monte Carlo event generator that is used to *simultaneously* unfold the multijet spectra to particle level, which will be the topic of the subsequent section. This comparison is done for all transverse momentum bins used in this analysis. Generally, all following results serve as a plausibility check to confirm that the reconstruction procedure works as expected. This is of particular importance as, for instance, slight differences in the preprocessing of the four-momenta of the jet's constituents between training and inference may result in considerable variations of the associated score. It is therefore absolutely necessary(!) to ensure that the training, testing and evaluation of the models are perfectly synchronized.

The following section shows the results for the detector-level distributions for the PFN – the corresponding results for the EFN are very similar.

12.5.1 PFN

As explained in Section 12.1, the PFN architecture according to Equation 12.1 implements a permutation invariant function over variable length sets, whereby the set is given by a collection of (charged) particles that constitute the reconstructed higher-level object, i.e., the jet. The first step is to measure the PFN's score assigned to each jet that passes the event selection according to Table 9.3 at *detector level* for both, data and simulation. The NN-score is measured over the full support $[0, 1]$ in several bins of transverse momentum as well as for jets more central and more forward in pseudorapidity (see Table 9.4). The distribution of the *continuous* NN-score is measured in a total number of 19 bins of non-equidistant width that is shown in Table 12.2. As it was the case for the track as well as SD multiplicity, the binning of the JSS observable is optimized in order to limit migrations between bins – resulting in more diagonal response matrices – while preserving as much shape information as possible for the extraction of "jet topics" as well as the quark-gluon fractions.

The measured number of inclusive multijet events in each individual phase space bin is shown in Figure 12.17 for central (12.17a) and forward jets (12.17b) for the full Run 2 data set and all Monte Carlo samples for the corresponding simulation campaigns. In this representation, the non-equidistant binning of the observables is, once again, not recognizable. Instead, as it was the case for the track and the SD multiplicity, the measured number of events is associated to a global measurement bin that uniquely (bijectively) maps the cross

Observable	Binning	Unit
$z_{\mathrm{PFN/EFN}}$	$[0.00, 0.04)$ $[0.04, 0.08)$ $[0.08, 0.10)$ $[0.10, 0.15)$ $[0.15, 0.20)$ $[0.20, 0.25)$ $[0.25, 0.30)$ $[0.30, 0.35)$ $[0.35, 0.40)$ $[0.50, 0.60)$ $[0.60, 0.65)$ $[0.65, 0.70)$ $[0.70, 0.75)$ $[0.75, 0.80)$ $[0.80, 0.85)$ $[0.85, 0.90)$ $[0.90, 0.93)$ $[0.93, 0.96)$ $[0.93, 1.00]$	none

Table 12.2: Binning of the NN-score z_{PFN} predicted by the PFN thought all cross section measurements.

section to an unique integer value (see Chapter 11.4.1 and Equation 11.18 in particular for further explanation). Each bin value in Figure 12.17a and 12.17b is normalized by the total "width", i.e., $\Delta p_{\mathrm{T}} \cdot \Delta y \cdot \Delta z_{\mathrm{PFN}}$ to correct for the "distortion" of the distribution due to the non-constant bin size.

Both recorded spectra in Figure 12.17 show the characteristic falling rate towards increasing transverse momenta. Upon closer inspection of the distribution of the NN-score, a strong dependence on the transverse momentum on the jets can be observed. This correlation is expected and intended as it reflects the changing flavor (light quarks and gluons) composition of multijet events in different regions in phase space. The different fractions of quark and gluon initialized jets also manifests itself in a clear difference between the spectra related to central and forward jets.

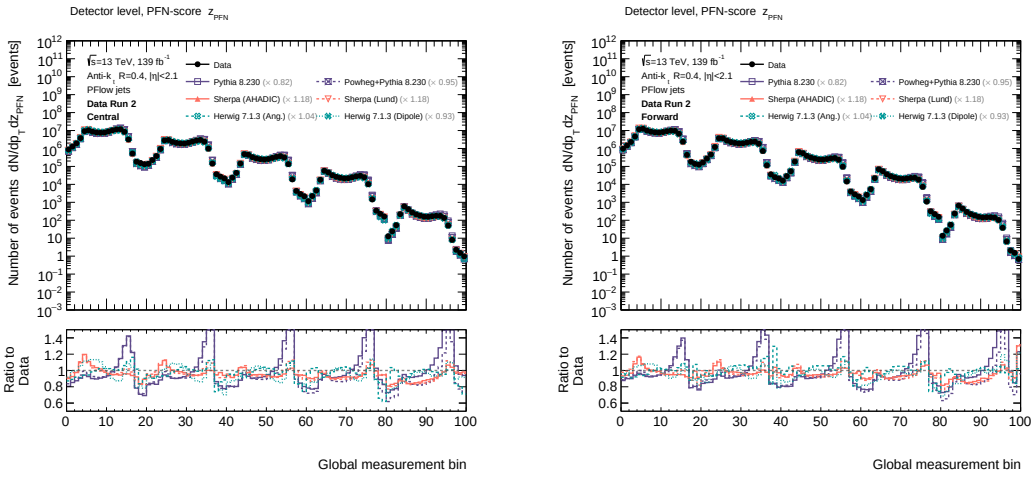
Overall, the simulation shows a good agreement with respect to the spectrum observed in data with deviations up to 50 % (and more) for certain regions in phase space. Furthermore, the degree of agreement varies significantly for the different Monte Carlo event generators because the neural network is sensitive to different physical effects like, for instance, parton shower or hadronization models.

Taking a closer look at the individual generators and their differences with respect to each other, the first property to notice is the close agreement of the two Pythia-based generators that differ regarding the calculation of the matrix element. In the lower-p_{T} regime, there is basically no difference concerning the predicted shape of the distributions. Only towards increasing transverse momenta a discrepancy can be observed.

Leaving aside the shape difference between the two samples, for both generators (the LO- as well as the NLO-based ones), the mismodeling with respect to data increases towards larger values of the NN-score ($z_{\mathrm{PFN}} \to 1$), which is associated with jets being classified as being more "gluon-like". This disagreement increases with p_{T}. The data-to-simulation agreement for phase space regions associated with quark-like jet ($z_{\mathrm{PFN}} \to 0$) does not change significantly over the whole range of transverse momentum, despite a noticeable mismodeling in the limit of low values of the score. Comparing the two (NLO) Herwig samples, the effect of different parton shower models can be studied. As expected, changing the parton shower model results in quite different shapes of the substructure variable as the angular-ordered and the dipole shower differ concerning the generated amount of emissions as well as the regions in phase space they populate. Therefore, the parton shower is expected to be one of the dominant uncertainties of the triple-differential multijet cross section measurement – as it was already the case for the counting-based observables.

Last but not least, the two spectra obtained from the Sherpa samples show the impact of different hadronization models on the predicted shape of the distributions. As one can see, the difference between the hadronization models is comparatively small. Compared to data, however, the degree of agreement changes significantly. For lower transverse momentum, both hadronization models tend to produce too many quark-like jets. As the score assigned by the neural network is highly correlated with number of jet constituents well as the jet width, this observation might be explained by a mismodeling (underestimation) of the number of hadrons in the final state, which results in a smaller number of constituents that are harder on average. For higher-p_{T} jets, the mismodeling shifts towards larger values of the NN-score, i.e., jets that are more likely to be initialized by gluons.

Figure 12.17 provides a valuable overview of the entire measurement, but the representation

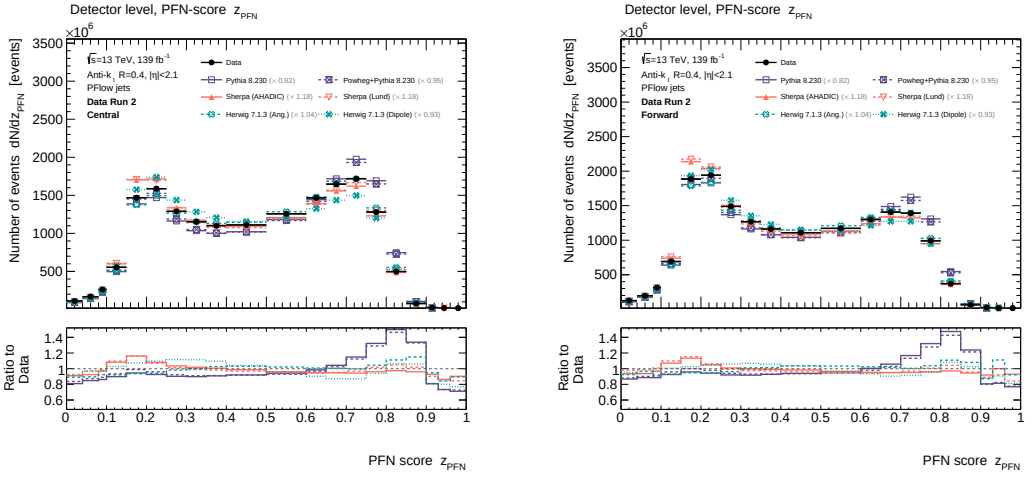


(a) Central jets *(b)* Forward jets

Plot 12.17: Full measurement of the number of multijet events in different regions in phase space with respect to the PFN-score for central (*left*) and forward (*right*) jets. The divided presentation of the multijet spectra for central and forward jets only serves the purpose of better visualization; indeed, the number of multijet events is *simultaneously* measured in both pseudorapidity regions and therefore accounts for migrations between the two.

based on the global bin indexing is fairly unsuited to gain insights concerning the distribution of the substructure variable z_{PFN}. For this purpose, Figure 12.18 shows the NN-score integrated over all transverse momentum bins for central (12.18a) and forward (12.18b) i.e., the information of the transverse momentum is marginalized. In this representation, the non-equidistant binning (summarized in Table 12.2) is explicitly visible.

Both distributions show the expected features. Figure 12.18a corresponds to the p_{T}-



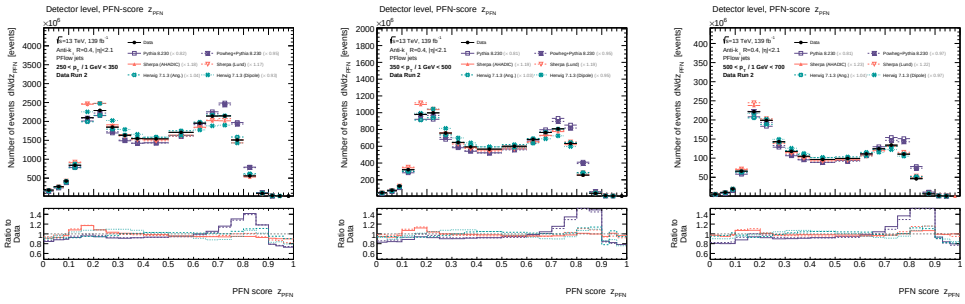
(a) Central jets *(b)* Forward jets

Plot 12.18: Distribution of the NN-score z_{PFN} for central (*left*) and forward (*right*) jets inclusive in transverse momentum.

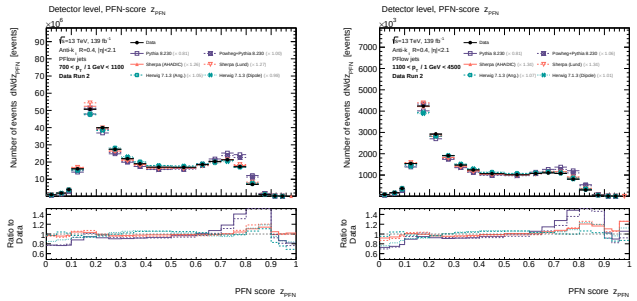
inclusive NN-score for *central* jets that clearly shows an enrichment of gluon-like jets as the bulk of the distribution is shifted towards higher values of z_{PFN}. In line with that, Figure 12.18b demonstrates that the category of *forward* jets is enhanced by jets being identified as

most likely initialized by quarks. On closer inspection, it is obvious that still a non-negligible fraction of quarks and gluons is present in both pseudorapidity categories. This is expected because quark and gluon jets can not perfectly be isolated into disjoint categories at detector level (not to mention the fact that the pseudorapidity is a comparatively weak discrimination variable with respect to features associated with hadronic showers). Undoubtedly, however, central and forward jets result in two orthogonal data sets with different *mixtures* of quark and gluon jets needed for the extraction of the underlying probability distributions of quark and gluon jets according to the method outlined in Chapter 7.

In case of the counting-based observables introduced in Chapter 11, the mean as well as the width of the probability distributions over the respective substructure observable(s) increased with the transverse momentum of the jets. This is expected as many counting-based observables can be approximated by a Poisson distribution at LL accuracy as was explicitly shown in Chapter 11.1.4 for the example of the SD multiplicity. In case of the NN-score, however, the evolution of the distribution with the transverse momentum of the jets is quite different. Generally, the change in the shape of the cross section should – to some extend – reflect the changing composition of quark- and gluon-initialized jets in the different regions of phase space, i.e., the different transverse momentum bins listed in Table 9.4. In order to gain an impression of the p_{T}-dependency of the multijet spectra, Figure 12.19 shows the single-differential spectra with respect to the NN-score as predicted by the PFN *inclusive* in central and forward jets for *all* transverse momentum bins of the measurement (see Table 9.4). As can be seen in Figure 12.19, compared to the counting-based observables, the



(a) $p_{\mathrm{T}} \in [250, 350)$ *(b)* $p_{\mathrm{T}} \in [350, 500)$ *(c)* $p_{\mathrm{T}} \in [500, 700)$



(d) $p_{\mathrm{T}} \in [700, 1100)$ *(e)* $p_{\mathrm{T}} \in [1100, 4500)$

Plot 12.19: Inclusive (in central and forward region) single-differential multijet distributions with respect to the NN-score as predicted by the PFN in all transverse momentum bins.

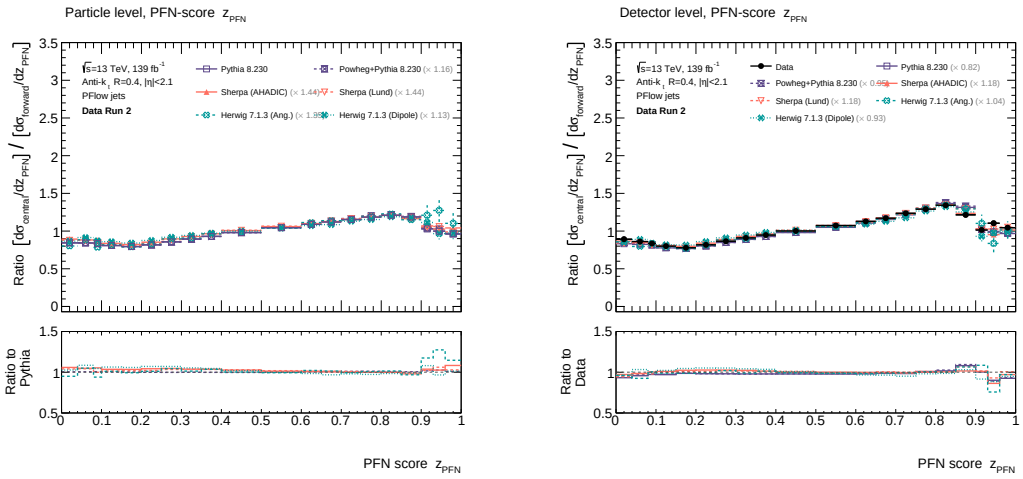
features of the distribution with respect to the NN-score *varies considerably* between different

[15] The labels in Figure 12.19 above are barely visible given the size and the number of figures, however, they are identical to the ones shown in Figure 12.18. Furthermore, this family of figures should only serve the purpose of providing an impression of the momentum-dependent evolution of the distributions.

p_{T} regimes. This is expected from Figure 9.1 that shows the fraction of quark and gluon jets as a function of the transverse momentum. Because of the NN-score being quite powerful regarding the discrimination between quark- and gluon-initialized showers, the z_{NN} distribution is very sensitive to changes with respect to variations in the quark-gluon composition of the different multijet samples. In the low-p_{T} regime in Figure 12.19, a clear shift towards higher NN-scores can be observed, indicating a higher fraction of jets originating from gluons in this region of phase space. However, as the transverse momentum of the jets increases, the "probability mass" of the NN-score shifts towards smaller values, which is associated with a higher probability of the respective showers being initialized by quarks. For jets in the very high-p_{T} regime, the mixture is basically dominated by quarks; extracting the "exact"' fraction of quarks and gluons in the respective phase space region using "jet topics" will be subject of the subsequent chapter.

With the extraction of the underlying quark and gluon distributions as well as the corresponding fractions in mind, the ratio between the measured cross section in central and forward region is of particular importance for the determination of the anchor bins. This cross section ratio between central and forward jets inclusive in transverse momentum is shown in Figure 12.20 for particle-level (12.20a) and detector-level (12.20b) inputs as it was done in case of the counting-based observables in Chapter 11 already.

Figure 12.20 shows a curve progression already familiar from the corresponding distributions of the counting-based observables: the ratio between central and forward cross section increase towards higher values of the NN-score[16], indicating an increased jet production rate in phase space regions associated with the gluon.



(a) Particle level *(b)* Detector level

Plot 12.20: Ratio between the spectra for central and forward jets at particle level (*left*) and detector level (*right*). In case of the particle level distributions, the ratio was taken with respect to the nominal sample (PYTHIA) as no data exists at truth level.

12.6 Unfolding of the Cross Section Measurement

After the triple-differential multijet cross section has been reconstructed with respect to the NN-score as predicted by the particle (and energy) flow network at detector level for simulation

[16]This similarity was one motivation for the definition of $z_{\mathrm{NN}} \to 1$ corresponding to gluon-like jets. Of course, the choice is completely arbitrary and does not affect the final result. Equivalently, the neural networks (EFN and PFN alike) could have just as well been trained such that $z_{\mathrm{NN}} \to 1(0)$ indicates the current jet to be initialized by a quark(gluon).

and real pp-collision data, the measurement must be unfolded to correct the spectra for detector and acceptance effects. With this goal in mind, this section basically follows the structure already used for the track multiplicity and the two configurations of the SD multiplicity in Chapter 11.5.

First and foremost (and for the sake of completeness), the first part of this section aims to very briefly show, once again, that the contributions from background sources (Drell-Yan) are negligible also in case of the NN-score given the degree of precision of the measurement and therefore do not need to be subtracted prior to the unfolding step. After that, the most important ingredients – according to Equation 5.9 – to the detector unfolding procedure, i.e., the actual response matrix that transform detector to particle level as well as the efficiency and fake factors that account for losses and fakes in the phase space volume, are presented as it was done in case of the track and SD multiplicity in the previous chapter.

Due to the very close correspondence of the PFN and the EFN, the technical aspects of the unfolding are analyzed only for the PFN, but the statements hold true for the EFN as well.

12.6.1 Background Processes

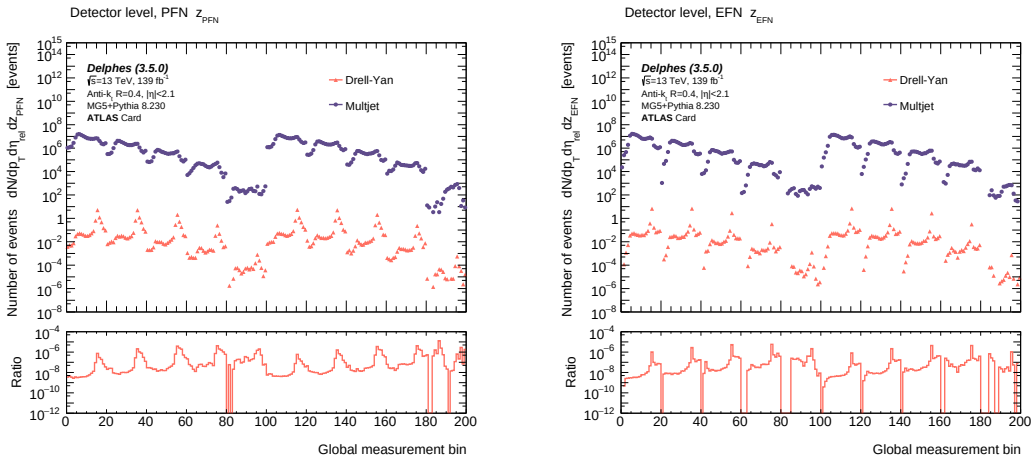
Like it was done in case of the counting-based observables in Chapter 11.5.1, this section is supposed to show that the contribution to the multijet cross section with respect to the NN-score due to the most dominant reducible background process, Drell-Yan, is completely negligible in all regions of phase space. The details of this superficial background-related study is not supposed to be repeated at this occasion (for a detailed explanation of all individual steps see Chapter 11.5.1).

In essence, a total number of $9.5 \cdot 10^6$ ($18.3 \cdot 10^6$) events for dijet (Drell-Yan) production has been generated at LO using MadGraph5_aMC@NLO, whereby the former one corresponds to the signal process while the latter gives rise to background due to the chance of misidentifying leptons as hadronic jets or those originating from hadronic tau decays. The matched tracks are used to reconstruct jets as described in Chapter 9.2.1. generator-level events are further processed via Pythia's interface for LHE files for subsequent parton showering (k_t-ordered) and following hadronization (Lund-string model); furthermore, Pythia's decay table is used to simulate the decay to final state particles. The set of final state particles is then processes by a simplified simulation of the whole ATLAS detector based on *Delphes*' fast detector simulation. The last step, i.e., the simulation of an imperfect measuring apparatus is essential for the estimation of the irreducible background, which arises due to the possibility of falsely identifying electromagnetic as hadonic showers (of course, there are also hadronic showers due to the hadronic decay modes of the tau).

After the detector simulation step, calorimeter-clusters and inner-detector tracks are used to reconstruct jets in the same vein as described in Chapter 9.2. Based on the set of tracks matched to the calorimeter-based jet (according to the matching scheme introduced in Chapter 9.2.1), the inputs to the neural networks are prepared in line with the procedure explained in Section 12.2. To obtain the NN-score (as predicted by the PFN as well as the EFN) for the simulated set of multijet and Drell-Yan events, the exact same models like in Section 12.5 are used to obtain the corresponding detector-level spectra.

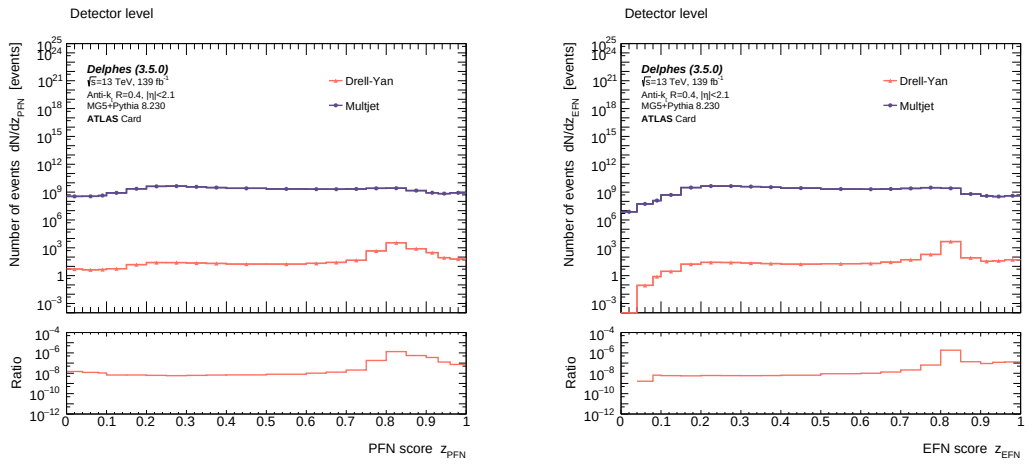
The result of the reconstructed triple-differential cross section at detector level for multijet events and Drell-Yan production is shown in Figure 12.21. Both distributions in Figure 12.21 have been correctly reweighted according to the weighting scheme described in Chapter 11.5.1 to account for the relative production rate between multijet and Drell-Yan events. As can be concluded by inspecting the ratio that shows the relative contribution of the *most dominant* background process to the signal, the contamination of background events is below 1‰ for all regions in phase space considered and therefore completely negligible given the overall degree of precision of the measurement.

To better analyze the different features of the NN-score for PFN and EFN, respectively, for multijet and Drell-Yan events, Figure 12.22 shows the single-differential distributions for both processes, which is obtained by marginalizing the transverse momentum as well as the pseudorapidity in the joint distributions in Figure 12.21. The reader is, once again, reminded of the fact that the distributions in Figure 12.22 are dominated by the low-p_{T} region due to the high



(a) PFN *(b)* EFN

Plot 12.21: Triple-differential cross section with respect to the NN-score as predicted by the PFN (*left*) and the EFN (*right*) for multijet and Drell-Yan production.

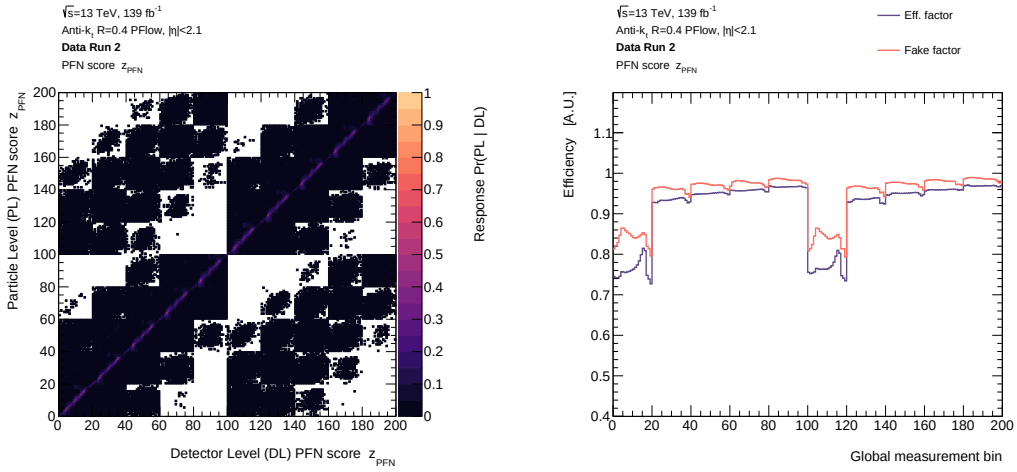


(a) PFN *(b)* EFN

Plot 12.22: Single-differential cross section with respect to the PFN (*left*) and the EFN (*right*) for multijet events and for the background due to Drell-Yan.

production rate of jets with lower transverse momenta. The multijet-related distributions look exactly as expected from the corresponding spectra in Section 12.5. Interestingly, both neural networks (PFN and EFN alike) tend to classify fake jets due to misidentified electromagnetic showers as gluon jets – while there is also a significant fraction of Drell-Yan-induced fake jets classified as being quark-like. This is interesting since the corresponding background-related study of the counting-based substructure observables showed that fake jets usually feature a small number of constituents/emissions (compare, for instance, with Figure 11.23), which is expected given the fact that they are induced by electromagnetic showers. Since the NN-score is highly correlated with the number of tracks inside of the jet (see Figure 12.16), one naively might expect an increased fraction of quark-like jets in case of Drell-Yan events. However, the NN-score is correlated to many "latent" variables besides the number of constituents like, e.g. the width of the shower to name just one out of many.

12.6.2 PFN

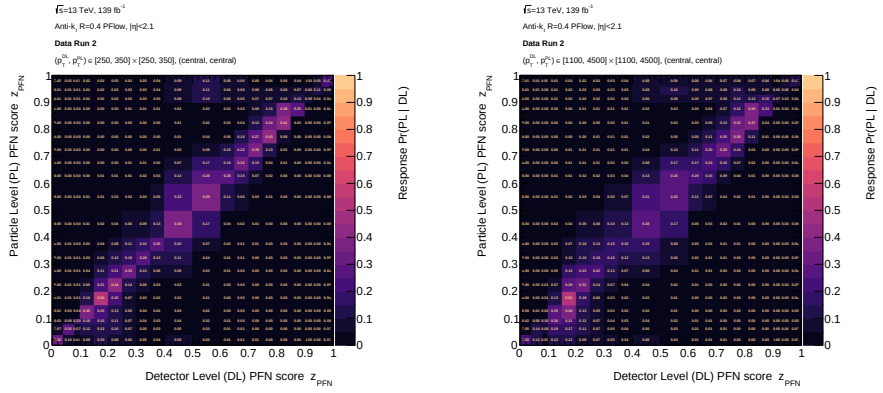


(a) Response matrix *(b)* Efficiency and fake factors

Plot 12.23: Full response matrix for the PFN (*left*) and the efficiency and fake factors (*right*).

As mentioned before, the first distributions that are considered in this section are the response matrix as well as the efficiency and fake factors for the PFN. The full response matrix (i.e., for all phase space bins of the measurement) is shown in Figure 12.23a. The two-dimensional distributions exhibits a structure that is already familiar from the corresponding figures of the track and SD multiplicity in Chapter 11.5, i.e., most jets at particle and detector level being reconstructed in the same phase space bin (no migrations). As explained previously, the asymmetry in case of the (first) off-diagonal elements is explained by the increased probability of the jet's transverse momentum being underestimated due to the steeply falling multijet cross section with the p_{T} of the jet.

Figure 12.24b shows the response matrices in the central pseudorapidity category (corresponding to the first "quadrant" in Figure 12.23a) for the PFN with the non-equidistant binning being explicitly visible. After the response matrix has been applied to the detector-level



(a) $p_{\mathrm{T}} \in [250, 350)$ *(b)* $p_{\mathrm{T}} \in [1100, 4500)$

Plot 12.24: Individual response matrices for different p_{T} regimes in central region for the PFN-score.

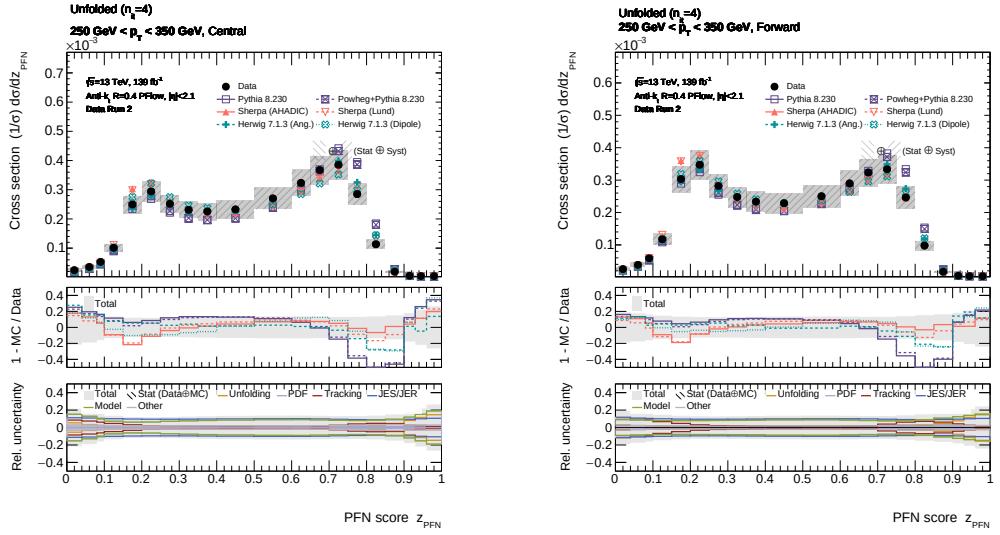
spectra, the triple-differential cross section must still be corrected for efficiency losses and fake jets; this is accounted for by the efficiency and fake factors shown in Figure 12.23b. Since the

efficiency and fake factors have already been described in great detail in Chapter 11.5 and the results basically corresponds to the ones obtained using the counting-based observables, there is no need to repeat the explanation at this point. Suffice it to say that the significant efficiency drop in the very first p_{T} bin for both pseudorapidity categories is a consequence of the steeply falling p_{T} spectrum and the fact of no lower momentum bin being present for the jets to migrate into.

12.7 Particle-Level Distributions

After the technical aspects of the unfolding step have been discussed, this section presents the unfolded results of the triple-differential multijet cross section with respect to the NN-score as predicted by the PFN including all relevant sources of systematic as well as statistical uncertainties. Furthermore, some derived quantities like the average score are considered, similarly to what was done in case of the track and SD multiplicity in Chapter 11.6. The unfolded result of the cross section measurement given in this section also serves as the basic input to the "jet topic" algorithm that is used in the upcoming chapter to extract the underlying quark/gluon fractions as well as the underlying probability distributions of the respective jet features.

12.7.1 PFN [17,18]



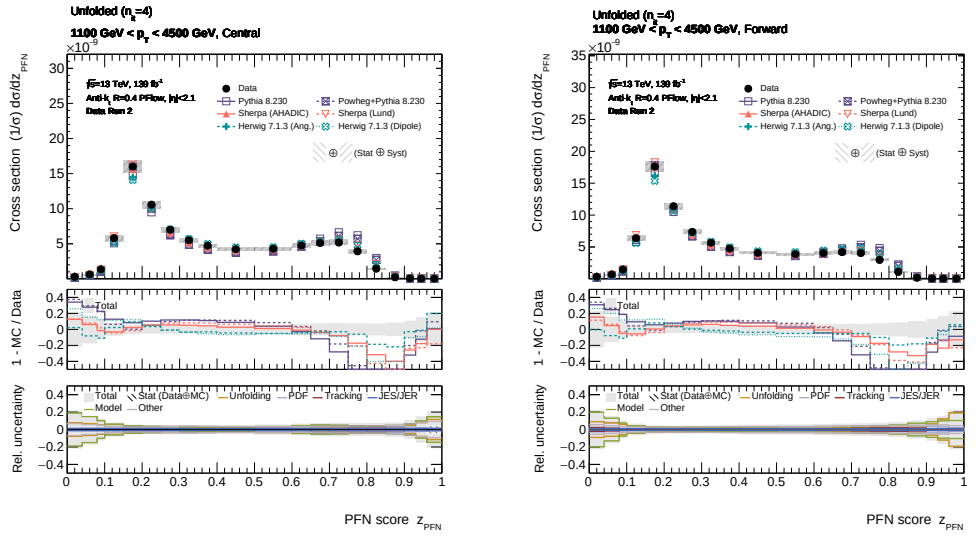
(a) Central *(b)* Forward

Plot 12.25: Individual single-differential multijet cross sections in bins of the PFN-score z_{PFN} for central (*left*) and forward (*right*) jets with $250\,\mathrm{GeV} \leq p_{\mathrm{T}} < 350\,\mathrm{GeV}$.

Before the full result of the unfolded triple-differential multijet cross section is presented, the distribution of the NN-score predicted by the PFN is analyzed in two bins of transverse momentum in central and forward jets with the non-equidistant binning of the score being explicitly visible to gain an intuition for the individual sources of systematic and statistical uncertainties for this observable. The unfolded single-differential cross sections for jets with a transverse momentum $250\,\mathrm{GeV} \leq p_{\mathrm{T}} < 350\,\mathrm{GeV}$ are shown in Figure 12.25 for central (12.25a) and forward (12.25b) jets. The two examples nicely illustrate the change of the shape of the

[17] The corresponding numerical values of the relative triple-differential multijet cross section with respect to this observable are provided in Table B.5 in the Appendix of this report.
[18] The results fro the EFN are very similar.

differential cross section as a function of the jet's transverse momentum: with increasing p_{T}, the "probability mass" of the distribution shifts towards smaller scores, indicating an increased fraction of quark-initialized jets as expected from the underlying PDF of the proton. This result is consistent with the expectations from the simulated multijet cross section and its composition of quarks and gluons in Chapter 9.1 as well as the insights gained from the detector-level distributions in the penultimate section. Regarding the systematic and statistical uncertainties, the measurement is – as expected – dominated by the modeling uncertainties as well as the uncertainty imposed by the jet energy scale calibration like it was the case for the counting-based observables. As explained in Chapter 10, the modeling uncertainty consists of three components, i.e., matrix element (ME), parton shower (PS) and hadronization (Had). The variation of the reconstructed spectra due to higher-order corrections of the perturbative calculation of the underlying hard (sub-)process is completely negligible because the internal structure of jets is mostly formed by the emission of additional radiation during the subsequent parton shower step. Also, the efficiency of the dijet-selection criteria does not vary significantly between the LO and NLO samples. This would change if at least three jets would be required to be present in the final state; in this case, the higher-order correction should indeed play an important role. As expected from the previous statements, the parton-shower-related uncertainty is among to most dominant ones for almost all regions in phase space over which the triple-differential cross section is measured. Furthermore, the degree of uncertainty due to the different parton shower models varies significantly between different regions in transverse momentum. The observation of the PS-related uncertainty being dominant is completely expected because the parton shower basically determines the internal structure of jets; hence, most substructure observables are *highly* sensitive to the variation of the underlying model that is used to simulate the emission of additional radiation in the jet. The strong dependency of the uncertainty immediately follows from the underlying premises of the parton shower philosophy explained in Chapter 2.4. With increasing transverse momentum of the jets, the average k_t of



(a) Central *(b)* Forward

Plot 12.26: Individual single-differential multijet cross sections in bins of the PFN-score z_{PFN} for central (*left*) and forward (*right*) jets with $1100\,\mathrm{GeV} \leq p_{\mathrm{T}} < 4500\,\mathrm{GeV}$.

the emissions increases (harder radiation); furthermore, the collimation angle of the radiation decreases too, i.e., the emissions are more collinear with respect to the jet's trunk. This implies that higher-p_{T} jets are closer to meet the underlying assumption of the parton shower approach of the generated emissions being collinear. Additionally, in case of collinear radiation (as it is the case for high-p_{T} jets), the different ordering schemes according to the evolution variable approach each other such that the difference between the various parton shower models reduces

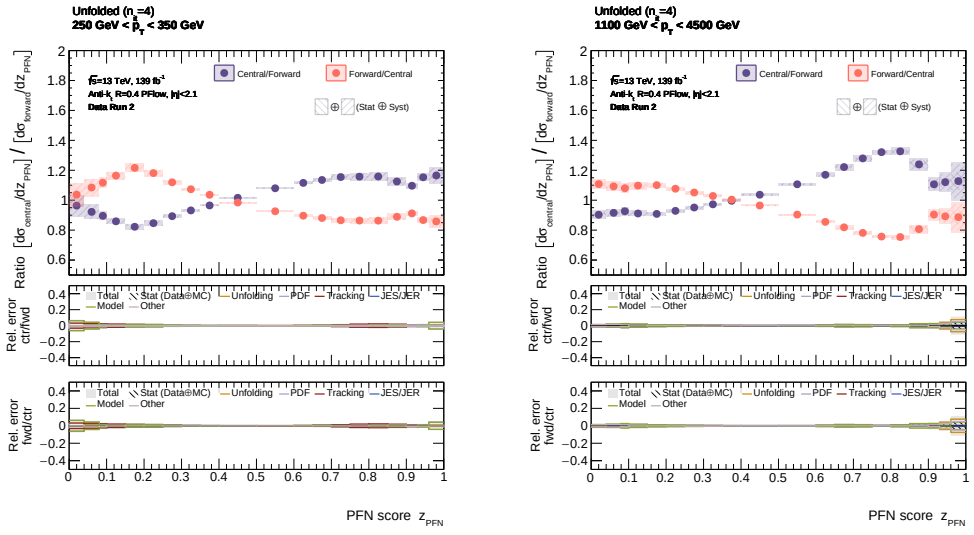
for high-p_{T} jets. The hadronization-related uncertainty fades compared to the parton shower one – at least in the low-p_{T} regime. Generally, the hadronization-related uncertainty is less dependent of the jet's transverse momentum. Because the PS-related uncertainty decreases with increasing transverse momentum p_{T}, the uncertainty due to different hadronization models becomes a non-negligible contribution to the total uncertainty of the cross section for higher-p_{T} jets. Regarding other sources of uncertainties, the statistical uncertainty is completely negligible for almost all phase space regions. Only in regions associated with a very small ($z_{\mathrm{NN}} \to 0$) or very large ($z_{\mathrm{NN}} \to 1$) NN-scores for low- and high-p_{T} jets shows some lack of statistics in both, data and simulation.

Contrary to the counting-based observables, the tracking-related uncertainties are rather small. In case of the track/SD multiplicty, the tracking-related variations resulted in one of the dominant components of the total uncertainty, which is expected as the counting-based observables are obviously very sensitive to variations in the number of tracks. The neural network on the other hand are much more stable in this regard. This can already be seen in Figure 12.8b that shows the AUC as a function of the upper limit of the number of charged constituents/tracks used as an input to the model.

The corresponding single-differential multijet cross sections in *central* and *forward* region for jets with $1100\,\mathrm{GeV} < p_{\mathrm{T}} < 4500\,\mathrm{GeV}$ are shown in Figure 12.26a–12.26b. Directly comparing the cross sections measured in the same p_{T} bin in central (12.25a/12.26a) and forward (12.25b/12.26b) region shows an increase in the fraction of quark-initialized jets. This simply reflects the fact of the forward region being statistically enriched by quark-initialized jets. Keep in mind that the "non-mutual information" (of quarks and gluons) encoded in the central and forward cross section are the basic information from which the underlying distributions as well as the quark/gluon fractions are derived.

According to the statements in Chapter 7, the anchor regions are determined based on the (likelihood-)ratio between the distributions in central and forward regions, respectively. This ratio is shown in Figure 12.27 for the two bins in transverse momentum in Figure 12.27.

The ratios show the desired behavior, i.e., they (mostly) monotonically increase – at least



(a) $p_{\mathrm{T}} \in [250, 350)$ *(b)* $p_{\mathrm{T}} \in [1100, 4500)$

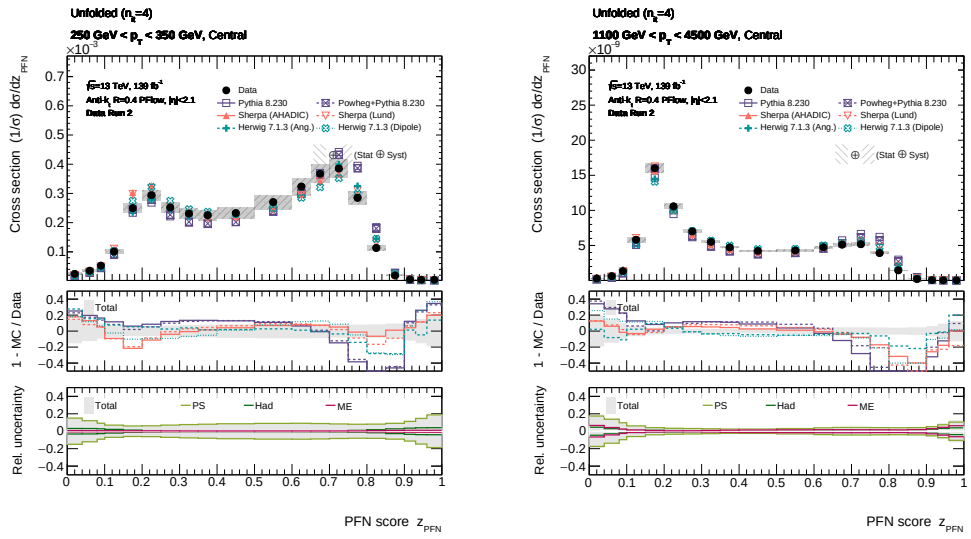
Plot 12.27: Individual ratios of the single-differential multijet cross section in central and forward region with respect to the NN-score according to the PFN for different transverse momentum regimes (in GeV).

in the regions relevant for the determination of the anchor bins/regions – for higher NN-scores, indicating that the corresponding phase space region is getting purer in gluon-like jets. The uncertainties of the cross section ratio are calculated assuming the systematic uncertainties of

the central and forward category being fully correlated, resulting in a high degree of cancellation and therefore overall small uncertainties associated with the ratio – the statistical uncertainties are, of course, taken to be uncorrelated.

Figure 12.28 shows the contribution of the modeling-related uncertainty to the total uncertainty of the multijet cross section. As explained previously, the uncertainty related to variations due to the underlying parton shower model is a dominant source of uncertainty of this measurement for almost all regions in phase space. Furthermore, a considerable reduction of the PS-related uncertainty for higher-momentum jets can be observed as can be seen by directly comparing Figure 12.28a and Figure 12.28b.

The next uncertainty to consider is the uncertainty related to tracks. In case of the SD and



(a) $p_T \in [250, 350)$ *(b)* $p_T \in [1100, 4500)$

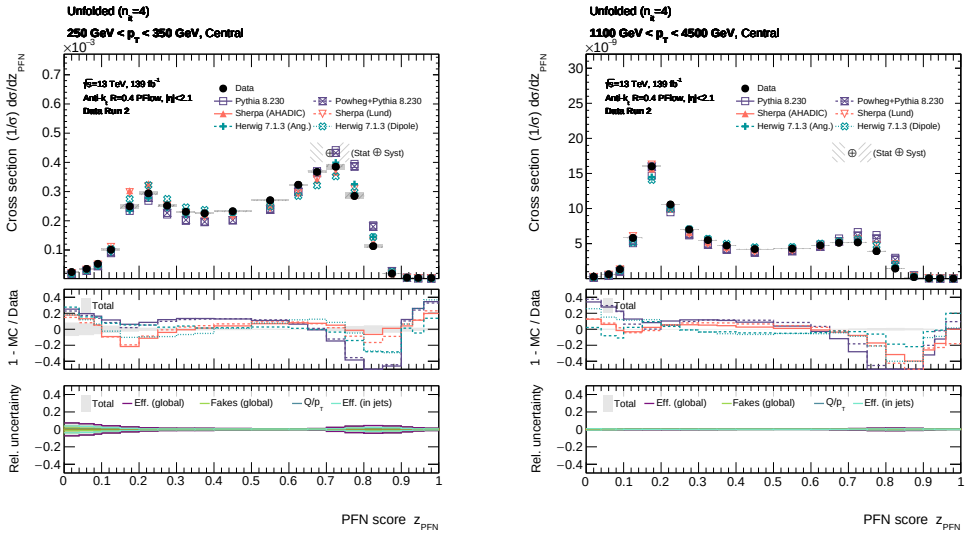
Plot 12.28: Modeling-related uncertainty for two individual single-differential multijet cross sections in bins of the PFN-score z_{PFN} for different p_T (in GeV) regimes in central region.

track multiplicity, the tracking-related uncertainty was – along with the uncertainty due to the parton shower model as well as the calibration of the jet energy scale – among the dominant sources of systeamtic uncertainties, which is expected given the fact that both jet substructure observables basically "count" tracks. The NN-score, on the other hand, turns out to be more robust concerning tracking-related variations as explained previously.

After the unfolded single-differential cross section has been analyzed for a few examples, it is about time to turn the focus of attention to the full, i.e., all phase space bins, measurement of the *triple*-differential multijet cross section with respect to the PFN-score z_{PFN}. Figure 12.30 shows the unfolded triple-differential multijet cross section between each phase space bin in terms of the global bin representation as well as the systematic and statistical uncertainties in all bins of the measurement. The full picture of the measurement basically confirms all insights gained from the single-differential distributions in Figure 12.25 and Figure 12.26. The total uncertainty of the measurement is obtained by assuming all individual uncertainties (listed in Figure 12.30) to be *uncorrelated* and taking the square root of the sum of the squares of all uncertainties. The resulting total uncertainty is basically determined by the choice of the parton shower model as well as the JES/JER uncertainty.

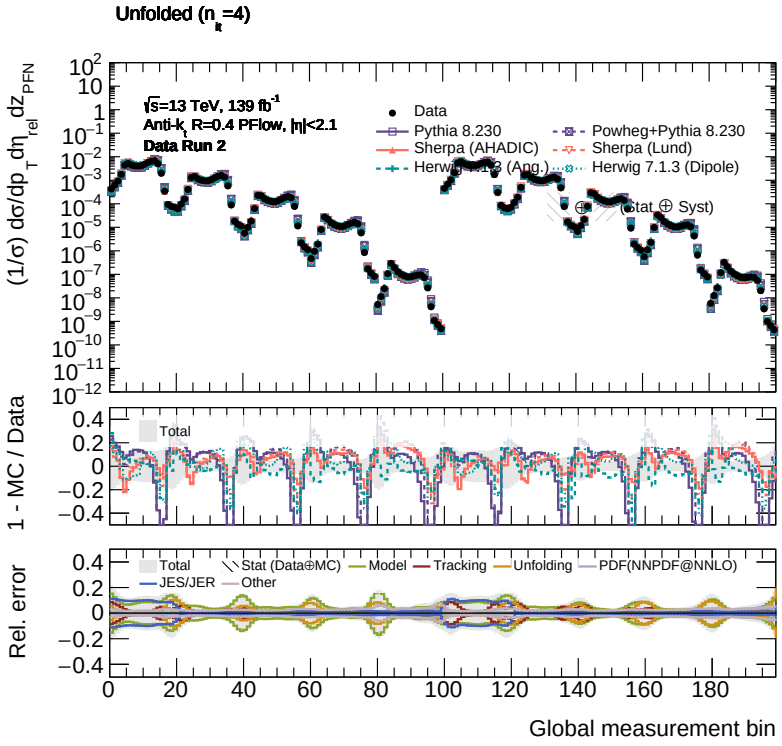
For better visualization, Figure 12.31 shows a summary of the uncertainty (statistical and systematic) of the triple-differential multijet cross section with respect to the NN-score predicted by the PFN.

In conclusion, the degree of uncertainty of the jet substructure measurement is in the order of roughly 10 % over the entire phase space. This degree of precision is expected in case of the

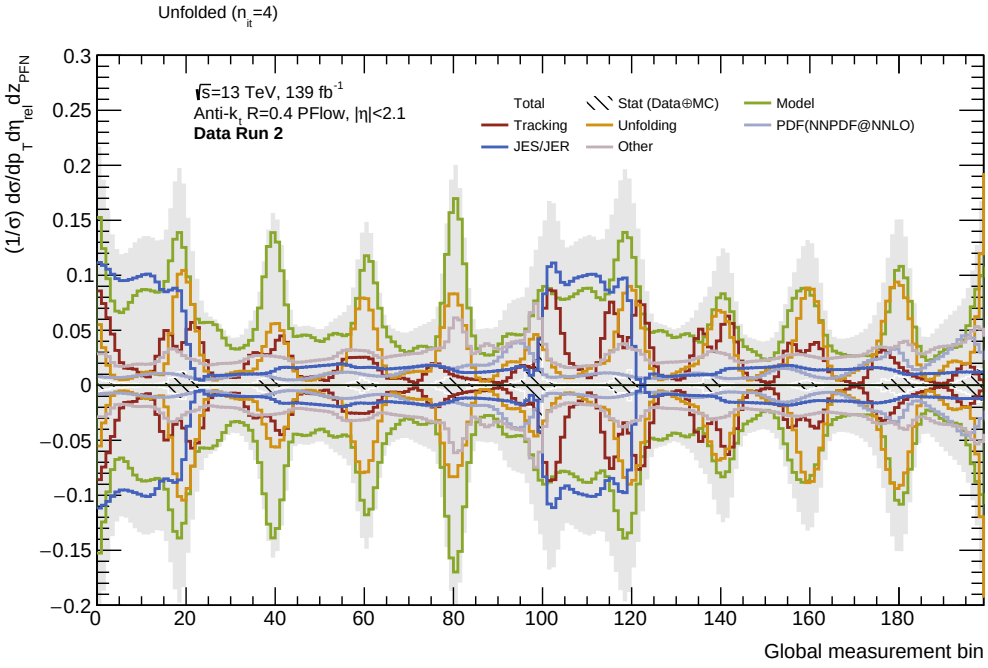


(a) $p_{\mathrm{T}} \in [250, 350)$ *(b)* $p_{\mathrm{T}} \in [1100, 4500)$

Plot 12.29: Tracking-related scale uncertainty for two individual single-differential multijet cross sections in bins of the PFN-score z_{PFN} for different p_{T} (in GeV) regimes in central region.



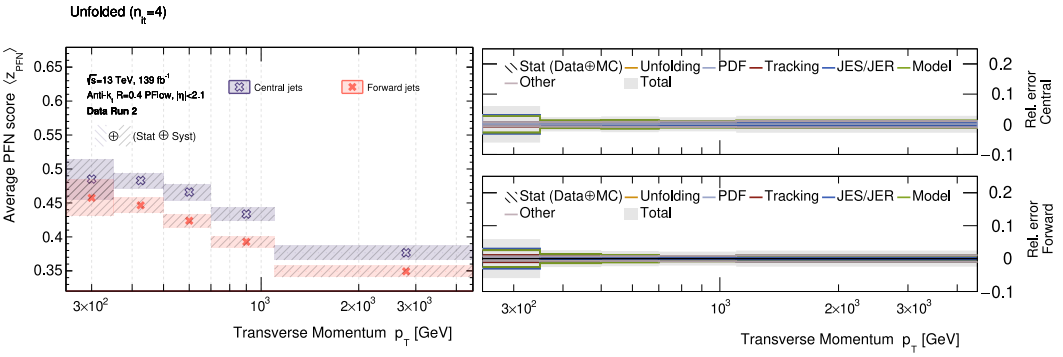
Plot 12.30: Full measurement of the unfolded triple-differential multijet cross section as a function of the global measurement bin $(p_{\mathrm{T}}, \eta_{\mathrm{rel}}, z_{\mathrm{PFN}})$ at particle level including all sources of systematic and statistical uncertainties.



Plot 12.31: All sources of systematic and statistical uncertainties of the triple-differential multijet cross section measurement as a function of the PFN-score.

measurement of a jet substructure observable and consistent with the results obtained in case of the counting-based observables.

The unfolded cross section can be used to derive some further quantities like, e.g., the average score as a function of the transverse momentum – like it was done in case of the counting-based observables in Chapter 11. In case of the counting-based observables, the average/mean of the probability distribution (which is given by the respective cross section with respect to n_{trk} or a configuration n_{SD}) reflects the growing number of constituents and emissions inside of a jet with increasing transverse momentum. In case of the NN-score (PFN as well as EFN), the interpretation is somewhat different; however, before elucidating the "evolution" of the average NN-score as a function of p_{T}, the reader is supposed to take a look at Figure 12.32, which shows the just-mentioned relation between the average score $\langle z_{\text{NN}} \rangle$ (the statements are valid for the PFN as well as the EFN alike) and the transverse momentum p_{T}. The first feature



Plot 12.32: Average of the NN-score as predicted by the PFN as a function of the jet's transverse momentum (in GeV) including all sources of uncertainties.

that immediately catches the eye is the close correspondence of the average NN-score in Figure 12.32 with the curve progression of the quark/gluon fractions in Figure 9.1 according to the jet-flavor definition based on the PDGID of the hardest constituent inside of the jet. This close correlation can be understood recalling that the neural network has been trained (in a supervised fashion) to implement a transformation of the input data such that quarks and gluons are maximally separated in the transformed representation. This representation might be quite complicated, however – by construction –, neural network scores with $z_{\mathrm{NN}} \to 0$ and $z_{\mathrm{NN}} \to 1$ will result in highly quark- and gluon-enriched regions of phase space, respectively. Furthermore, following the basic assumption from Chapter 7 (assuming that the measured distributions in multijet events are a weighted combination of the underlying distributions belonging to quark and gluon jets according to $P_{M_\alpha}(z_{\mathrm{NN}}) = f_q^{M_\alpha} P_q(z_{\mathrm{NN}}) + f_g^{M_\alpha} P_g(z_{\mathrm{NN}})$, whereby M_α with $\alpha \in \{\mathrm{central}, \mathrm{forward}\}$ denoting the two mixtures of jets), the average NN-score over jet mixture α is given by

$$\begin{aligned}
\langle z_{\mathrm{NN}} \rangle_{M_\alpha} &= \int_{[0,1]} \mathrm{d}z_{\mathrm{NN}}\, z_{\mathrm{NN}} P_{M_\alpha}(z_{\mathrm{NN}}) && (12.19)\\
&= \int_{[0,1]} \mathrm{d}z_{\mathrm{NN}}\, \left\{ f_q^{M_\alpha} P_q(z_{\mathrm{NN}}) + f_g^{M_\alpha} P_g(z_{\mathrm{NN}}) \right\} && (12.20)\\
&= f_q^{M_\alpha} \int_{[0,1]} \mathrm{d}z_{\mathrm{NN}}\, P_q(z_{\mathrm{NN}}) + f_g^{M_\alpha} \int_{[0,1]} \mathrm{d}z_{\mathrm{NN}}\, P_g(z_{\mathrm{NN}}) && (12.21)\\
&= f_q^{M_\alpha} \langle z_{\mathrm{NN}} \rangle_q + f_g^{M_\alpha} \langle z_{\mathrm{NN}} \rangle_g, && (12.22)
\end{aligned}$$

with $P_{M_\alpha}(z_{\mathrm{NN}})$ being the probability distribution in central or forward region. Ideally (i.e., a perfect classifier), the mean score of quarks and gluons is given by $\langle z_{\mathrm{NN}} \rangle_q \approx 0$ and $\langle z_{\mathrm{NN}} \rangle_g \approx 1$. In this case, Equation 12.22 reduces to $\langle z_{\mathrm{NN}} \rangle_{M_\alpha} \approx f_g^{M_\alpha}$, i.e., the average NN-score in Figure 12.32 closely corresponds to the gluon fraction in the respective region in phase space – if the network would have been trained such that $z_{\mathrm{NN}} \to 1$ would correspond to quark-like jets, the average score would correspond to the quark fraction $\langle z_{\mathrm{NN}} \rangle_{M_\alpha} \approx f_q^{M_\alpha}$. Of course, this is only an approximation (because the learned representation does not result in a perfect separation of quarks and gluons), and the average score is at most highly correlated with the fraction of gluon jets. To obtain the actual fraction of quark/gluon jets, the underlying jet features must be extracted using the "jet topic" algorithm explained in Chapter 7, which will be the subject of the next chapter.

13. Disentangle Quark and Gluon Jets

A long journey is about to approach its end ... But, this is certainly not the time to get maudlin. Shortly before the finishing line, it is appropriate to pause for a brief moment of time, turn around and cast a brief look at most important stations of the expedition that is behind us. This provides some little time to catch one's breath in order to prepare for the last stage of the trip, that is, the extraction of the underlying quark/gluon distributions/themes for all jet substructure observables (counting- and NN-based) as well as the fraction of quarks/gluons in the different transverse momentum bins of the measurement. But, first a quick review of what have been done so far.

Chapter 11 and Chapter 12 of this report form a complete *self-contained* analysis on its own. The triple-differential cross section in multijet events was measured with respect to several jet substructure variables that are supposed to provide "sufficient" sensitivity to the quark- and gluon-like characteristics of the hadronic shower that gave rise to the jet (quark-gluon discrimination variable). The class of jet substructure observables that are investigated in this thesis can be broadly subdivided into two categories: *firstly*, a set of counting-based jet shapes like the track and SD multiplicity (with the latter one counting the number of valid SD emissions in a C/A reclustered jet); and, *secondly*, a score assigned on a jet-by-jet basis predicted by neural networks based on the "Deep Sets" architecture that allows to implement permutation invariant functions over sets. The choice of those two classes of observables is by no means arbitrary – on the contrary: the underlying motivation was to obtain one class of observables featuring a simple interpretation yet decent discrimination power that is not limited by Casimir scaling, while the second class of multivariate observables focuses on the aspect of classification performance with the waiver of transparency on the model's decision-making process.

For each jet substructure observable, the total number of multijet events is measured in several bins of transverse momentum p_{T}, two bins of (relative) pseudorapidity η_{rel} as well as one of the aforementioned jet shapes $\mathcal{O}_{\mathrm{JSS}}$. The measurement is normalized by the total number of events to cancel systematic uncertainties to the degree of their correlation, and the relative event rate is simultaneously unfolded to hadron level to correct for detector and acceptance effects.

Several sources of systematic uncertainties are taken into account such as, e.g., modeling-related uncertainties accounting for higher-order corrections in the calculation of the hard interaction, the utilization of different parton shower and hadronization models; the uncertainty on the jet energy scale due to the calibration of the jet energy and its resolution; the variation of the simulation's prediction due to a different choices of the proton's PDF set; the simulation of different pile-up conditions, uncertainties due to tracking-related variations etc. In a nutshell, it was shown that the precision of the measurement for almost *all* regions in phase space is determined by the parton-shower-related uncertainty as well as the limited knowledge of the jet energy scale and – in case of the counting-based observables – the tracking-related uncertainties. The total degree of uncertainty of the cross section measurement is roughly between $10 - 20\,\%$ in the low- and intermediate- to high-p_{T} regime.

As mentioned before, the results presented to this point represents a fully self-contained analysis with plenty of applications already – especially in the context of the tuning of Monte Carlo event generators. Especially the sensitivity of the cross section measurement to gluon-like jets is charming in this context since, as previously stated, most of the state-of-the-art

generators have been extensively tuned to quark-enriched samples from LEP data but not on gluons – at least not to the same extend. It should be noted (to not create false expectations), however, that the distributions measured do *not* solve the issue because the samples (over which the cross section was measured) are sill mixtures of *both*, quark and gluon jets. But, the distributions can still be used to improve the data-to-MC agreement within certain regions of phase enriched by gluon-like jets.

With the triple-differential multijet cross section being measured for all jet substructure observables, the next (and final) step is to extract the underlying probability distributions of quark and gluon jets as well as their corresponding fractions in the given mixture of hadronic jets. This problem is ill-defined ... *unless* ... the jet substructure observables fulfill a certain mathematical property known as mutual irreducibility (see Chapter 7.3 for a quick repetition of this property in the context of QCD), i.e., it is not possible to write the underlying quark and gluon distributions as a non-trivial mixture of each other including some (arbitrary) other distribution. All observables in this analysis are chosen in light of this criterion. This will be done for *all* observables. In general, the motivation of the various cross section measurement presented in this report was to obtain "independent" measurements of the quark/gluon fractions in order to verify whether they are consistent with each other or not[1].

In this chapter, the "jet topics" (the underlying quark/gluon "themes"/distributions) for each jet substructure observable (n_{trk}, n_{SD} and its two configurations; z_{PFN} and z_{EFN}) are extracted using the method(s) introduced in Chapter 7. Some examples of the probability distributions corresponding to quark- and gluon-like jets are presented; this corresponds to the interpretation of quark- and gluon-like jets on distribution level according to the operational definition of quark and gluon jets, whereby the exact composition is unknown up to this point. Along with the distributions, the fraction of quark and gluon jets and their corresponding uncertainties are calculated and quoted for each jet substructure variable individually.

13.1 Preparatory Work

The first part of this chapter provides some introductory work that is indispensable for the further course of the discussion. As explained previously, the purpose of this chapter is to extract the underlying probability distributions (in line with the operational definition of quark and gluon jets) as well as the quark/gluon fractions in each bin of transverse momentum. While the underlying quark and gluon distributions are – turning a blind eye to the issue of sample-dependence of jets – are universal, the extracted fractions of quarks and gluons are, of course, associated to a *specific*, well defined mixture of jets. This implies that the quark/gluon fractions finally stated in this thesis are *specific* to the event selection an the resulting phase space volume (see Table 9.3) over which the triple-differential multijet cross section is measured. This means, in particular, that the extracted fractions are *not* transferable to other event selections. This means, the analysis presented in this thesis must be repeated entirely if one needs the quark/gluon fractions for a different selection of multijet events[2].

13.1.1 Extraction of the Anchor Region

An important aspect is the determination of the anchor bins (or regions), respectively, that are used to extract the underlying quark and gluon probability distributions as well as their fractions, given a certain bin in transverse momentum. In principle, the reducibility factors are determined based on the ratio of the *unfolded* single-differential multijet cross sections in central and forward region according to the prescription described in Chapter 7. However, the "naive" application of Equation 7.10 to obtain the reducibility factors results in a problem concerning the extracted histograms and their interpretation in terms of probability distributions associated with quark- and gluon-like jets. Furthermore, given the comparatively coarse binned histograms

[1] This is important in order to estimate the effect of different binnings on the extracted fractions of quarks and gluons.

[2] They are transferable if and only if the additional selection cut is ecaxtly equally efficient in quark and gluon jets (like, for instance, a cut on the azimuthal angle).

(which does not take the evolution of the spectra with increasing transverse momentum into account) that approximate the single-differential cross section, the calculation of the reducibility factors themselves involves substantial complications.

To illustrate the first problem regarding the interpretation of the extracted histograms in terms of proper probability distributions, consider, for instance, the unfolded multijet cross section for some p_{T} bin k in the rapidity category $\alpha(\in \{\text{central}, \text{forward}\})$ with respect to one of the jets substructure observables $\mathcal{O}_{\mathrm{JSS}}$ measured in this report that is approximated by a finitely binned histogram $P^k_{M_\alpha}(\mathcal{O}_{\mathrm{JSS}})$, whereby M_α corresponds to the (jet) "mixture" in the rapidity category α. Given the reducibility factors $\kappa^k_{M_1 M_2}$ and $\kappa^k_{M_2 M_1}$ according to Equation 7.10, the underlying "jet topics"/probability distributions associated with quarks and gluons are obtained using Equation 7.19 by (maximally) subtracting both histograms (central and forward) bin-by-bin, taking into account the correct factors and normalization. This subtraction, however, may result in negative (probability) "mass" in certain bins of the extracted quark-gluon distributions of the respective jet substructure observable, which is – of course – in conflict with the obtained histogram corresponding to a probability distribution because the resulting distribution is not positive definite after all. Instead, in order for the histograms (which are considered to be an approximator of the underlying quark/gluon distributions) to be positive definit, they *must* have a positive probability mass (or vanish) in each bin of the substructure observable; otherwise, the interpretation of the distribution as probabilities is not valid anymore.

To illustrate the problem in mathematical terms, consider the measured distribution $P^k_{M_\alpha}(\mathcal{O}_{\mathrm{JSS}})$ in central ($\alpha = 0$) and forward ($\alpha = 1$) region, whereby $P^k_{M_\alpha, i}$ denotes the probability "mass" in the ith bin of the discretized support (the bin) of the JSS observable $\mathcal{O}_{\mathrm{JSS}}$. According to Equation 7.10, the reducibility factor $\kappa^k_{M_\alpha M_\beta}$ is given by

$$\kappa^k_{M_\alpha M_\beta} = \min_{\mathcal{O}_{\mathrm{JSS}}} \frac{P^k_{M_\alpha}(\mathcal{O}_{\mathrm{JSS}})}{P^k_{M_\beta}(\mathcal{O}_{\mathrm{JSS}})}, \tag{13.1}$$

whereby $\alpha \neq \beta$. Furthermore, the associated anchor bin $\mathcal{O}^{\downarrow}_{\mathrm{JSS}\alpha\beta}$ of the JSS observable is given by

$$\mathcal{O}^{\downarrow}_{k,\alpha\beta} := \arg\min_{\mathcal{O}_{\mathrm{JSS}}} \frac{P^k_{M_\alpha}(\mathcal{O}_{\mathrm{JSS}})}{P^k_{M_\beta}(\mathcal{O}_{\mathrm{JSS}})}, \tag{13.2}$$

with the anchor bin $\mathcal{O}^{\downarrow}_{k\alpha\beta}$ not necessarily corresponding to the limits of the support (largest and smallest value of $\mathcal{O}_{\mathrm{JSS}}$). The latter aspect is responsible for the extracted distributions may not being positive definite. To see this, consider, once again, the definition of the extracted (probability) distributions associated with the quark and the gluon according to Equation 7.19. By construction, the probability distributions *must* vanish for the respective anchor bin, i.e., $P^k_{M_\alpha}(\mathcal{O}^{\downarrow}_{k,\beta\alpha}) = 0$ and $P^k_{M_\beta}(\mathcal{O}^{\downarrow}_{k,\alpha\beta}) = 0$, as can easily bee seen by substituting the definition of the reducibility factor according to Equation 13.1 into Equation 7.19 for the jet topics. This property is desired because the underlying assumption of the method is that the observable results in (almost) pure samples of either quarks or gluons given the corresponding anchor bin. However, if the determined anchor bin does not coincide with the limits of the support of the JSS observable, the probability "mass" of the extracted jet distributions $P^k_{q/g,i}$ in some bin i below (or above) the anchor bin can easily get negative due to the subtraction step in Equation 7.19 if $P^k_{M_\alpha}(\mathcal{O}_{\mathrm{JSS}}) < \kappa^k_{M_\alpha M_\beta} P^k_{M_\beta}(\mathcal{O}_{\mathrm{JSS}})$. This problem can be avoided if the anchor bin is guaranteed to *always* correspond to the limits of the support[3]

[3]This is, of course, very reasonable because the statement of the anchor bin also is a statement about all bins above or below it (provided that the JSS observable has been constructed accordingly), depending on whether it corresponds to the upper or the lower limit of the phase space. For example, if $\mathcal{O}^{\downarrow}_{\mathrm{JSS}q}$ corresponds to the anchor bin associated with the quark, the probability of measuring a contribution from the gluon vanishes, i.e., $P_g(\mathcal{O}^{\downarrow}_{\mathrm{JSS}q}) = 0$. Obviously, if the anchor bin is pure in quarks, all bins below $\mathcal{O}^{\downarrow}_{\mathrm{JSS}q}$ must be pure as well, i.e., $P_g(i < \mathcal{O}^{\downarrow}_{\mathrm{JSS}q}) = 0$, provided that the fraction of gluons decreases with decreasing values of the jet substructure observable.

To address this problem, the algorithm for the determination and calculation of the reducibility factors is modified to ensure that the extracted distributions are indeed positive definit. This can be archived by merging bins and repeating the calculation of the reducibility factors till all bins of the extracted distributions are positive such that the resulting histograms can be consistently interpreted as probability distributions.

The (modified) procedure goes as follows: in a first step, the reducibility factors $\kappa_{M_\alpha M_\beta}$ are calculated for the unfolded single-differential multijet cross section with respect to the JSS variable $\mathcal{O}_{\mathrm{JSS}}$ in the kth transverse momentum bin. Once the reducibility factors are known, the underlying quark/gluon distributions are extracted and checked for negative probability masses in each bin. If the distributions turn out to be positive definite, the algorithm terminates and the extracted distributions are considered to be the final probability distributions of the underlying quark/gluon features; otherwise, bins are successively merged (starting from the phase space boundaries) and the reducibility factors are recalculated for the rebinned distributions.

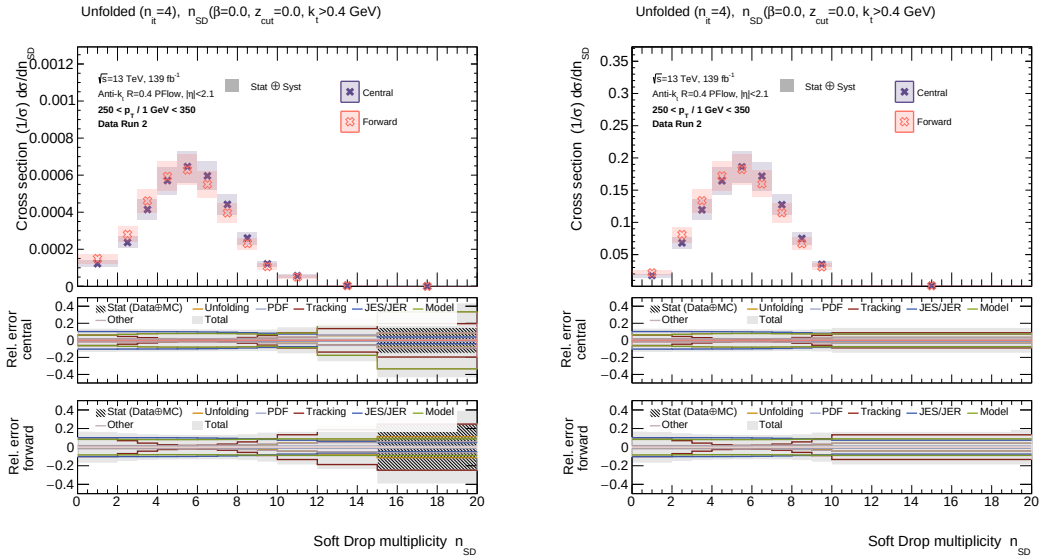
Besides the aforementioned rebinning of the histogram, the calculation of the reducibility was modified in order to account for bins with low statistical significance as well. The problem of bins with low statistical significance was partially already encountered in the pioneering publication in Ref. [29] that introduced the method. In order to obtain meaningful results, they excluded all bins below a certain number of events from the calculation of the reducibility factors to avoid the anchor bin being chosen due to statistical noise. The reason is that bins that are selected as being anchor bins according to Equation 7.11 might not correspond to bins that are considered statistically significant but rather noise; furthermore – compared to the aforementioned phenomenological study that did not include systematic uncertainties –, the binning of the substructure observables is much coarser to allow for the measurement to be unfolded. This removes a significant amount of quark-gluon-related information, which therefore complicates the extraction of related features.

In the end, the scheme for the determination of the anchor region was modified to account for the *statistical* uncertainties of the measurement in each bin according to

$$\mathcal{O}^{\downarrow}_{k,\alpha\beta} = \arg\min_{\mathcal{O}_{\mathrm{JSS}}} \left(r^k_{\alpha\beta}(\mathcal{O}_{\mathrm{JSS}}) + \sigma^{k,\mathrm{stat}}_{\alpha\beta}(\mathcal{O}_{\mathrm{JSS}}) \right), \tag{13.3}$$

whereby $r^k_{\alpha\beta} := P^k_{M_\alpha}(\mathcal{O}_{\mathrm{JSS}})/P^k_{M_\beta}(\mathcal{O}_{\mathrm{JSS}})$ and $\sigma^{k,\mathrm{stat}}_{\alpha\beta}$ denotes the statistical uncertainty on the cross section ratio between central and forward jets in the kth transverse momentum bin in the respective bin of the JSS observable. The corresponding reducibility factor is then given by $\kappa^k_{M_\alpha M_\beta} = r^k_{\alpha\beta}(\mathcal{O}_{\mathrm{JSS}} = \mathcal{O}^{\downarrow}_{k,\alpha\beta})$ of the *rebinned* histogram. According to this modification, bins with very large statistical uncertainties(=low statistical significance) are less likely to be selected as an anchor bin. After each consecutive merging step, Equation 13.3 is once again evaluated until a final configuration has been found.

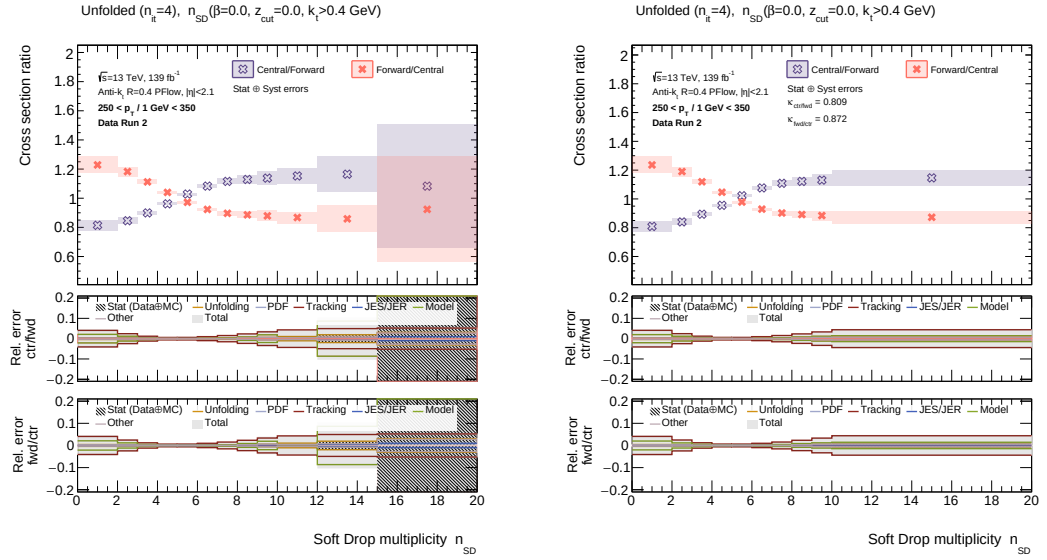
The above-mentioned prescription is best illustrated with the aid of an example. Consider, for instance, the single-differential multijet cross section in central and forward region with respect to the SD multiplicity for emissions with $k_t > 400\,\mathrm{MeV}$ for jets with a transverse momentum of $250\,\mathrm{GeV} < p_{\mathrm{T}} \leq 350\,\mathrm{GeV}$ in Figure 13.1a. The uncertainty of the cross section corresponds to the total uncertainty including all systematic and statistical sources. In the low-momentum regime, the average number of emissions is smaller compared to jets with large transverse momentum; hence, bins corresponding to larger multiplicities come with larger statistical uncertainties. The rebinned distributions according to the procedure described above are shown in Figure 13.1b. As explained above, the updated binning is determined based on the ratio of the single-differential multijet cross sections in central and forward region, which is shown in Figure 13.2, according to the modified algorithm to extract the anchor bins/regions. As one can see, the binning of the phase space differs between Figure 13.1a and Figure 13.1b, which is a consequence of the aforementioned rebinning scheme in the context of the determination of the reducibility factors. The arranged new binning of the histograms according to the example that is shown in Figure 13.1 is quite intuitive given distribution of the SD multiplicity, i.e., higher tails of the distribution are sparsely populated (low statistical significance) and are therefore not suited for the calculation of the reducibility factors. Adjacent bins are therefore successively merged till a stable (i.e., a positive definite) and valid (i.e.,



(a) Single-differential cross section in central and forward region with the *original* binning.

(b) Rebinned single-differential cross section in central and forward region with *updated* binning.

Plot 13.1: Illustration of the described *re*binning procedure of the measured single-differential cross sections in central and forward region (*left*) for the extraction of the rebinned distributions (*right*) .



(a) Ratio between central and forward region with the *original* binning.

(b) Ratio between central and forward region with *updated* binning.

Plot 13.2: Ratio of the single-differential multijet cross sections in central and forward region (*left*) and for the rebinned distributions (*right*).

physical and meaningful) configuration is found.

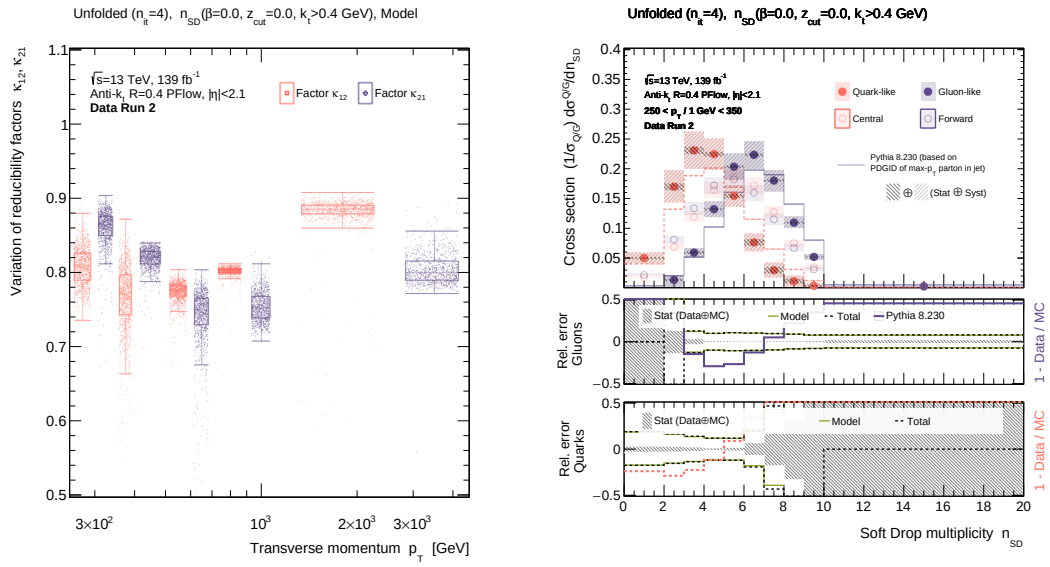
The updated binning of the jet substructure observable follows naturally its evolution with transverse momentum. For example, in case of the counting-based observables (like in Figure 13.1), one expects the anchor region (i.e., the phase space region that is supposed to be pure in

either quark or gluon jets) to increase with transverse momentum of the jets as the number of tracks/emissions increases with p_{T} too. Therefore, the selection of the anchor bins/region must somehow account for this behavior to be not limited by statistics. This is particularly true for the counting-based observables for which the highest-multiplicity bin is sparsely populated except for very high-p_{T} jets (or the lowest multiplicity bins in case of very high momentum jets).

13.1.2 Errors on Quark/Gluon Distributions and Fractions

The last part of this section is dedicated to the propagation of the uncertainties on the unfolded triple-differential multijet cross section onto the extracted underlying quark/gluon probability distributions as well as the corresponding fractions. The general procedure was already explained in Chapter 10.10.

As a quick reminder, the determination of the uncertainties rely on a variational approach that is based on a repetition of the aforementioned extraction algorithm for each component of the total uncertainty on the cross section measurement by mimicking the sampling process to obtain a number of pseudo experiment from which the difference with respect to the nominal distribution can be calculated. In more concrete terms, a total number of (N_{toy} =)1000



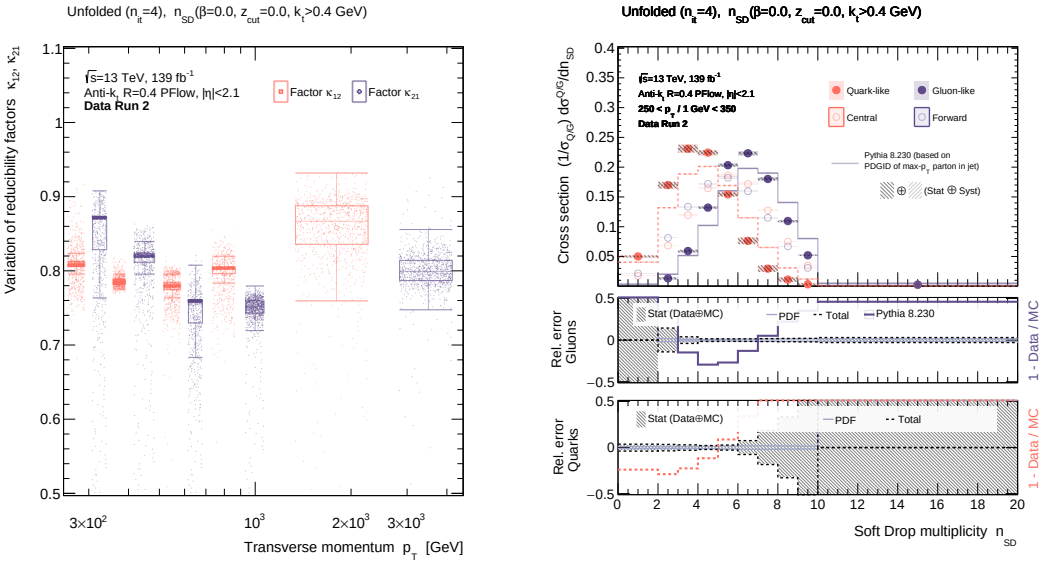
(a) Varied reducibility factors

(b) Extracted distributions

Plot 13.3: Combined model uncertainty (Me, PS, Had)

variations of the unfolded multijet cross section in central and forward region are generated for each individual component of the total uncertainty (which are assumed to be uncorrelated). This process results in $2N_{\mathrm{toy}}$ spectra (central and forward) for each component of the uncertainty $\sigma_{\ell}^{\pm}$, with $\ell \in \{\mathrm{JES}, \mathrm{PS}, \mathrm{Had}, ...\}$ whereby $\pm$ denotes the upper and lower variation, respectively. For each variation of the central and forward spectrum as well as each uncertainty, the reducibility factor $\kappa^{\sigma_{\ell}^{\pm},k}$ is recalculated (whereby k denotes the respective transverse momentum bin), resulting in a new set of reducibility factors $\{\kappa_{ij}^{\sigma_{\ell}^{\pm},k,m}\}_{m\in[1,N_{\mathrm{toy}}]}$ associated to each component of the uncertainty and associated variations of the underlying distributions $\{P(\mathcal{O}_{\mathrm{JSS}})_{g(q)}^{\sigma_{\ell}^{\pm},k,m}\}_{m\in[1,N_{\mathrm{toy}}]}$. Figure 13.3 show some examples of the variation of the reducibility factor for the respective component of the uncertainty (*left*) as well as the corresponding uncertainty on the extracted quark/gluon distributions (*right*). For instance, Figure 13.3 shows the model-related variation of the reducibility factors in each transverse momentum

bin of the measurement (13.3a) as well as the model-related uncertainty on the extracted distributions (13.3b) specific for all jets with $250\,\text{GeV} < p_\text{T} < 350\,\text{GeV}$ to stay consistent with the previous examples. In Figure 13.3a, the "candles" show the mean (circle), the median/50 % quartile (horizontal bar) as well as the lower (Q1/25 %) and higher (Q3/75 %) quartile of the distribution of reducibility factors in each transverse momentum bin. As yet another example



(c) Varied reducibility factors

(d) Extracted distributions

Plot 13.3: Examples for the variation of the reducibility factors (central over forward and *vice versa*) in each transverse momentum bin (*left*) as well as the corresponding uncertainties on the extracted quark/gluon distributions for jets with $250\,\text{GeV} < p_\text{T} < 350\,\text{GeV}$ (*right*) for variations related to the modeling (*top*) as well as the PDF uncertainty (*bottom*).

for a smaller uncertainty, Figure 13.3c-13.3d shows the PDF-related uncertainty. As it was shown in the previous chapters for the triple-differential multijet cross section as a function of the counting- and NN-based JSS observables, this uncertainty is small; consequently, the uncertainty on the underlying distributions due to this component turns out to be small as well.

To get the uncertainty on the extracted underlying quark/gluon distributions associated with $\sigma_\ell^\pm$, the lower and upper quartile is calculated *bin-by-bin* based on the variations with respect to the nominal value.

At the end of this section it should be noted, that (according to the algorithm described in Chapter 7) for each variation of the distribution and the concomitant variation of the reducibility factor as well as the corresponding variation in the underlying quark/gluon distributions, a varied fraction of gluon-like jets in the respective transverse momentum bin (see Equation 7.4) is obtained – keep in mind, that the underlying distributions are assumed to be *universal*, i.e., the measured spectra in central and forward regions are given by the weighted sum of the underlying probability distributions, whereby the weights are given by the fraction of quarks and gluons in the sample (see Equation 7.4). This implies that there is a set of new gluon(quark) fractions $\{f_{g(q)}^{\sigma_\ell^\pm,k,m}\}_{m\in[1,N_\text{toy}]}$ for each component of the uncertainty. The uncertainty on the gluon fraction associated to the variation σ_ℓ is then simply given by the upper and lower quartile of the sample.

13.2 Extraction of Quark/Gluon-Related Features

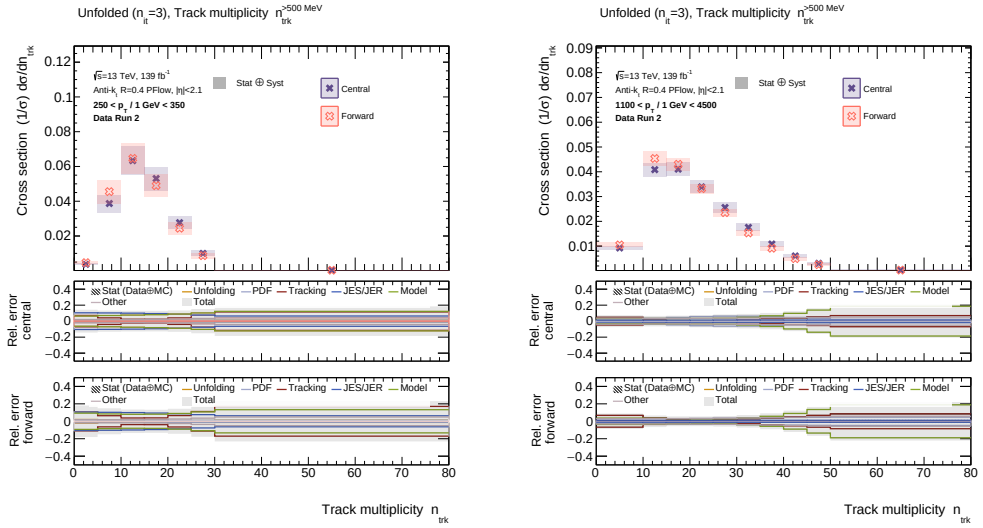
With the triple-differential multijet cross sections for all jet substructure observables being unfolded and all preparatory work being completed, this section *finally* presents the results for the extracted quark/gluon probability distributions as well as the derived quark/gluon fractions in all transverse momentum bins. The main focus of this section is the fraction of gluon-like jets according to the operational definition introduced in Chapter 7.2 for the counting-based (13.2.1) as well as the NN-based (13.2.3) jet shapes; additionally, the extracted underlying quark/gluon (probability) distributions are shown for some selected examples.

The jet themes are extracted following the procedure explained in Chapter 7 including the modifications in the determination of the anchor regions and the reducibility factors derived therefrom (see previous section). The "propagation" of the uncertainties of the unfolded multijet cross section onto the extracted distributions as well as the quark/gluon fractions was explained in Chapter 10 as well as in the previous section.

13.2.1 Track Multiplicity n_{trk}

This section presents results for the track multiplicity n_{trk}. Figure 13.4 shows the measured single-differential multijet cross section with respect to the track multiplicity in central and forward region in two bins of transverse momentum including all sources of statistical and systematic uncertainties, which corresponds to results already shown in Chapter 11.6.1, but in a different representation.

The ratio between both spectra in central and forward region, which are used to calculate

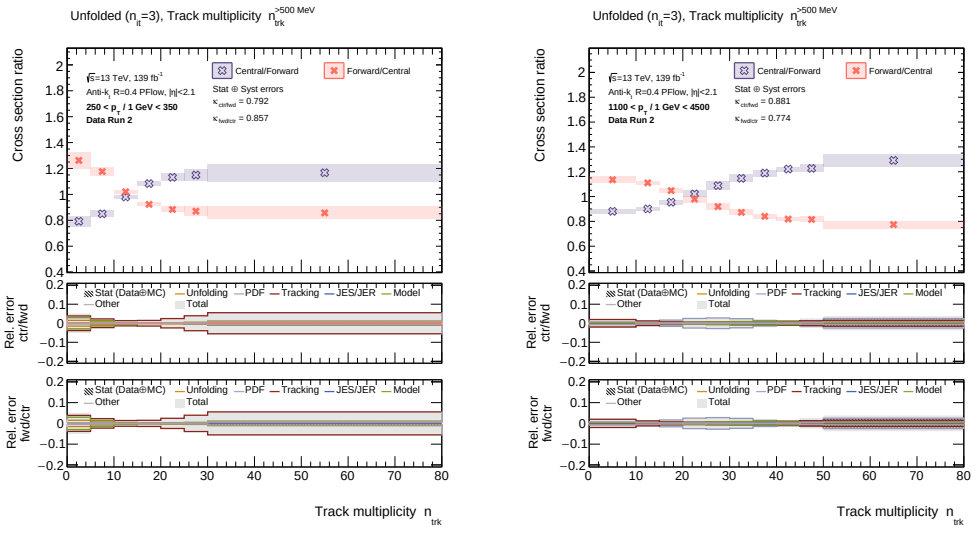


(a) $p_{\mathrm{T}} \in [250, 350)$ *(b)* $p_{\mathrm{T}} \in [1100, 4500)$

Plot 13.4: Rebinned distributions in central and forward region with respect to the track multiplicity n_{trk} for different transverse momentum (in GeV) bins including uncertainties.

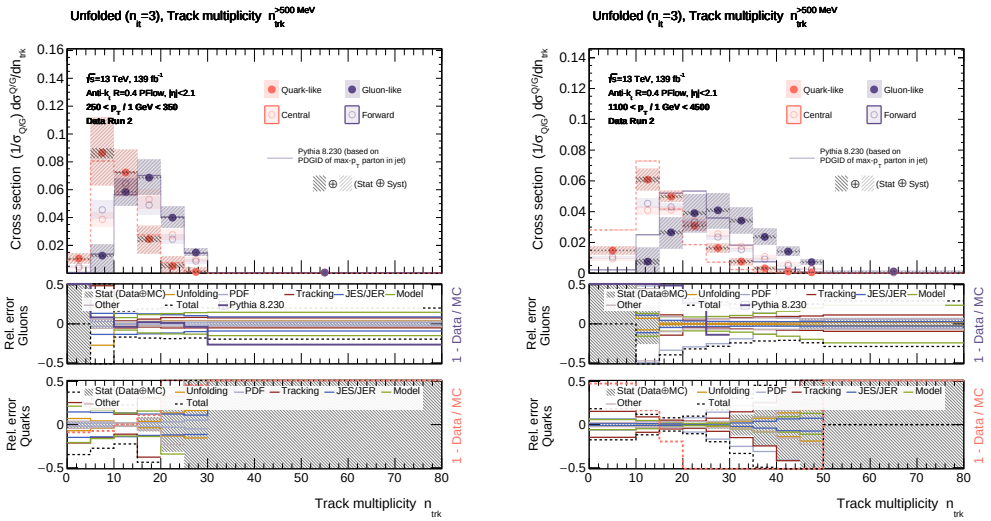
the reducibility factors, are shown in Figure 13.5 for the very same p_{T} bins.

Once the reducibility factors have been inferred from Figure 13.5, both spectra in central and forward region are subtracted from each other according to Equation 7.21 to obtain the underlying quark and gluon probability distributions according to the operational definition of quark- and gluon-like jets (see Definition 2). The extracted quark/gluon distributions along with all sources of statistical and systematic uncertainties are shown in Figure 13.6 for the already familiar transverse momentum bins $250\,\mathrm{GeV} < p_{\mathrm{T}} \leq 350\,\mathrm{GeV}$ (13.6a) and $1100\,\mathrm{GeV} < p_{\mathrm{T}} \leq 4500\,\mathrm{GeV}$ (13.6b). Additionally, the quark/gluon distributions as predicted by Pythia 8.230 (at LO) have been included for the purpose of comparing the result of the operational definition of jets with the prediction of the maximum-p_{T} approach to the jet flavor.



(a) $p_{\mathrm{T}} \in [250, 350)$ *(b)* $p_{\mathrm{T}} \in [1100, 4500)$

Plot 13.5: Ratio between central and forward jets with respect to the track multiplicity n_{trk} for different transverse momentum (in GeV) bins including uncertainties.

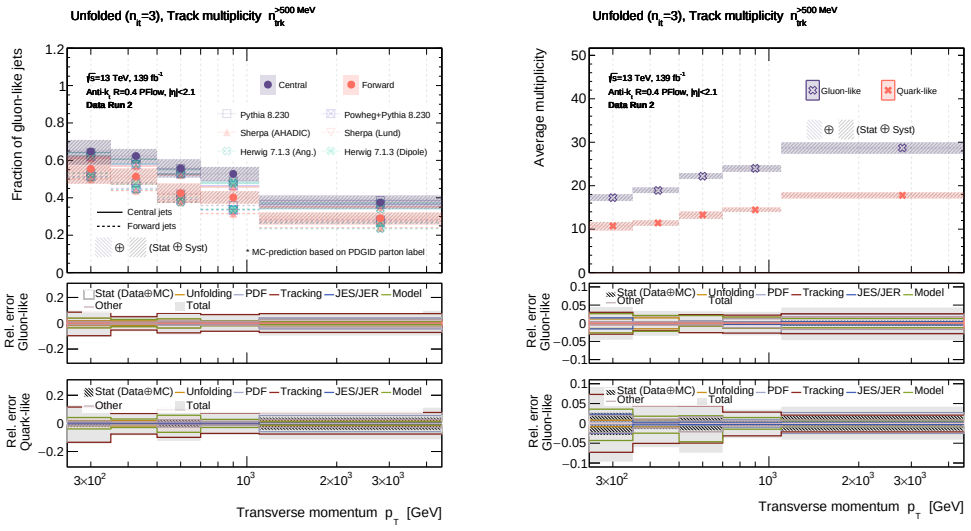


(a) $p_{\mathrm{T}} \in [250, 350)$ *(b)* $p_{\mathrm{T}} \in [1100, 4500)$

Plot 13.6: Extracted underlying quark/gluon(-like) distributions with respect to the track multiplicity n_{trk} for different transverse momentum (in GeV) bins including uncertainties.

Regarding the uncertainties in Figure 13.6a and 13.6b, keep in mind that the cross section of quark- and gluon-like jets in the anchor regions vanishes *by construction* (compare to Equation 12.19e). I.e., according to the premises of the operational definition of quark and gluon jets, the first multiplicity bin in Figure 13.6 is assumed to be pure in quark-like jets; consequently, the *relative* – not the absolute! – uncertainty in this bin is infinite by construction.

Each distribution in Figure 13.6 show the desired features, i.e., an (on average) smaller number of tracks associated to jets originating from quarks and a larger number of emissions for gluon-like jets. Furthermore, the number of tracks in the jet increases for increasing transverse momentum as can be seen by comparing Figure 13.6a and Figure 13.6b.



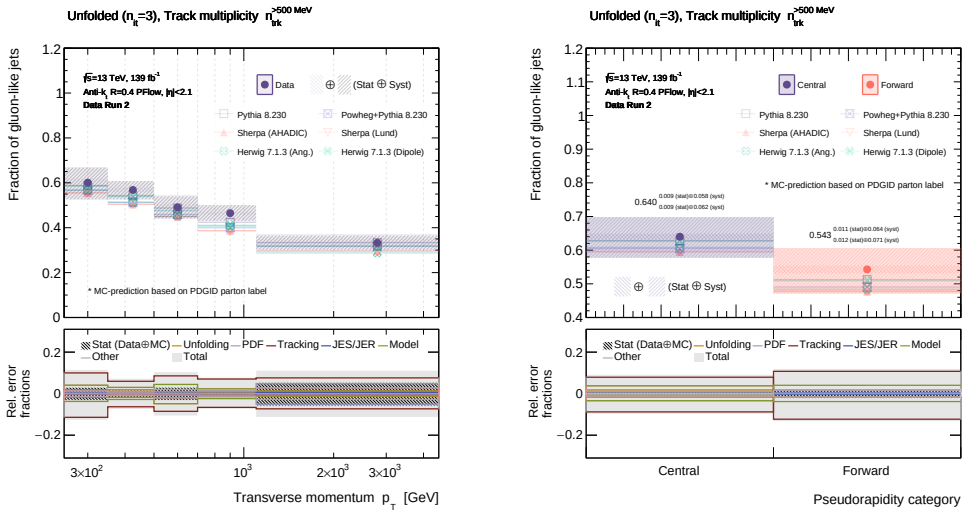
(a) Gluon fractions *(b)* Average multiplicity

Plot 13.7: Extracted gluon fractions (*left*) and average number of tracks for quark- and gluon-like jets (*right*) as a function of transverse momentum including uncertainties.

Besides the underlying probability distributions of quark and gluons, their fractions within each bin of transverse momentum are extracted using the approach outlined in Chapter 7.2. The result of the extracted fractions of gluon-initialized jets using the triple-differential multijet cross section with respect to the track multiplicity is shown in Figure 13.7a. Besides the fraction of gluon-like jets in each p_{T} bin of the measurement, the average number of tracks for quark- and gluon-like jets was determined using the extracted distributions shown in Figure 13.6. The average number of tracks including all sources of uncertainties is shown in Figure 13.7.

Figure 13.7a, on the other hand, shows the fraction of gluon-like jets in both categories of pseudorapidity as well as in each bin of transverse momentum. Figure 13.8a shows the fraction of gluon jets *inclusive* in pseudorapidity, i.e., for the combined sample of central and forward jets. Similarly, Figure 13.8b shows the fraction of gluons in central and forward region inclusive in transverse momentum, i.e., over all p_{T} bins. Figure 13.8b nicely supports the statement of multijet events generally being generally enriched in gluon-initialized jets.

To give some impression of the actual numerical values of the fraction of gluon-initialized jets (according to the operational definition of quarks and gluons), Table 13.1 shows the extracted fraction based on the cross section as a function of the track multiplicity for the central and forward jet sample in each bin in transverse momentum of the measurement including its uncertainties. Furthermore, a combined result for central and forward jets is presented.



(a) Inclusive in η_{rel} *(b)* Inclusive in p_{T}

Plot 13.8: Extracted gluon fractions inclusive in (relative) pseudorapidity (*left*) and inclusive in transverse momentum (*right*) including uncertainties.

	Transverse Momentum p_{T} [GeV]					
	[250, 350)	[350, 500)	[500, 700)	[700, 1100)	[1100, 4500)	[250, 4500)
Rapidity category η_{rel}	Fraction of gluon(quark)-like jets $f_g(1-f_g)^{+\sigma^{\text{up}}_{\text{stat}}\oplus\sigma^{\text{up}}_{\text{syst}}}_{-\sigma^{\text{dn}}_{\text{stat}}\oplus\sigma^{\text{dn}}_{\text{syst}}}$ obtained from $n_{\text{trk}}^{>500\,\text{MeV}}$					
Central	$0.647^{+0.012\oplus0.063}_{-0.012\oplus0.068}$	$0.623^{+0.006\oplus0.038}_{-0.006\oplus0.039}$	$0.557^{+0.009\oplus0.048}_{-0.009\oplus0.047}$	$0.528^{+0.004\oplus0.038}_{-0.004\oplus0.036}$	$0.375^{+0.012\oplus0.034}_{-0.013\oplus0.034}$	$0.640^{+0.009\oplus0.058}_{-0.009\oplus0.062}$
Forward	$0.554^{+0.014\oplus0.070}_{-0.015\oplus0.079}$	$0.513^{+0.007\oplus0.041}_{-0.008\oplus0.043}$	$0.424^{+0.012\oplus0.049}_{-0.012\oplus0.051}$	$0.402^{+0.004\oplus0.033}_{-0.004\oplus0.032}$	$0.290^{+0.013\oplus0.030}_{-0.014\oplus0.030}$	$0.543^{+0.011\oplus0.064}_{-0.012\oplus0.071}$
Combined	$0.601^{+0.019\oplus0.066}_{-0.019\oplus0.073}$	$0.568^{+0.010\oplus0.040}_{-0.010\oplus0.041}$	$0.491^{+0.015\oplus0.049}_{-0.016\oplus0.049}$	$0.465^{+0.006\oplus0.035}_{-0.006\oplus0.034}$	$0.333^{+0.018\oplus0.032}_{-0.020\oplus0.032}$	$0.591^{+0.014\oplus0.061}_{-0.015\oplus0.066}$

Table 13.1: Numerical values of the extracted gluon fractions including its uncertainties for the unfolded measurement of the tripe-differential multijet cross section as a function of the track multiplicity n_{trk}.

13.2.2 SD Multiplicities n_{SD}

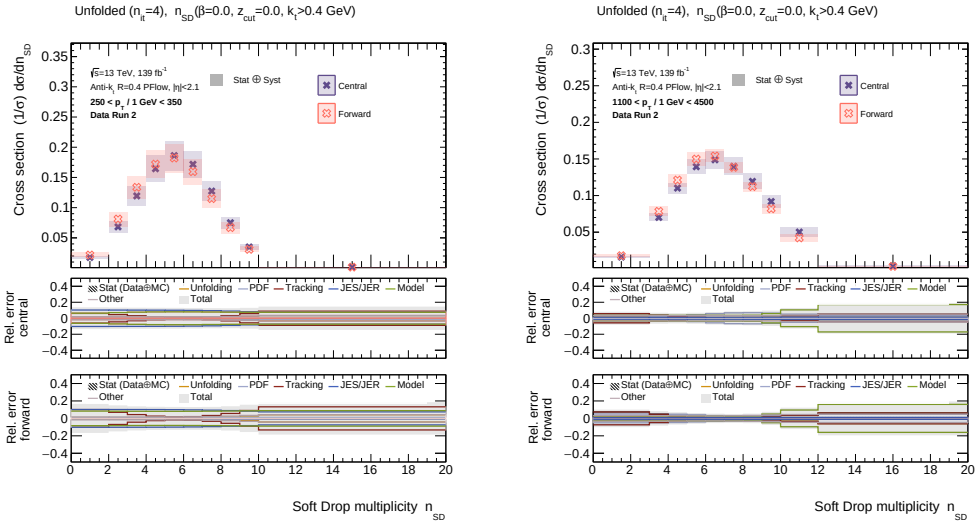
This section basically follows the previous one, but shows the corresponding results for the extracted underlying quark/gluon distributions and gluon fractions for one out of two configurations of the SD multiplicity introduced in Chapter 11.1.3 (both configurations of the SD multiplicity give very similar results).

The first two figures in Figure 13.9 show the measured single-differential multijet cross section as a function of the SD multiplicity $n_{\text{SD}}(k_t \geq 400\,\text{MeV})$ in central and forward region including all sources of statistical and systematic uncertainties.

The corresponding ratio between the two pseudorapidity categories is shown in Figure 13.10.

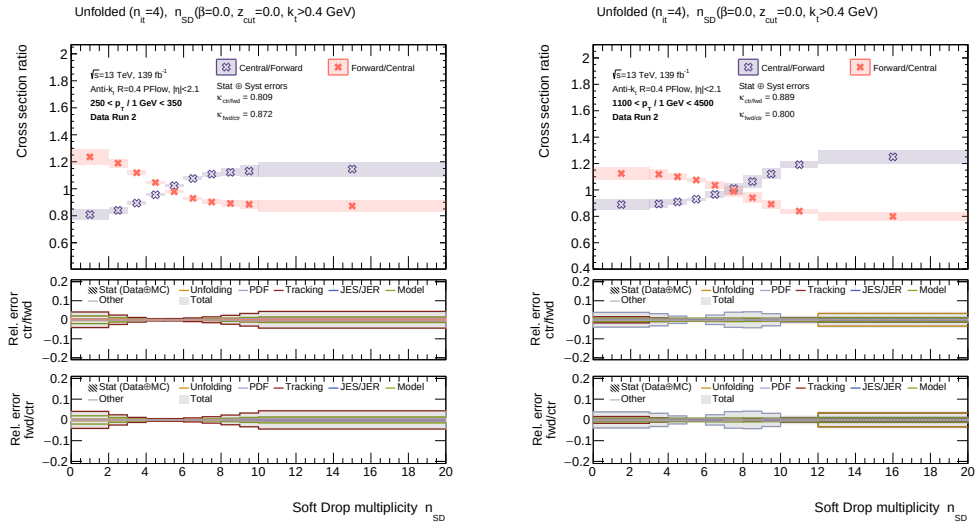
With the relative multijet event rate in central and forward region being measured and the reducibility factors being calculated based on it, the underlying probability distributions associated with quark-like and gluon-like jets can be extracted; the results are shown in Figure 13.11 for $n_{\text{SD}}(k_t \geq 400\,\text{MeV})$ and for $n_{\text{SD}}(z_{\text{cut}} = 0.1, \beta = 1, \theta_{\text{cut}} = 0.002)$ in Figure 13.12.

The extracted quark and gluon distributions exhibit the expected features. While the spectra in central and forward region only differ marginally due to the similar fractions of quark-



(a) $p_T \in [250, 350)$ *(b)* $p_T \in [1100, 4500)$

Plot 13.9: Rebinned distributions in central and forward region with respect to the SD multiplicity $n_{SD}(k_t \geq 400\,\text{MeV})$ for different transverse momentum (in GeV) bins including uncertainties.

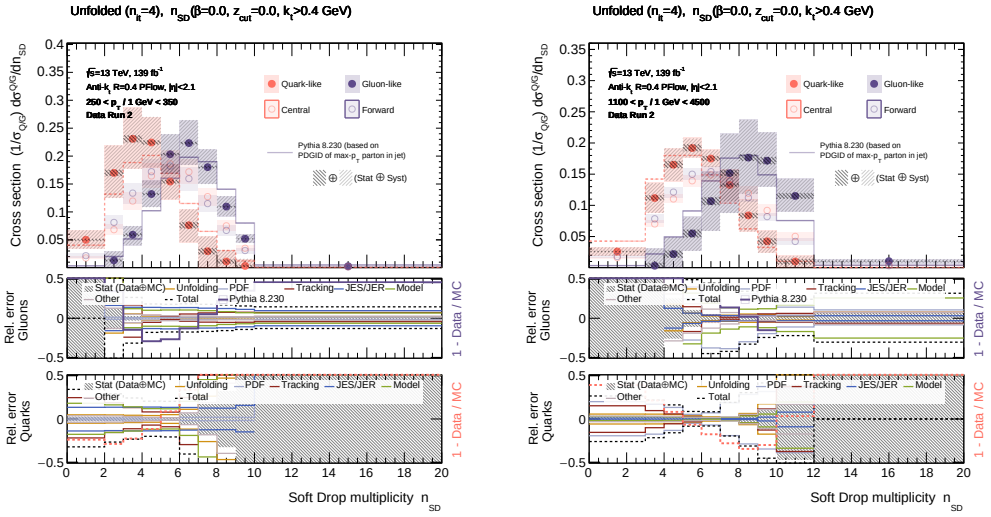


(a) $p_T \in [250, 350)$ *(b)* $p_T \in [1100, 4500)$

Plot 13.10: Ratio between the single-differential multijet cross section as a function of $n_{SD}(k_t \geq 400\,\text{MeV})$ in central and forward region for different transverse momentum (in GeV) bins including uncertainties.

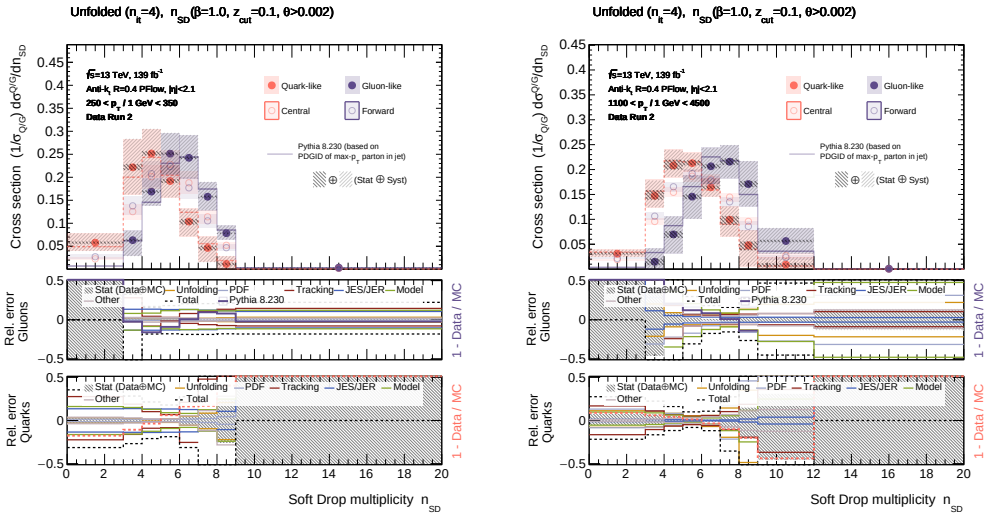
and gluon-initialized jets in both rapidity categories, the extracted underlying distributions ("jet topics") differ noticeably for quark- and gluon-like jets. Furthermore, the mean of the quark and gluon distributions increase with transverse momentum, reflecting the growing number of emissions like it was already observed in the case of n_{trk}.

The extracted fraction of gluon-like jets in each pseudorapidity and transverse momentum bin is shown in Figure 13.13 for $n_{SD}(k_t > 400\,\text{MeV})$ (13.13a) and $n_{SD}(z_{cut} = 0.1, \beta = 1, \theta_{cut} = 0.002)$ (13.13b).



(a) $p_{\mathrm{T}} \in [250, 350)$ *(b)* $p_{\mathrm{T}} \in [1100, 4500)$

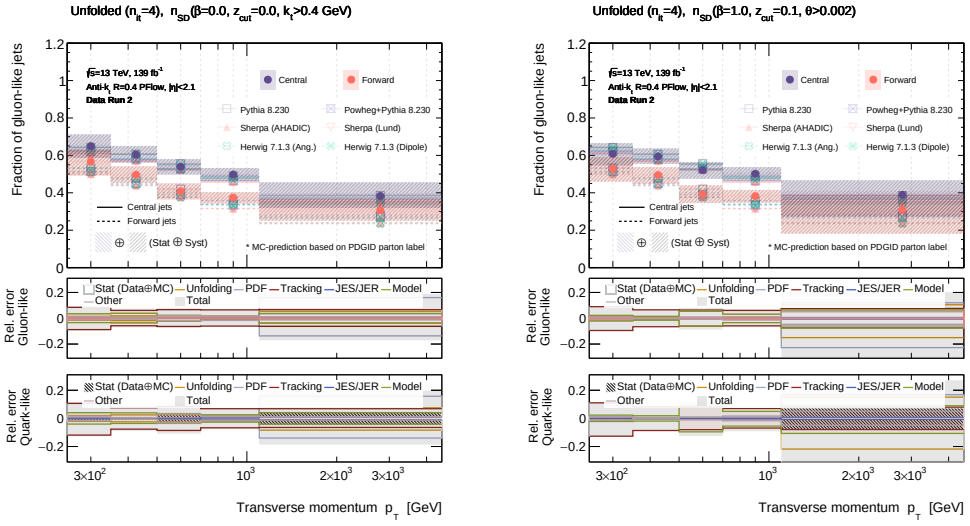
Plot 13.11: Extracted underlying quark/gluon distributions with respect to the SD multiplicity with $n_{\mathrm{SD}}(k_t \geq 400\,\mathrm{MeV})$ for different transverse momentum (in GeV) bins including uncertainties.



(a) $p_{\mathrm{T}} \in [250, 350)$ *(b)* $p_{\mathrm{T}} \in [1100, 4500)$

Plot 13.12: Extracted underlying quark/gluon distributions with respect to the SD multiplicity with $n_{\mathrm{SD}}(z_{\mathrm{cut}} = 0.1, \beta = 1, \theta_{\mathrm{cut}} = 0.002)$ for different transverse momentum (in GeV) bins including uncertainties.

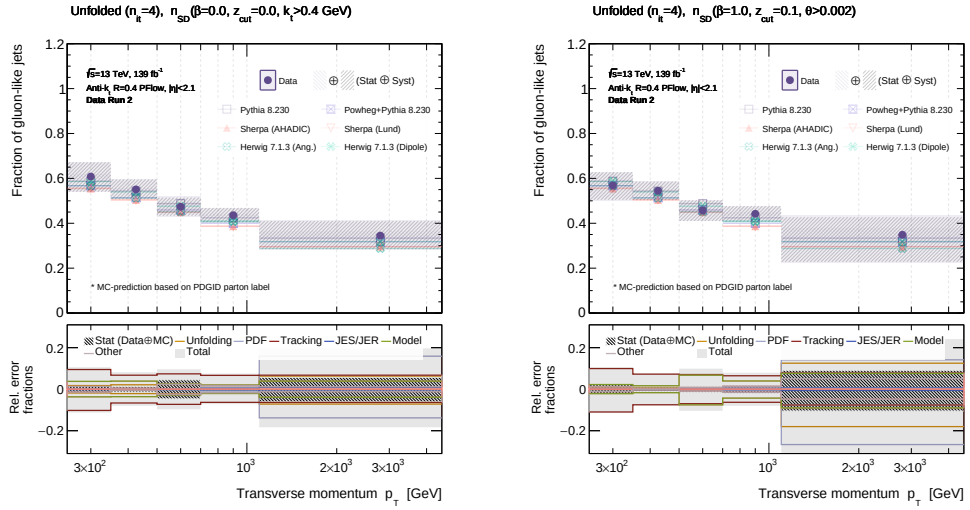
As one can see, simulation and model-independent/data-driven measurement are in good agreement given the uncertainties, indicating that the procedure of calculating the reducibility factors by means of rebinning the distributions provides reasonable results indeed. The experimental uncertainty of the extracted gluon fraction strongly depends on the uncertainty associated to the anchor bin/region, which may vary between different bins of transverse momentum. As expected from the underlying PDFs of the quarks and the gluon, the fraction of gluon-initialized jets in the central rapidity category is larger compared to forward jets.



(a) $n_{\mathrm{SD}}(k_t \geq 400\,\mathrm{MeV})$ *(b)* $n_{\mathrm{SD}}(z_{\mathrm{cut}} = 0.1, \beta = 1, \theta_{\mathrm{cut}} = 0.002)$

Plot 13.13: Extracted gluon fractions for $n_{\mathrm{SD}}(k_t \geq 400\,\mathrm{MeV})$ (*left*) and for $n_{\mathrm{SD}}(z_{\mathrm{cut}} = 0.1, \beta = 1, \theta_{\mathrm{cut}} = 0.002)$ (*right*) including uncertainties.

Last but not least, Figure 13.15 shows the fraction of gluon-like jets in both pseudorapidity

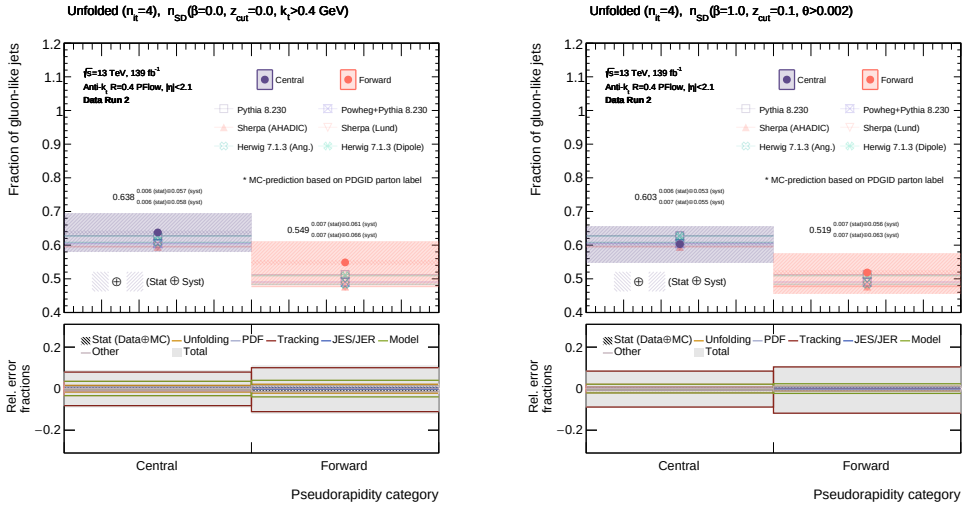


(a) $n_{\mathrm{SD}}(k_t \geq 400\,\mathrm{MeV})$ *(b)* $n_{\mathrm{SD}}(z_{\mathrm{cut}} = 0.1, \beta = 1, \theta_{\mathrm{cut}} = 0.002)$

Plot 13.14: Extracted gluon fractions inclusive in pseudorapidity for $n_{\mathrm{SD}}(k_t \geq 400\,\mathrm{MeV})$ (*left*) and for $n_{\mathrm{SD}}(z_{\mathrm{cut}} = 0.1, \beta = 1, \theta_{\mathrm{cut}} = 0.002)$ (*right*) including uncertainties.

categories inclusive in p_{T} for all configurations of the SD multiplicity including all sources of uncertainties.

The extracted underlying probability distributions of quark and gluon jets can also be used to derive a ROC curve for each transverse momentum bin in a solely data-driven manner to quantify the, e.g., efficiency in gluon-like jets versus the rejection of quark-like jets – or vice versa. This is tempting because it allows to define working points (including systematic uncertainties) for the selection of gluons and quarks, respectively. Two examples for ROC curves are shown in Figure 13.16 for two bins in transverse momentum for the track multiplicity

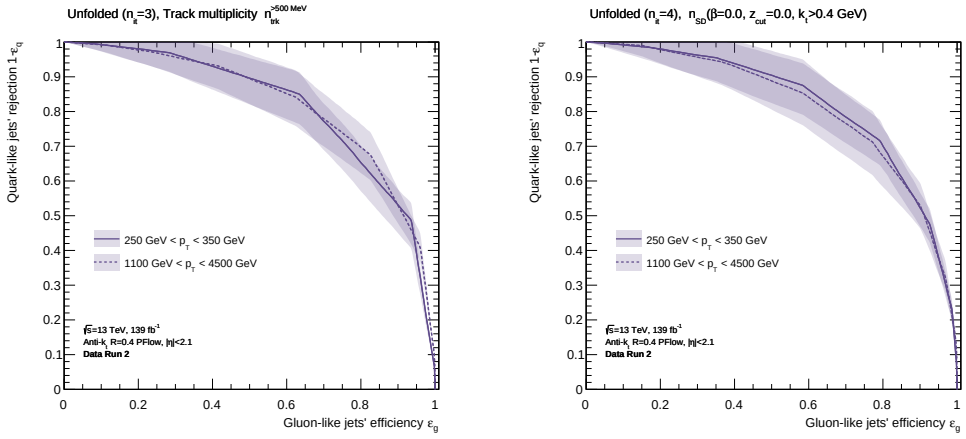


(a) $n_{\mathrm{SD}}(k_t \geq 400\,\mathrm{MeV})$ *(b)* $n_{\mathrm{SD}}(\beta = 1, z_{\mathrm{cut}} = 0.1, \theta_{\mathrm{cut}} = 0.002)$

Plot 13.15: Extracted gluon fractions inclusive in transverse momentum for $n_{\mathrm{SD}}(k_t \geq 400\,\mathrm{MeV})$ (*left*) and for $n_{\mathrm{SD}}(z_{\mathrm{cut}} = 0.1, \beta = 1, \theta_{\mathrm{cut}} = 0.002)$ (*right*) including uncertainties.

(13.16a) and the SD multiplicity $n_{\mathrm{SD}}(k_t \geq 400\,\mathrm{MeV})$ (13.16b) to provide a comparison for both observables.

To also provide a numerical example of the extracted gluon fractions based on the unfolded



(a) Track multiplicity n_{trk} *(b)* SD multiplicity $n_{\mathrm{SD}}(k_t \geq 400\,\mathrm{MeV})$

Plot 13.16: Data-based ROC curve for jets with $250\,\mathrm{GeV} < p_{\mathrm{T}} \leq 350\,\mathrm{GeV}$ (solid) and $1100\,\mathrm{GeV} < p_{\mathrm{T}} \leq 4500\,\mathrm{GeV}$ (dashed) for the track multiplicity (*left*) and SD multiplicity $n_{\mathrm{SD}}(k_t \geq 400\,\mathrm{MeV})$ (*right*).

multijet cross section with respect to the SD multiplicity, Table 13.2 shows the fraction of jets originating from gluons that was calculated based on the measured triple-differential multijet cross section as a function of the default configuration of the SD multiplicity (see corresponding results in Chapter 11.6.2).

For most transverse momentum bins, the extracted fractions of gluon-initialized jets are consistent within their uncertainties.

	Transverse Momentum p_T [GeV]					
	[250, 350)	[350, 500)	[500, 700)	[700, 1100)	[1100, 4500)	[250, 4500)
Rapidity category η_rel	Fraction of gluon(quark)-like jets $f_g(1-f_g)^{+\sigma^\mathrm{up}_\mathrm{stat}\oplus\sigma^\mathrm{up}_\mathrm{syst}}_{-\sigma^\mathrm{dn}_\mathrm{stat}\oplus\sigma^\mathrm{dn}_\mathrm{syst}}$ obtained from $n_\mathrm{SD}(k_t > 400\,\mathrm{MeV})$					
Central	$0.649^{+0.007\oplus0.061}_{-0.007\oplus0.062}$	$0.605^{+0.005\oplus0.045}_{-0.005\oplus0.043}$	$0.539^{+0.012\oplus0.040}_{-0.013\oplus0.039}$	$0.498^{+0.003\oplus0.035}_{-0.003\oplus0.034}$	$0.383^{+0.014\oplus0.071}_{-0.014\oplus0.064}$	$0.638^{+0.006\oplus0.057}_{-0.006\oplus0.058}$
Forward	$0.567^{+0.008\oplus0.066}_{-0.009\oplus0.073}$	$0.496^{+0.006\oplus0.044}_{-0.006\oplus0.044}$	$0.409^{+0.016\oplus0.040}_{-0.016\oplus0.041}$	$0.374^{+0.003\oplus0.030}_{-0.003\oplus0.029}$	$0.306^{+0.014\oplus0.059}_{-0.015\oplus0.055}$	$0.549^{+0.007\oplus0.061}_{-0.007\oplus0.066}$
Combined	$0.608^{+0.011\oplus0.063}_{-0.012\oplus0.067}$	$0.551^{+0.008\oplus0.044}_{-0.008\oplus0.043}$	$0.474^{+0.021\oplus0.040}_{-0.022\oplus0.040}$	$0.436^{+0.005\oplus0.032}_{-0.005\oplus0.031}$	$0.345^{+0.020\oplus0.065}_{-0.021\oplus0.060}$	$0.593^{+0.009\oplus0.059}_{-0.009\oplus0.062}$

Table 13.2: Numerical values of the extracted gluon fractions including its uncertainties for the unfolded measurement of the relative tripe-differential multijet cross section as a function of the SD multiplicity with $k_t > 400\,\mathrm{MeV}$.

	Transverse Momentum p_T [GeV]					
	[250, 350)	[350, 500)	[500, 700)	[700, 1100)	[1100, 4500)	[250, 4500)
Rapidity category η_rel	Fraction of gluon(quark)-like jets $f_g(1-f_g)^{+\sigma^\mathrm{up}_\mathrm{stat}\oplus\sigma^\mathrm{up}_\mathrm{syst}}_{-\sigma^\mathrm{dn}_\mathrm{stat}\oplus\sigma^\mathrm{dn}_\mathrm{syst}}$ obtained from $n_\mathrm{SD}(z_\mathrm{cut}=0.1,\beta=1,\theta_\mathrm{cut}=0.002)$					
Central	$0.608^{+0.008\oplus0.057}_{-0.008\oplus0.060}$	$0.595^{+0.003\oplus0.040}_{-0.003\oplus0.040}$	$0.521^{+0.004\oplus0.044}_{-0.004\oplus0.045}$	$0.502^{+0.005\oplus0.035}_{-0.005\oplus0.035}$	$0.389^{+0.021\oplus0.076}_{-0.023\oplus0.115}$	$0.603^{+0.006\oplus0.053}_{-0.007\oplus0.055}$
Forward	$0.529^{+0.009\oplus0.061}_{-0.009\oplus0.069}$	$0.495^{+0.004\oplus0.042}_{-0.004\oplus0.044}$	$0.393^{+0.005\oplus0.046}_{-0.005\oplus0.049}$	$0.382^{+0.006\oplus0.034}_{-0.006\oplus0.034}$	$0.308^{+0.023\oplus0.081}_{-0.027\oplus0.126}$	$0.519^{+0.007\oplus0.056}_{-0.007\oplus0.063}$
Combined	$0.569^{+0.012\oplus0.059}_{-0.013\oplus0.064}$	$0.545^{+0.005\oplus0.041}_{-0.005\oplus0.042}$	$0.457^{+0.006\oplus0.045}_{-0.006\oplus0.047}$	$0.442^{+0.008\oplus0.034}_{-0.008\oplus0.034}$	$0.349^{+0.032\oplus0.078}_{-0.037\oplus0.121}$	$0.561^{+0.010\oplus0.055}_{-0.010\oplus0.059}$

Table 13.3: Numerical values of the extracted gluon fractions including its uncertainties for the unfolded measurement of the relative tripe-differential multijet cross section as a function of the SD multiplicity $n_\mathrm{SD}(z_\mathrm{cut}=0.1,\beta=1,\theta_\mathrm{cut}=0.002)$.

13.2.3 NN-Score z_NN

This brief section shows the quark/gluon fractions that are derived based on the unfolded cross sections with respect to the NN-score predicted by the two energy flow networks. Due to the great similarity of both networks (keep in mind that the score predicted by the PFN and the EFN is highly correlated as it was shown in Chapter 12.4), most results presented in this section are obtained using the PFN, but are fully transferable to the corresponding EFN ones if not explicitly mentioned otherwise.

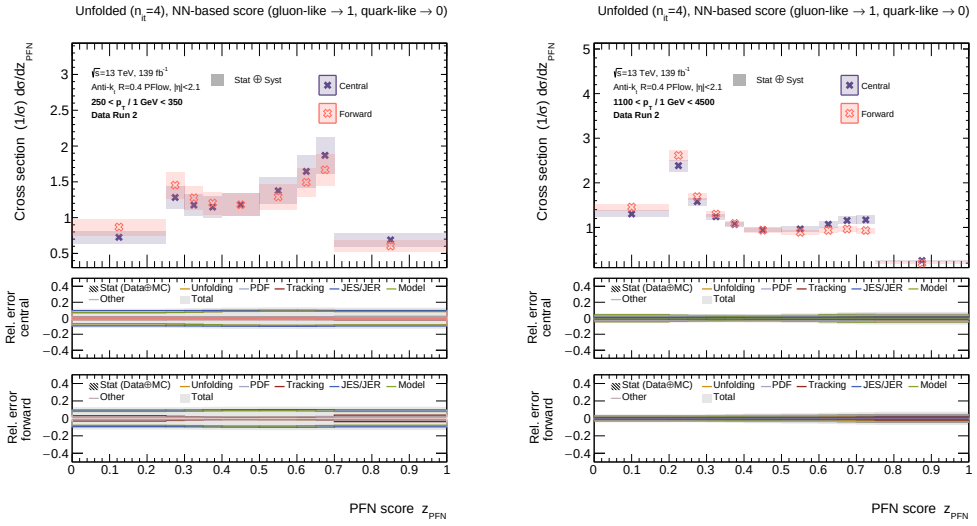
Like in case of the counting-based observables, Figure 13.17 shows the rebinned distributions for the single-differential multijet cross section in central and forward region including all sources of uncertainties for two transverse momentum bins.

The ratios between the spectra of the two pseudorapidity categories for the corresponding transverse momentum bins are shown in Figure 13.18, which are used to calculate the reducibility factors.

The next Figure 13.19 shows the extracted underlying quark/gluon probability distributions with respect to the PFN-score in two transverse momentum bins. The corresponding distributions with respect to the score predicted by the EFN look very similar and are therefore omitted at this point.

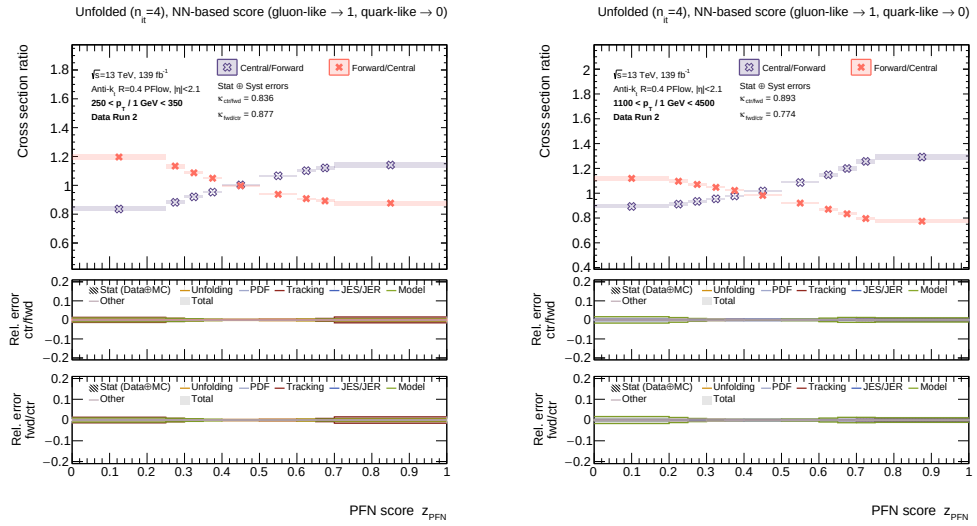
The corresponding fractions of gluon-initialized jets are shown in Figure 13.20 for central and forward jets (13.20a) as well as inclusive in the pseudorapidity category (13.20b). The total fraction of gluon-like jets (combined sample of central and forward jets) in each p_T bin is summarized in Figure 13.21a for the PFN; furthermore, Figure 13.21b shows the ROC curves for jets with $250\,\mathrm{GeV} < p_\mathrm{T} \leq 350\,\mathrm{GeV}$ (solid) and $1100\,\mathrm{GeV} < p_\mathrm{T} \leq 4500\,\mathrm{GeV}$ (dashed).

Finally, Table 13.4 summarizes the fraction of gluon-initialized jets[4] in central and forward



(a) $p_{\mathrm{T}} \in [250, 350)$ *(b)* $p_{\mathrm{T}} \in [1100, 4500)$

Plot 13.17: (Rebinned) distributions in central and forward region with respect to the NN-score predicted by the PFN for different transverse momentum (in GeV) bins including uncertainties.

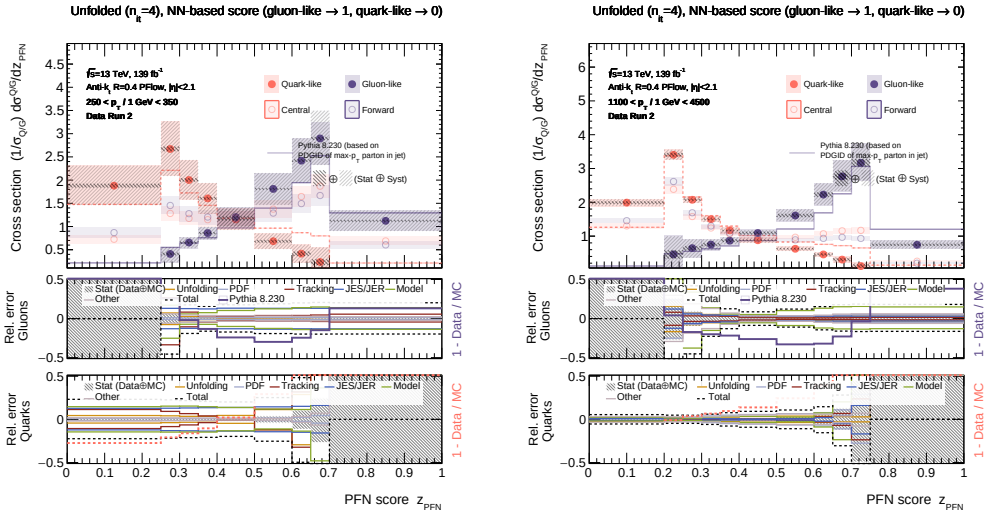


(a) $p_{\mathrm{T}} \in [250, 350)$ *(b)* $p_{\mathrm{T}} \in [1100, 4500)$

Plot 13.18: Ratio between central and forward jets with respect to the NN-score as predicted by the PFN for different transverse momentum (in GeV) bins including uncertainties.

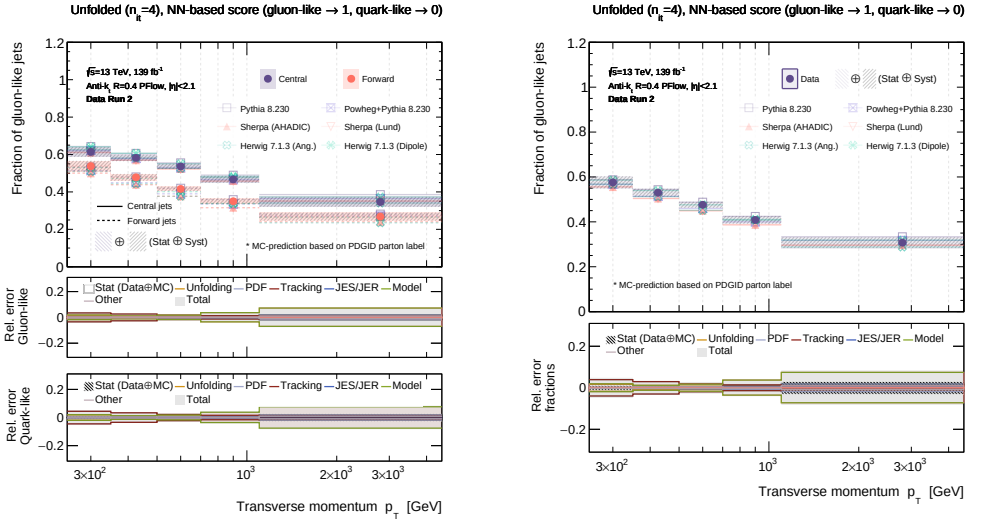
region, the fraction for the combined samples of central and forward jets as well as the total fraction of gluons in both rapidity categories over all transverse momentum bins.

[4]Keep in mind that corresponding numbers for the fractions of (light) quark-like jets can easily be obtained by using the relation $f_q^k = 1 - f_g^k$, whereby the upper index denotes the kth transverse momentum bin. Consequently, the uncertainties remain unchanged in case of quark-initialized jets.



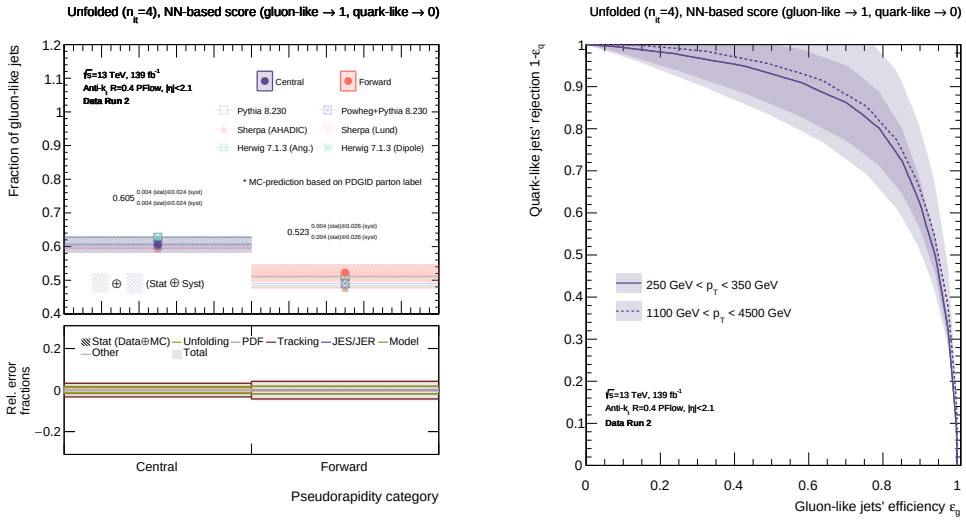
(a) $p_{\mathrm{T}} \in [250, 350)$ *(b)* $p_{\mathrm{T}} \in [1100, 4500)$

Plot 13.19: Extracted (rebinned) underlying quark/gluon distributions with respect to the NN-score as predicted by the PFN for different transverse momentum (in GeV) bins including uncertainties.



(a) *(b)*

Plot 13.20: Extracted gluon fractions according to the PFN for central and forward jets (*left*) and for the combined sample of central and forward jets (*right*) as a function of transverse momentum including uncertainties.



(a) *(b)*

Plot 13.21: Extracted gluon fractions inclusive in transverse momentum according to the PFN (*left*) as well two ROC curves (*right*).

	Transverse Momentum p_{T} [GeV]					
	[250, 350)	[350, 500)	[500, 700)	[700, 1100)	[1100, 4500)	[250, 4500)
Rapidity category η_{rel}	Fraction of gluon(quark)-like jets $f_g(1-f_g)^{+\sigma^{\mathrm{up}}_{\mathrm{stat}} \oplus \sigma^{\mathrm{up}}_{\mathrm{syst}}}_{-\sigma^{\mathrm{dn}}_{\mathrm{stat}} \oplus \sigma^{\mathrm{dn}}_{\mathrm{syst}}}$ obtained from z_{PFN}					
Central	$0.614^{+0.005 \oplus 0.026}_{-0.005 \oplus 0.026}$	$0.582^{+0.003 \oplus 0.017}_{-0.003 \oplus 0.017}$	$0.536^{+0.001 \oplus 0.014}_{-0.001 \oplus 0.013}$	$0.467^{+0.002 \oplus 0.019}_{-0.002 \oplus 0.018}$	$0.346^{+0.007 \oplus 0.026}_{-0.007 \oplus 0.025}$	$0.605^{+0.004 \oplus 0.024}_{-0.004 \oplus 0.024}$
Forward	$0.538^{+0.005 \oplus 0.028}_{-0.005 \oplus 0.029}$	$0.478^{+0.003 \oplus 0.018}_{-0.003 \oplus 0.018}$	$0.416^{+0.001 \oplus 0.013}_{-0.001 \oplus 0.013}$	$0.350^{+0.002 \oplus 0.015}_{-0.002 \oplus 0.015}$	$0.268^{+0.006 \oplus 0.022}_{-0.007 \oplus 0.021}$	$0.523^{+0.004 \oplus 0.026}_{-0.004 \oplus 0.026}$
Combined	$0.576^{+0.007 \oplus 0.027}_{-0.007 \oplus 0.027}$	$0.530^{+0.004 \oplus 0.018}_{-0.004 \oplus 0.018}$	$0.476^{+0.001 \oplus 0.013}_{-0.001 \oplus 0.013}$	$0.408^{+0.002 \oplus 0.017}_{-0.002 \oplus 0.017}$	$0.307^{+0.009 \oplus 0.024}_{-0.010 \oplus 0.023}$	$0.564^{+0.005 \oplus 0.025}_{-0.005 \oplus 0.025}$

Table 13.4: Numerical values of the extracted gluon fractions including its uncertainties for the unfolded measurement of the relative tripe-differential multijet cross section as a function of the NN-score as predicted by the PFN.

13.2.4 Topics for (Non-)Mutual Irreducible Observables

As explained in Chapter 7.4, in order for the demixing procedure to be applicable, the jet substructure observable that is used to measure the rate of multijet events must be mutually irreducible (see Chapter 7.3 for a quick reminder of this concept); otherwise, the extracted fractions and probability distributions will not correspond to features related to quark- and gluon-like jets. In fact, the unmixing procedure will fail in most circumstance like resulting in non-physical results as, for instance, negative fractions etc. However, as mentioned previously, the fractions of gluon-initialized jets over some phase space that have been extracted using a suited jet substructure observable (like the track or SD multiplicity) are *universal* and therefore can be – once they are known – used to extract the underlying quark/gluon probability distributions for observables that are not mutual irreducible provided that the cross section measurement has been performed over the *exact* same phase space volume. The latter point is of major importance, because different selection cuts may come with different efficiencies in quark and gluon jets and therefore change the fraction of those particles in the respective sample of jets. Furthermore, the extracted fractions are both infrared and collinear safe and therefore can be compared to theoretical calculations.

The universality of the extracted gluon/quark fractions as well as their property of being IRC safe opens a vast range of possible applications some of which will be further outlined in the proposal of "future research suggestions" in the end of this report. This final section merely serves the purpose to demonstrate how the extracted quark/gluon fractions can be used to obtain the underlying probability distributions related to quark- and gluon-like jets for a jet (substructure) observable that is not mutually irreducible by means of one examples: the (p_{T}-)normalized jet mass ρ. Furthermore, the underlying quark/gluon themes are extracted for the SD multiplicity with $n_{\mathrm{SD}}(z_{\mathrm{cut}} = 0.1, \theta_{\mathrm{cut}} = 0.002)$ (with $\beta = 0$) based on the fractions that have been measured from the triple-differential multijet cross section despite being mutibile irreducible by itself and therefore suited to directly unmix the distributions.

Normalized Jet Mass

The normalized jet mass is given by the invariant mass of an SD groomed jet divide by the transverse momentum of the *non-groomed* jet according to

$$\rho_{\mathrm{trk}} = \log\left((m_{\mathrm{trk}}^{\mathrm{SD}})^2/(p_{\mathrm{T}})^2\right), \tag{13.4}$$

whereby the p_{T} of the ungroomed jet is used because groomed transverse momentum is not IRC safe[5]. The normalized jet mass possesses some interesting features like the factorization of the distribution related to different physical processes. The lower tail of the distribution is usually associated with hadronization effects that can not be calculated analytically using perturbation theory (non-perturbative regime). In the middle section, on the other hand, the cross section is mostly described by means of resummation. The binning of the normalized jet mass is summarized in Table 13.5.

Two examples for the normalized jet mass are shown in Figure 13.22 for the familiar

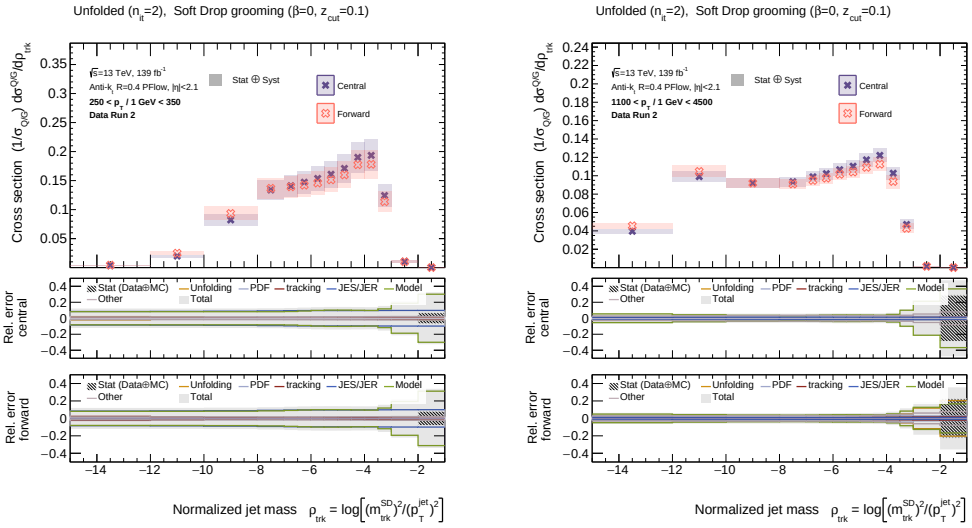
Observable	Binning	Unit
ρ_{trk}	$[-15,-12)$ $[-12,-8)$ $[-8,-7)$ $[-6.5,-6)$ $[-6,-5.5)$ $[-5.5,-5)$ $[-5,-4.5)$ $[-4.5,-4)$ $[-4,-3.5)$ $-3.5,-3)$ $[-3,-2)$ $[-2,-1]$	none

Table 13.5: Binning of the normalized jet mass.

transverse momentum bins $250\,\mathrm{GeV} < p_{\mathrm{T}} \leq 350\,\mathrm{GeV}$ (13.22a) and $1100\,\mathrm{GeV} < p_{\mathrm{T}} \leq 4500\,\mathrm{GeV}$ (13.22b).

Like it was the case for all observables studied within the context of this thesis, the relative

[5] In this example, the invariant mass of the jet is calculated from tracks and therefore not IRC safe in the first place.

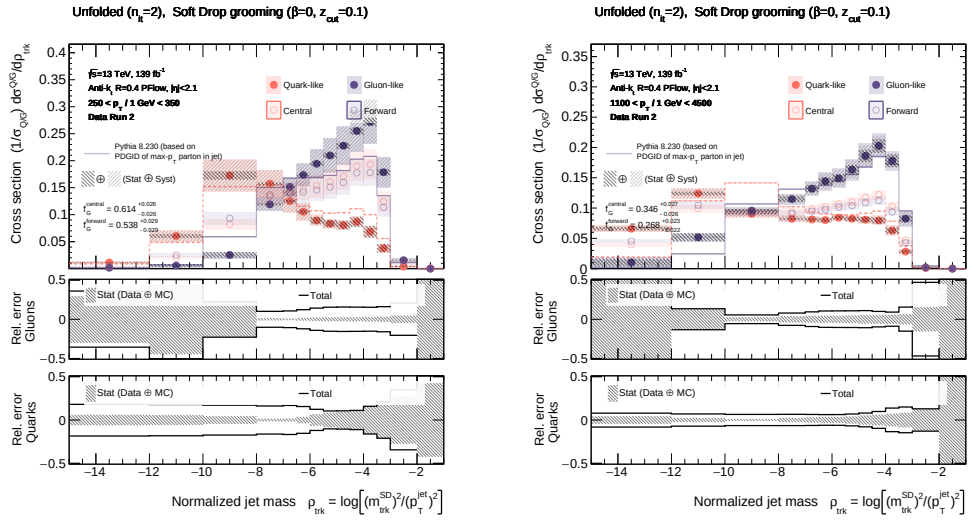


(a) $p_{\mathrm{T}} \in [250, 350)$ *(b)* $p_{\mathrm{T}} \in [1100, 4500)$

Plot 13.22: Distributions in central and forward region as a function of the normalized jet mass for different transverse momentum (in GeV) bins including uncertainties.

event rate was measured for central and forward jets. Closer inspecting both distributions in Figure 13.22 by eye reveals very little differences between the measured cross section for central and forward jets, indicating low sensitivity to features related to quark- and gluon initialized jets – only the resummation region shows some noticeable differences between both rapidity categories due to the sensitivity to the color ratio $C_{\mathrm{A}}/C_{\mathrm{F}}$.

The small sensitivity concerning quark- and gluon-related features already suggests that



(a) $p_{\mathrm{T}} \in [250, 350)$ *(b)* $p_{\mathrm{T}} \in [1100, 4500)$

Plot 13.23: Extracted underlying quark/gluon distributions with respect to the normalized jet mass for different transverse momentum (in GeV) bins including uncertainties.

the (normalized) jet mass being not mutual irreducible; hence, the underlying probability distributions of quarks and gluons can not be extracted using the procedure outlined in Chapter 7.2. However, because the fractions of gluon-like jets are known in each p_{T} bin (after the

analysis that was presented in this report), the degeneracy of the system of equations 7.4 is broken and can be solved for the unknown distributions related to quarks and gluons by means of Equation 7.56 given the known spectra in central and forward region.

The result for the extracted quark/gluon distributions in case of the normalized jet mass are shown in Figure 13.23, whereby the quark/gluon fractions from Table 13.4 (PFN) have been used to solve Equation 7.4 for the unknown distributions.

The extracted distributions in Figure 13.23 show some interesting features whereby special attention is paid to the gluon-related spectra in the resummation region. As can be seen in Figure 13.23a, the gluon-related distribution increases mostly linearly in the resummation region; the slope of the linear growth is proportional to the strong coupling constant and therefore opens a potential window to extract α_{S} based on a jet substructure measurement, which is the idea outlined in Ref. [46] by M. LeBlanc, B. Nachman, and C. Sauer.

SD Multiplicity $n_{\mathrm{SD}}(z_{\mathrm{cut}} = 0.1, \theta_{\mathrm{cut}} = 0.002)$

The last observable to consider is the SD multiplicty with $n_{\mathrm{SD}}(z_{\mathrm{cut}} = 0.1, \theta_{\mathrm{cut}} = 0.002)$. The mutual irreducibility of the SD multiplicity has been explicitly shown in Chapter 7.3 by means of an analytic calculation of the cross section at LL. Therfore, the unmixing procedure could be directly applied in order to obtain the underlying distributions of quark- and gluon-initialized jets as well as the fraction of quark and gluons in each bin of transverse momentum. However, because the fractions have already been consistently measured already, the underlying jet features can be extracted the very same way as it was done in case of the normalized jet mass as well as the momentum fraction.

The next configuration in line of all SD multiplicities corresponds to a lower cut on

Observable	Binning								Unit
z_g	$[0,2)$	$[2,3)$	$[3,4)$	$[4,5)$	$[5,6)$	$[6,7)$	$[7,8)$	$[8,9)$ $[9,20)$	none

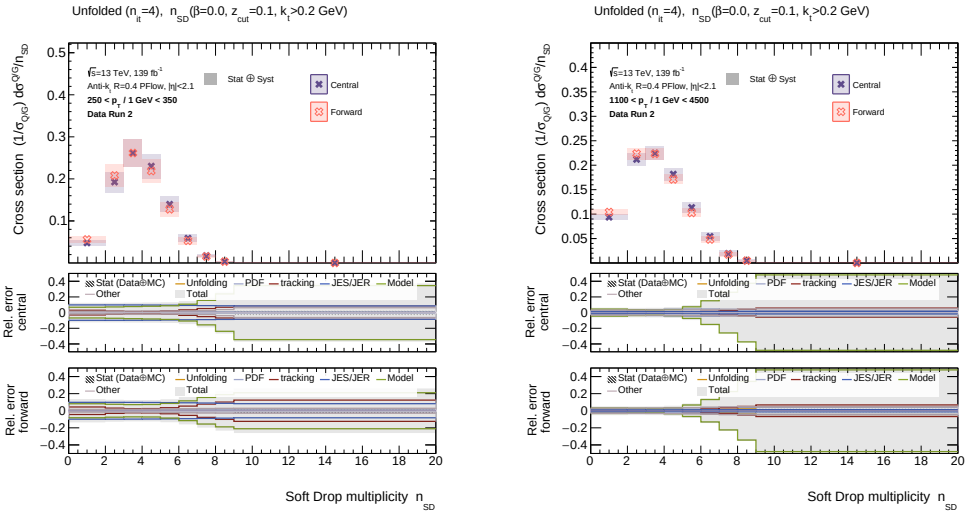
Table 13.6: Binning of the SD Multiplicity $n_{\mathrm{SD}}(z_{\mathrm{cut}} = 0.1, \theta_{\mathrm{cut}} = 0.002)$.

the momentum fraction $z_{\mathrm{cut}} = 0.1$ while keeping the algorithm completely insensitive to the emission angle[6] ($\beta = 0$). The binning is shown in Table 13.6.

As previously state in Section 11.2, this corresponds to a very *aggressive* form of grooming that gets rid of all emissions inside of the jet whose momentum fraction with respect to the jet's "trunk" is below 10 % irregardless of the radiation angle. This configuration of the (I)SD algorithm puts a tight constraint on the region of possible emissions as can be seen in Equation 11.7; hence, a significant shift of the SD multiplicity distribution towards lower multiplicities compared to the other configurations of the ISD algorithm is expected. Furthermore, as the jet's transverse momentum increases, so does (on average) the emissions' k_t. It is therefore anticipated that the multiplicity distributions only change slightly in the high-p_{T} regime compared to softer jets that are accompanied by softer radiation on average. The resulting spectra in central and forward region are shown in Figure 13.24 for two p_{T} bins.

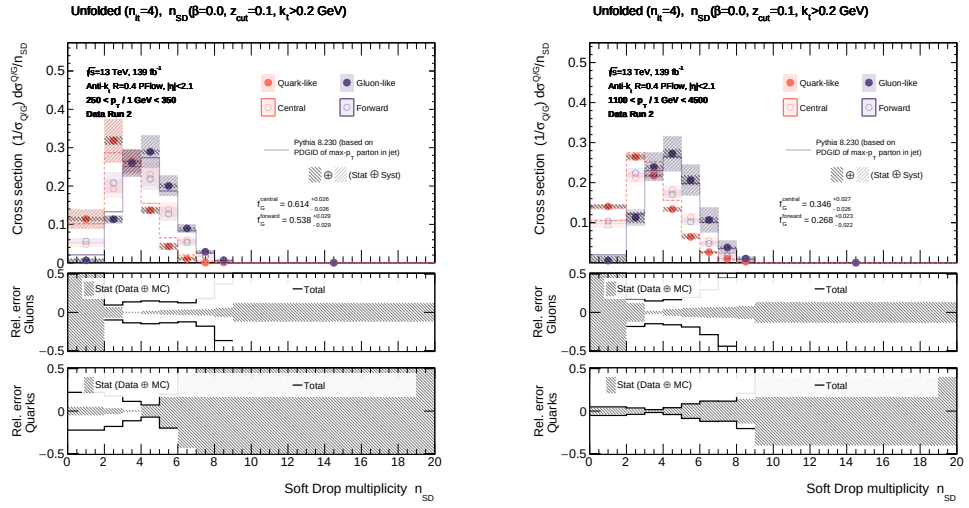
The extracted underlying probability distributions for quark- and gluon-like jets are shown in Figure 13.25, whereby the fractions from Table 13.4 (PFN) have been used to solve Equation 7.56 for the unknown distributions. Regarding the data-to-MC agreement, the exclusion of soft radiation (no matter if very collinear or wide-angled) results in excellent agreement between data and simulation for most regions of phase space.

[6] The only sensitivity to the angular component is a lower cut of $\theta = 0.002$ on the emission angle



(a) $p_{\mathrm{T}} \in [250, 350)$ *(b)* $p_{\mathrm{T}} \in [1100, 4500)$

Plot 13.24: Distributions in central and forward region as a function of the SD multiplicity $n_{\mathrm{SD}}(z_{\mathrm{cut}} = 0.1, \theta_{\mathrm{cut}} = 0.002)$ for different transverse momentum (in GeV) bins including uncertainties.



(a) $p_{\mathrm{T}} \in [250, 350)$ *(b)* $p_{\mathrm{T}} \in [1100, 4500)$

Plot 13.25: Extracted underlying quark/gluon distributions with respect to the SD Multiplicity $n_{\mathrm{SD}}(z_{\mathrm{cut}} = 0.1, \theta_{\mathrm{cut}} = 0.002)$ for different transverse momentum (in GeV) bins including uncertainties.

14. Conclusion

With the fractions as well as the underlying probability distributions of gluon-initialized jets being measured in all bins of transverse momentum and both categories of the relative pseudorapidity, the analysis is concluded; thus, this document is also approaching its end. Hopefully, it was an exciting and educational journey! Before shutting this report for good and turning the focus of attention to more important facets of life, this very last chapter is supposed to provide a brief discussion (14.1) of the insights gained within the course of this analysis along with some critical reflection thereof. In doing so, the most important stations are walked through once again, starting with a quick reminder of the individual steps and their motivation but with a main focus on the physical insights along with their interpretation. After that, the scope of this measurement is broadened by going beyond the boundaries and constraints of this analysis with the objective to embed the results of this study into a larger overall context and to provide some outlook for further research projects (14.2), which may build on (or benefit from) the insight and results presented in this document.

14.1 Summary and Discussion

This analysis can roughly be divided into two consecutive steps that build on top of each other, i.e., the measurement of the triple-differential multijet cross section as a function of several jet substructure observables at a centre-of-mass energy of 13 TeV using an integrated luminosity of $139\,\text{fb}^{-1}$ of data collected with the ATLAS detector in proton-proton collisions at the LHC during Run 2 (2015–2018) of data taking, *and* the extraction of the underlying quark/gluon probability distributions as well as the fractions of quark- and gluon-initialized jets in several regions of phase space. Accordingly, this discussion will focus on both parts individually.

The overall idea of this analysis was, very generally speaking, to study (light) quark- and gluon-initialized jets at the LHC. This is a challenging task in general as at hadron-hadron colliders (like the LHC), it is almost impossible to obtain pure samples in either quark- or gluon-initialized hadronic showers, because one always deals with a mixture of jets due to the underlying PDFs and the resulting statistical nature of the sampling process of quarks and gluons that constitute the proton as an emergent object. This implies that the fractions of quarks and gluons in a given sample of jets are generally *unknown*, which makes it very challenging to study quark/gluon-related properties at hadron colliders in general. Furthermore, the concept of quarks and gluons is theoretically ill defined (see Chapter 7.1.1) at the level of *individual* jets beyond leading order approximation.

The analysis that has been presented in this thesis aims to break the aforementioned degeneracy of mixed samples from hadron-hadron collisions based on a multi-differential cross section measurement over multijet events that provides some sensitivity to the quark- or gluon-like nature of the hadronic shower at the level of the hadronic cross section. This is achived by measuring the number of multijet events as a function of a jet *substructure observable* $\mathcal{O}_{\text{JSS}}$ that provides some discrimination power between quarks and gluons. Broadly speaking, the JSS variables, which have been studied within the context of this thesis, can be divided into two categories: the first set of JSS observables is constructed by iteratively declustering and counting charged constituents of a jet that has been (re-)clustered using the C/A algorithm; the second set is based the output predicted by DNNs derived from the "deep sets" paradigm (see Chapter 6.3) to implement permutation invariant functions over sets and therefore naturally

respecting symmetries associated with jets. Besides the discrimination variable, the (relative) event rate is determined in several regions of transverse momentum (see Table 9.4) with the objective to extract the fraction of quark- and gluon-like jets as a function of p_{T}. The fraction of gluon-initialized jets is known to decrease with increasing transverse momentum of the jets as the momentum of the jets is highly correlated with the energy scale $x \propto p_{\mathrm{T}}/\sqrt{s}$ at which the initial proton has been probed. Besides that, the unmixing procedure requires *two* orthogonal samples of jets that are enriched by either quarks and gluons, respectively, to extract the underlying features associated with the two classes of partons. Therefore, the cross section is also measured in two disjoint categories of different "regions" that are defined by the relative orientation of the two leading-p_{T} jets with $250\,\mathrm{GeV} < p_{\mathrm{T}} < 4.5\,\mathrm{TeV}$ as well as $|\eta| < 2.1$ in the event. Both orthogonal categories of relative pseudorapidity result in two mixtures, M_α with $\alpha \in \{\mathrm{central}, \mathrm{forward}\}$, of jets that differ concerning their composition of quarks and gluons. This is the case because the momentum fraction x is not only correlated with the momentum of the jet but also with the hyperbolic cosine of the pseudorapidity[1]. The difference in quarks and gluons between both categories of relative pseudorapidity is not very pronounced, but it is sufficient[2] to be able to extract their underlying fractions. The final cross section measurement is inclusive because all events with *at least* two jets that pass the event selection contribute to the event yield.

The input that enters the measurement are jets that have been reconstructed from PFlow objects (see Chapter 4.4) using the anti-k_{t} algorithm with a radius parameter of $R = 0.4$. Those PFlow-based jets are used to calculate the transverse momentum as well as the relative pseudorapidity category (see Figure 9.1). The JSS observable, on the other hand, is calculated from Inner Detector tracks with $|\eta| < 2.5$ that have been matched to the PFlow jets (see Chapter 9.2.1), which is why the measurement is sensitive to *charged particles only* at the level of the substructure observable. The JSS observable is reconstructed from tracks to maintain a high degree of correlation between detector- and particle-level objects, which is essential for the subsequent unfolding step to succeed as single-particle resolution is needed to obtain sufficiently diagonal response matrices.

Once the number of multijet events in each phase space bin was measured, the obtained spectra are normalized by the total number of events to cancel systematic uncertainties to the degree of their correlation. For instance, the uncertainty related to the Luminosity drops out completely in this measurement of the relative multijet event rate. Due to the normalization step, the measurement is not sensitive to the total event yield but "only" to the shape of the distributions. A non-normalized measurement of the cross section, however, would be desirable as it comes with an increased sensitivity to quark/gluon-related effects. After that, all spectra are unfolded to particle level (using the nominal response matrix) and are corrected for efficiency and acceptance effects. The uncertainty of the unfolded result is determined by taking several sources of systematic and statistical uncertainties into account and fully propagating them through the full unfolding procedure onto the final cross section. Several state-of-the-art Monte Carlo event generators (see Table 8.1) are used to derive systematic and statistical uncertainties corresponding to different steps of the event generation chain (see Chapter 2.2.2) like, for instance, the calculation of the underlying matrix element, the simulation of parton showers, the underlying hadronization model, the uncertainty associated to the calibration of the energy scale/resolution of the jets etc. (see Chapter 10).

After the unfolding step, the triple-differential multijet cross section is available at particle/hadron level with respect to the track multiplicity n_{trk}, two configurations of the SD multiplicity n_{SD} ($k_t > 400\,\mathrm{MeV}$; $z_{\mathrm{cut}} = 0.1, \beta = 1, \theta_{\mathrm{cut}} = 0.002$), as well as the NN-score z_{PFN} (with the results in case of the EFN being basically identical) including all sources of statistical

[1] The momentum fraction x is approximately given by $x \propto p_{\mathrm{T}}/\sqrt{s} \cdot \cosh\eta$ (which can easily be derived using momentum conservation), i.e., the momentum fraction increases with the pseudorapidity of the jets according to $x \propto \cosh\eta$, implying more quark-initialized jets being present in the forward category, because the probability of interacting with a quark increases for larger x's due to the structure of the underlying PDF of the proton.

[2] The need for two mixtures of jets with different fractions of quarks and gluons can easily be seen by taking a look at the denominator of Equation 7.19, which is singular for $f_{g/q}^{M_1^k} \approx f_{g/q}^{M_2^k}$ because the reducibility factors approach unity.

and systematic uncertainties. Generally, the data-to-MC agreement varies considerably between the different JSS observables and different phase space elements. As expected, the uncertainty associated with the variation of the underlying parton shower model is among the most dominant sources of systematic uncertainties along with the uncertainty related to the calibration of the jet's energy scale and resolution. Also, the tracking-related uncertainties contribute significantly to the total uncertainty in certain regions of phase space: the tracking-related uncertainties are particular pronounced in case of the counting-based observables (n_{trk} and n_{SD}). This is, of course, expected as the tracking-related uncertainties are derived from variations in the number of tracks according to some probability that, in turn, depends on the transverse momentum as well as the pseudorapidity of the jets. Since the counting-based observables are highly correlated with the number of charged particles matched to the reconstructed jet (this is what is counted by the track multiplicity), variations in the number of tracks translates into large systematic uncertainties. Generally, the tracking-related uncertainties are systematically larger in case of forward jets as jets in the forward category experience, on average, an increased material budget compared to the central region (see Figure 9.10a). Broadly speaking, the *total* degree of uncertainty varies between 5 % and 20 %, which is comparatively large but expected in case of the measurement of a jet substructure observable.

A possible approach to reduce the sensitivity of the measurement to modeling-related effects (and consequently increase the sensitivity to other systematic effects) would be to extend the multi-differential cross section by a "latent" variable like, e.g., the track multiplicity (or some other JSS observable) that captures some features of the internal structure of jets. After that, the measurement is reweighted (similar to the reweighing-scheme described, for the DDNC uncertainty as explained in Chapter 10.8) such that the data-to-simulation agreement improves in case of the latent observable. This way, the sensitivity of the measurement to the non-model-related uncertainties (like, for instance, the PDF) would increase[3].

It should be noted, that the measurement of the triple-differential multijet cross section with respect to the various JSS observables forms a *complete* analysis on its own that requires several iterations of the individual analysis steps. For instance, the binning of the observables must be optimized to obtain response matrices that are sufficiently diagonal for the unfolding to succeed; furthermore, the number of iterations used for the unfolding of the spectra at detector level must be adjusted such that the (total) uncertainty is minimal – and plenty more. Especially the binning of the jet substructure observables might be – justifiably – subject to criticism since it is identical for all regions in transverse momentum and therefore does *not* account for the evolution of the differential cross section with the p_{T} of the jets. As explained in Chapter 13.1.1, this is particularly problematic in the context of the extraction of the underlying quark/gluon distributions and their associated fractions. Further studies should therefore focus on an optimized binning of the JSS observable *in each transverse momentum bin* for which the fractions of quark- and gluon-initialized jets are measured. This binning should be chosen to reduce the statistical and systematic uncertainties in the upper and lower limits of the phase space in light of the extraction of the underlying quark/gluon-related features.

Once the triple-differential multijet cross section has been measured for different JSS observables, the underlying probability distributions associated with quark- and gluon-initialized jets as well as their fractions are extracted using the method(s) described in Chapter 7. To cut a long story short, the unmixing procedure relies on the *assumption* of pure regions in phase space of quark- and gluon-jets being present. This does not imply that the measured spectra in central and forward region must yield pure regions in phase space in terms of quarks and gluons – this is (almost) impossible –: it rather means that the respective JSS observable is, *in principle*, sufficiently powerful in terms of quark-gluon discrimination to result in phase space regions that are pure in quarks and gluons, respectively. Based on the unfolded multijet cross sections in central and forward region, the irreducibility factors are calculated and used to maximally subtract the spectra in both rapidity categories to obtain the underlying probability

[3]In practice, this would mean that the measurement step needs to be executed twice: one time to obtain the spectra for multijet events with respect to the latent variable to calculate weights; and a second time for the actual measurement that accounts for the latent variable by means of an additional event weight that increases the agreement between data and simulation with respect to the latent variable.

distributions of quark- and gluon-initialized jets. This picture of quarks and gluons is based on an *operational* definition that does not seek to assign a flavor on a jet-on-jet basis (which is ill-defined in most circumstances as explained in great detail in Chapter 7.1.2) but on the level of higher-level features, i.e., the cross section at hadron level.

The fraction of gluon-initialized jets has been extracted for each jet substructure variable individually in each transverse momentum bin as well as in both categories of relative pseudorapidity. As expected from the underlying PDFs of the (light) quarks and the gluon, the central region is generally enriched by gluon-initialized jets, which is a consequence of the more forward (in terms of absolute (pseudo)rapidity) particles being correlated with the parton of larger longitudinal momentum fraction that is more likely to be a quark due to the structure of the underlying PDFs (see Figure 1.2). Furthermore, the fraction of quark-initialized showers increases with the transverse momentum of jets, which is in accordance with expectation of larger transverse momenta corresponding to higher energies at which the protons are probed; as before, in case of larger momentum fractions, the underlying parton is more likely to be a quark.

The uncertainty on the unfolded triple-differential multijet cross section was propagated onto the extracted fraction and the probability distributions associated with quark- and gluon-like features for each component of the total uncertainty from with the corresponding variations in the reducibility factors (see Equation 7.10) are calculated and translated into an uncertainty on the underlying probability distributions and fractions associated to quark- and gluon-related features.

Generally, the uncertainties on the extracted distributions related to quarks and gluons are large, sometimes being even larger than the multijet cross section in central and forward region. This result reflects the strong dependence of the extracted distributions on the selected anchor bin/region. During the calculation of the uncertainties, variations in the cross section may (in some cases) result in different anchor bins/regions selected, which may cause significant changes in the shape of the extracted distributions. Fixing the anchor bn/region results in *considerably* smaller systematic variations, but the variations of the anchor bin/region must be taken into account for the result to be conclusive. Furthermore, the uncertainty associated with the quark-related distributions are systematically larger than for the corresponding spectra related to the gluon. This can easily be understood by the fact of the forward region of jets being enriched by quarks that (on average) see an increased material budget.

The extracted fractions of gluon-initialized jets for each JSS observable is compared with the prediction from several state-of-the-art Monte Carlo event generators (see Table 8.1), whereby the jet flavor in simulation is assigned using the PDGID of the highest-p_{T} constituent associated with the jet to compare the result with the operation definition of quark and gluon jets. Given the degree of uncertainty, the predictions from simulation are mostly consistent within their uncertainties with the extracted fractions from the Run 2 data set. Furthermore, the extracted underlying distributions have been compared with predictions obtained from the baseline generator (Pythia 8.320), whereby once again, the PDGID of the maximum-p_{T} constituent was used to define the "species" of the jet on a jet-on-jet basis. In general, measurement and simulation are in good agreement given the degree of uncertainty. In case of the track multiplicity, however, some substantial deviations between data and simulations can be observed; the discrepancy increases with the transverse momentum of the jets. This observation may be explained by the high sensitivity of the track multiplicity to non-perturbative effects, which are generally hard to model as no calculations from first principles are possible. This is confirmed by the corresponding results that have been obtained in case of the two configurations of the SD multiplicity, both of which show a significant better agreement between data and simulation for almost all regions in phase space. Both configurations of the ISD algorithm are chosen with the objective to reduce the sensitivity to non-perturbative effects (see corresponding emission planes shown in Figure 11.4), which translates into an overall better data-to-MC agreement as the resulting cross section is dominated by resummation.

It is important to highlight, once again, that a direct comparison of the extracted underlying single-differential multijet cross sections associated with quarks and gluons with predictions obtained from analytic calculations is *not* possible. In case of the track multiplicity, an analytic calculation of the cross section is – for all that is currently known – a priori impossible as it

is not collinear nor (provided that there is no lower-momentum cut) infrared safe; the track multiplicity is not even Sudakov safe! In case of the SD multiplicity, perturbative calculations are, in principle, possible (see an example of a calculation of the SD multiplicity at LL accuracy in Chapter 11.1.4), *however*, only if both, charged and neutral components of the jets are taken into account in the measurement. Analytic calculations of the contribution from charged particles solely are not feasible so far. This problem can in principle be avoided by using objects that account for charged and neutral particles (like, e.g., calorimeter cells) in the calculation of the JSS observable. This solution, however, is ruled out at the present time, because the unfolding of a jet substructure observable requires single particle resolution.

Contrary to the probability distributions, the extracted underlying fractions of quark- and gluon-initialized jets are indeed *IRC-safe*, because the JSS-related information is marginalized and the transverse momentum as well as the relative pseudorapidity are calculated from a jet object that accounts for charged and neutral particles alike.

The numerical values for the extracted fractions of quark and gluon jets for all JSS observables considered in this analysis can be found in Chapter 13. At this point, attention is paid only to the measured fraction of gluons in central and forward region inclusive in transverse momentum obtained from the cross section as a function of the NN-score as predicted by the PFN, because it results in the smallest total uncertainty (in case of the counting-based observables, the total uncertainty is larger due to the increased sensitivity to the tracking related uncertainty). The fraction of gluons in the sample of *central* jets has been measured to be $60.5 \pm 0.4(\text{Stat}) \oplus 2.4(\text{Syst})\,\%$, while the fraction in the *forward* category of jets was measured to be $52.3 \pm 0.4(\text{Stat}) \oplus 2.6(\text{Syst})\,\%$, whereby the numbers refer to the inclusive sample, i.e., all jets with $250\,\text{GeV} \leq p_\text{T} \leq 4500\,\text{GeV}$. This results confirms the initial statement of multijet events generally being enriched in gluons[4]. The fractions of gluons inclusive in transverse momentum are mostly dominated by low-p_T jets due to the steeply falling event rate with the transverse momentum in case of multijet events. Furthermore, the measured fractions nicely demonstrate the central region being enriched by gluons compared to the forward category due to the underlying PDF of the proton. Also, the forward region comes with slightly larger systematic uncertainties, which is expected and explained by the increased material budget in case of more forward jets. In summary, the total uncertainty for the fraction of gluon-like jets is about 4 % and 5 % for central and forward jets, respectively, which roughly corresponds to the uncertainty of the bulk of the distributions.

At this point, it is worth pointing out that the extraction of the fraction of quarks and gluons *in multijet events* according to the method outlined in Chapter 7 does not work anymore for mixtures of jets with $p_\text{T} \lessapprox 100\,\text{GeV}$ and $p_\text{T} \gtrapprox 2\,\text{TeV}$[5], respectively, because the fraction of quark/gluon-initialized jets in central $f_{q/g}^{M_1^k}$ and forward region $f_{q/g}^{M_2^k}$ for the kth momentum bin approach each other $f_{q/g}^{M_1^k} \approx f_{q/g}^{M_2^k}$, resulting in the reducibility factors approaching unity $\kappa^k_{M_1 M_2 / M_2 M_1} \to 1$, with the result of Equation 7.56 being ill defined. This is the case for the two mixtures of jets being defined in multijet events solely. In order to extend the method to higher (and lower) regions of transverse momentum, the two generic mixtures of jets, M_1 and M_2, must be re-defined. A possible solution (already proposed in the pioneering publication in Ref. [29]) would be to define the two mixture of jets over two different processes that differ with respect to the fraction of quarks and gluons in the final state. As it is known (and has been shown in this analysis), multijet events are generally enriched in gluon-like jets; hence, a sample of multijet events M_multijet (no distinction between central and forward jets being made!) can be used to define a mixture of jets that is correlated to features associated with gluons – a mixture of jets that is enriched by quark-like jets $M_\text{Z+jet}$, on the other hand, could, for instance, be obtained from (inclusive) $\text{Z} + \text{jet}$ events. The two mixtures M_multijet and $M_\text{Z+jet}$ naturally come with different fractions of quarks and gluons (see Expression 7.3), however, they add another layer of complexity to the analysis because two different processes

[4] A sample of jets that is statistically enriched in quarks could be, for instance, obtained by using Z+jets instead of multijet events.

[5] Those limits are not based on a systematic, methodical analysis but rather on insights acquired empirically and heuristically.

(different trigger, different physical objects etc.) must be analyzed individually. Furthermore, two mixtures that are defined over two different physical processes may violate the assumption of the jet being an universal object, i.e., no sample-dependence of the jets being studied, which indirectly enters the underlying premises of the linear decomposition according to Equation 7.4. This statement just refers to the circumstance of no *a priori* justification of quark/gluon jets born from multijet and Z + jet events being fundamentally identical; in fact, there are good arguments indeed that suggest that this is not the case – not even in multijet events due to, e.g., color connection. Using two mixtures defined by two processes, however, certainly increases the problem of sample dependence, but would allow to extend the unmixing method to a larger phase space.

The extracted fractions within the same group of jet substructure observables, i.e., counting- and NN-based, agree very well with each other for all transverse momentum bins. Between the different groups of observables, some deviations can be observed that are, however, consistent given the degree of uncertainty. There are two main explanations for the observed discrepancy: *firstly*, both observables differ considerably regarding their discrimination power of quark- and gluon-initialized jets; *secondly*, the granularity of the discretized phase space volume in terms of the binning is quite different between both groups of variables. Starting with the latter argument, the measurement of counting-based jet substructure observables is *generally* challenging. In order to be able to unfold the measurement to particle level in the first place, the binning of the counting-based observables must be comparatively coarse in order to reduce the amount of migrations between bins especially due to tracking-related inefficiencies. A coarser binning, however, reduces the amount of information concerning the shape of the distribution and consequently also decreases the discrimination power between quark- and gluon-initialized jets. Additionally, large bins (especially at the tails of the distributions) may cause problems concerning the purity (the underlying assumption of the unmixing method used to extract the underlying quark/gluon distributions as well as their fractions) of the anchor bins/regions in terms of quarks and gluons. This was one motivation to measure two configurations of the SD multiplicity with respect to different binnings of the underlying phase space.

The NN-based observables, on the other hand, are much more robust in terms of tracking- and model-related variations and therefore allow for a finer binning, which preserves more information regarding the shape of the spectra and provides more freedom concerning the selection of anchor bins/regions. The "resilience" of the NN-score with respect to tracking-related inefficiencies and the consequential variation in the number of tracks thereof is mostly explained by the correlations learned by the model between the individual constituents (and their components) that still allow for a high degree of sensitivity to quark- and gluon related features even if the number of tracks varies. This statement is supported by Figure 12.8b that shows the AUC as a function of the number of tracks provided as an input to the model. Indeed, the robustness with respect to tracking-related variation (along with the increased sensitivity to quark- and gluon-related features) was one motivation for using a NN-based jet substructure observable in the first place. Furthermore, the discrimination power between quark- and gluon-like jets is significantly large in case of the NN-based observables. Consequently, the curve progression of the extracted gluon-fractions as a function of the transverse momentum appears much smoother in case of the NN-based jet substructure observables and also results in overall smaller systematic uncertainties.

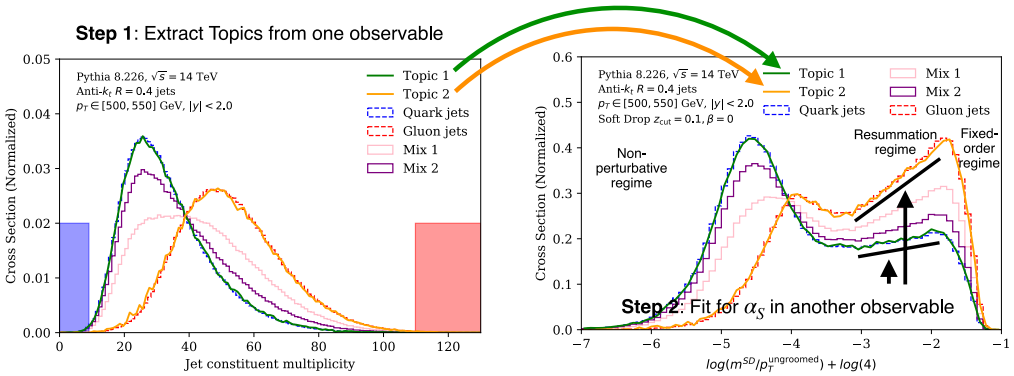
14.2 Future Research Suggestions

This final section outlines some future research suggestions that may be build upon the results presented in this thesis.

Extraction of α_S from a JSS-based Measurement

The possible application of the statistical demixing procedure used in this analysis in the context of a measurement of the strong coupling constant α_S was already suggested and indicated at several occasions in this thesis. The analysis strategy is fully outlined in Ref. [46] by M. LeBlanc, B. Nachman, and C. Sauer; in the following, the principle idea of the measurement is conceptually outlined. The abstract of the paper can be found in the Preface of this document.

The principle idea is related to the discussion in Chapter 13.2.4 in the context of the

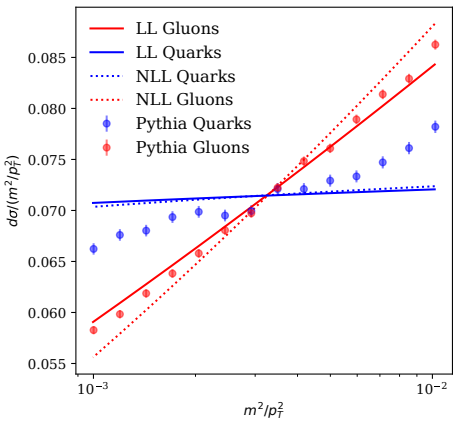


Plot 14.1: "An overview of the proposed extraction of α_S . First (left), the topics are extracted from an approximately mutually irreducible observable. Illustrated above is the case of constituent multiplicity. Two samples with different quark/gluon compositions are used to extract the topics. These topics are interpreted as quarks and gluons and the fractions of the two topics are then used as the quark and gluon fractions. Next (right), an observable with a precisely known cross section is used to fit for aS . We consider the groomed jet mass and focus on the resummation regime at intermediate mass where non-perturbative and fixed- order effects are small. The predictions as a function of aS are then fit to the data, with the proper mixture of the quark and gluon components[.]" (adapted and modified from Ref. 46, Fig. 6, p. 14).

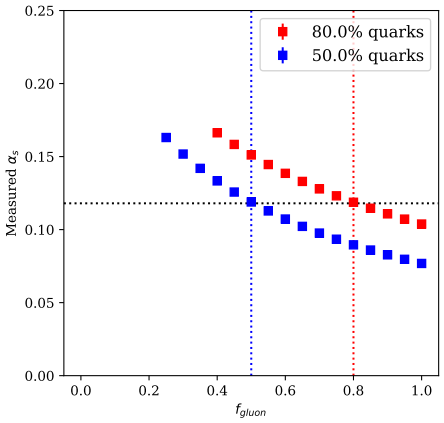
extraction of the underlying quark/gluon distributions for non-mutual irreducible jet substructure observables. The individual steps are outlined in Figure 14.3. The first step is to measure the underlying fractions of quark- and gluon-like jets over some phase space volume in terms of a cross section measurement with respect to a jet substructure observable that is mutually irreducible and therefore allows for the unmixing procedure to be applied. This first part is an entire analysis in itself and basically corresponds to the studies conducted within the context of this thesis[6].

Once the fractions are known, the normalized jet mass is measured and the underlying probability distributions associated with quarks and gluons are extracted according to Equation 7.56 using the fractions from the previous step similar to the results shown in Figure 13.23. As explained previously, the normalized jet mass naturally factorizes into different physical processes, whereby the linearly increasing part of the cross section is mostly determined by resummation. The slope of the linear resummation regime is proportional to the strong coupling constant, allowing for an extraction of α_S in a cross section fit within this region provided that the underlying quark/gluon distributions are known. In principle, this can be done using

[6]Keep in mind that the entire measurement must be repeated once a different selection of jets is used!

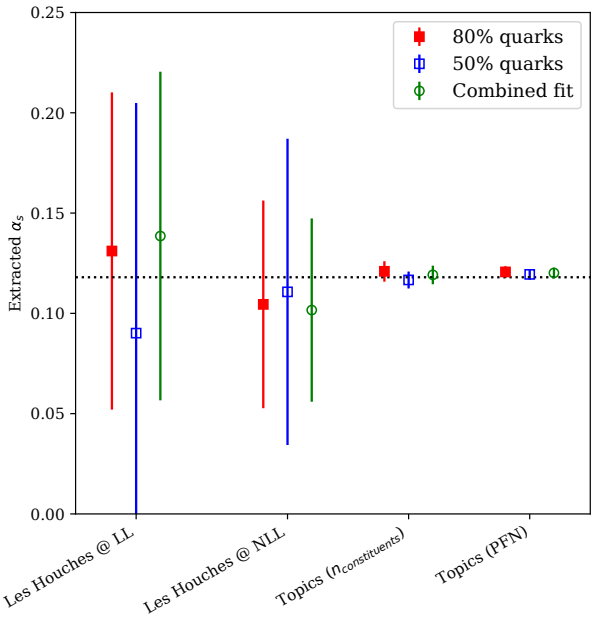


(a) "Comparison of Pythia soft-drop mass prediction to LL and NLL theory predictions. [...]" (adapted from Ref. 46, Fig. 4a, p. 12)



(b) "Characteristic curves obtained by extracting α_S from the groomed jet mass and varying the underlying sample quark/gluon composition using [...] NLL theory predictions." (adapted from Ref. 46, Fig. 5b, p. 12)

Plot 14.2: LL and NLL theory predictions soft-drop mass (*left*) and extracted string coupling α_S as a function of the varied quark/gluon composition (*right*).



Plot 14.3: "Value of α_S obtained [...] with theoretical predictions at leading-logarithmic and next-to-leading-logarithmic accuracy and no a priori knowledge of the quark/gluon fractions, compared to the value obtained using next-to-logarithmic theory and flavor composition information obtained with the jet topics procedure" (adapted from Ref. 46, Fig. 6, p. 14).

Monte Carlo, i.e., measure the fractions of quarks and gluons (according to the ill-defined definition of the jet flavor in terms of the maximum-p_T constituent inside of the jet) to break the degeneracy of the system of linear equations 7.4 and to obtain the underlying spectra related to quarks and gluons. This works fine, however, no meaningful uncertainties can be associated to the extracted fractions of quark/gluons jets. In this analysis, however, the

extracted the underlying fractions of quark- and gluon-like jets are well-defined and include statistical as well as systematic uncertainties. As it was shown, the MC-based prediction of the fractions is consistent with this measurement given the degree of precision.

The fit requires some analytic calculation of the cross section (which is a composition of contributions from gluons and light quarks) within the resummation region. A simple example of such a calculation is illustrated in Figure 14.2a that show the normalized jet mass in the resummation region in simulation for quarks and gluons in comparison with analytic calculations at LL as well as NLL accuracy.

If the underlying composition of quark/gluon jets is unknown, the fraction of quark and gluon jets is unconstrained and consequently results in large uncertainties given the extracted numerical value of α_{S} as it is shown in Figure 14.2b, which shows the degeneracy of the extracted strong coupling and the composition of quarks and gluons. If the fraction is known, however, the fit can be further constrained.

Figure 14.3 summarizes the results for the different extractions of the strong coupling constant. In case of the first two results, the fraction of quark- and gluon-initialized jets was not constrained in the fits that differ concerning the degree of accuracy of the analytic calculation that was used to perform the fit. (The uncertainty decreases for higher-order calculations because non-linearities add additional constraints on the possible values of quark and gluon fractions.) The last two results show the extracted values of α_{S} with the fraction of quarks and gluons being fixed by the values that have been determined based on the unmixing procedure for the track multiplicity as well as the NN-score. Keep in mind that the results shown in Figure 14.3 do *not* include any systematic uncertainties at all!

Reduction of Flavor-Related JES uncertainties

The measurement presented in this thesis may contribute to a reduction of the flavor-related component of the uncertainty due to the calibration of the jet energy scale as well as the Global Sequential Correction, which is supposed to reduce differences in the calorimeter response due to quark- and gluon-initialized showers. As explained in Chapter 4.4.3, the flavor composition uncertainty accounts for differences in the response for quark $\mathcal{R}_q$ and gluon $\mathcal{R}_g$ jets, with $\mathcal{R}_q$ and $\mathcal{R}_g$ being the responses as determined in simulation using Pythia. The uncertainty of the flavor composition is given by Equation 4.13, whereby σ_g^f denotes the uncertainty of the fraction of *gluon*-initialized jets in a sample, with the corresponding uncertainty on the fraction of quark-initialized jets being negligible. This uncertainty is proportional to the residual in the response and the uncertainty on the gluon fraction [45]. So far, all elements in Equation 4.13 are based on information obtained from simulation, i.e., the response for quarks and gluons $\mathcal{R}_{q/g}$ are obtained by performing a full simulation of the ATLAS detector, whereby the flavor of the jet is determined on an event-by-event basis using the PDGID of the maximum-p_{T} constituent of the jet; the fractions of gluon-initialized jets (entering Equation 4.13) are obtained from simulation as well similar to the results shown in Figure 9.1b.

Using the results/method presented in this study, however, allows to determine the flavor composition uncertainty in a solely data-driven manner without any input from simulation. For this, the combined response of quark- and gluon-initialized jets must be measured in data[7] over two samples of jets one of which is statistically enriched in quarks and gluons, respectively, Furthermore, the response is measured in several bins of transverse momentum for which the fraction of quark- and gluon-like jets is known, for instance, after a measurement like the one presented in this document has been performed. After that, the distributions are unmixed the

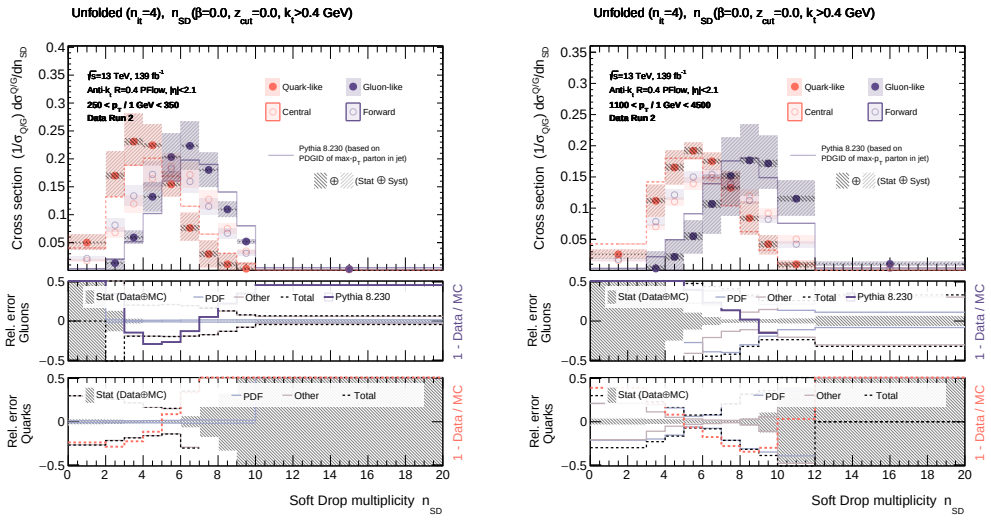
[7]This requires a change in the definition of the response that is usually defined by the ratio of the energy (or the transverse momentum) of the jet reconstructed in the calorimeter of the detector with respect to the energy (p_{T}) of a truth-level jet. Alternatively, the ratio between tracks that have been matched to a calorimeter-based jet could be used, because measurements in the Inner Detector and the calorimeter of the detector are assumed to be uncorrelated. This, however, only accounts for charged components of the jet. Alternatively, the response can still be determined in simulation, but using the fractions of gluon-initialized jets and its uncertainties measured in data using the unmixing method presented in this analysis. Of course, the algorithm for the extraction of the underlying quark- and gluon-related features can be applied to data and simulation alike; hence, the operational definition of quark and gluon jets may be used to avoid common pitfalls regarding the definition of those objects.

same way like it has been done in case of jet observables that are non-mutual irreducible. This results in two responses $\mathcal{R}_{q/g}^{\text{topic}}$ corresponding to quark- and gluon-like jets. Keep in mind that no statement was made about the an individual jet being of quark or gluon origin, instead, features associated with quarks and gluons are extracted rather on the level of the distributions, avoiding common pitfalls concerning the definition of quark and gluon jets. The uncertainty σ_g^f as well as the fraction of gluon-initialized jets are obtained from the unmixing method presented in this analysis.

Reduction of Gluon-Related PDF Uncertainty

Another possible application is the reduction of the uncertainties of the gluon-related PDF. This proposal is based on the observation of the PDF-related component of the uncertainty being dominant in the highest transverse momentum bin. Contrary to most other systematic uncertainties, the PDF uncertainty is increasing with the transverse momentum of the jets, reflecting the increasing uncertainty in the region of larger momentum fractions according to $x \propto p_{\text{T}}/\sqrt{s}$. Furthermore, the PDF-related uncertainty increases for increasing multiplicities, which is connected to the gluon-related PDF being less well constrained compared to the PDFs associated to the different quark flavors. This result opens a possible window to reduce the gluon-related PDF uncertainties.

Figure 14.4 shows the extracted quark- and gluon-like probability distributions as a



(a) $p_{\text{T}} \in [250, 350)$ *(b)* $p_{\text{T}} \in [1100, 4500)$

Plot 14.4: Extracted quark- and gluon-like distributions as a function of the SD multiplicity for jets with $250\,\text{GeV} \leq p_{\text{T}} < 350\,\text{GeV}$ (*left*) and $1100\,\text{GeV} \leq p_{\text{T}} \leq 4500\,\text{GeV}$ (*right*).

function of the SD multiplicity for $k_t > 400\,\text{MeV}$ for two bins in transverse momentum along with the statistical uncertainty, the uncertainty related to variations of the underlying PDF of the proton as well as the total uncertainty (including all statistical and systematic uncertainties). As can be seen by directly comparing the PDF-related uncertainties for jets with $250\,\text{GeV} \leq p_{\text{T}} < 350\,\text{GeV}$ (14.4a) and jets with $1100\,\text{GeV} \leq p_{\text{T}} \leq 4500\,\text{GeV}$ (14.4b), the component of the total uncertainty associated with the PDF increases considerably; in multiplicity bins with the PDF-related uncertainty being dominant, one possibly gains sensitivity to the underlying PDF. This is particularly true for the PDF uncertainty obtained for the gluon-related distribution in Figure 14.4b.

It should be noted, however, that in order to constrain the PDF-related uncertainty, the measurement of the cross section has be IRC safe, which is not the case for this measurement

due to the utilization of tracks in the jet substructure observable.

The different transverse momentum regimes in Figure 14.4 correspond to different regions of the momentum fraction at which the proton has been probed. A (very!) naïve calculation of the momentum fraction shows that the distributions in Figure 14.4a corresponds to a momentum fraction of roughly $x \sim 0.01$, while the spectra in Figure 14.4b correlates with a proton being probed at an energy scale of approximately $x \sim 0.1$.

"Certainly we shall all rise again, certainly we shall see each other and shall tell each other with joy and gladness all that has happened!"

– Fyodor Mikhailovich Dostoevsky (1821 – 1881), *"The Brothers Karamazov"* (1880)

A. Connection to “Topic Modeling”

The method to extract the underlying distributions for different jet types based on the measurement of two (jet) mixtures that leads towards the operational definition of quark and gluon jets (see Section 7.2) has some mathematical resemblances to *“Probabilistic Topic Models – ‘Surveying a suite of algorithms that offer a solution to managing large document archives”’* [28], i.e., *“[...] a paradigm for discovering emergent themes in a corpus of documents[...]“* [29] – quote for the pioneering publication *“On the Topic of Jets”* (p. 1) by Eric M. Metodiev & Jesse Thaler who first adopted this method to jet physics.

Probabilistic topic models are statistical algorithms that aim to discover latent semantic or thematic (like, for instance, categories of different topics) structures in large collections or archives of text documents.

The most simplest probabilistic topic models are based on the so-called *latent Dirichlet allocation*, which assumes that analyzed collection of text documents exhibit multiple topics. Consider, for example, the (real-world) article shown in Figure A.1 with the title *“Seeking Life’s Bare Genetic) Necessities“* [232] that aims to address the question of *“[...] [w]hat’s the minimum number of genes needed for a simple organism to survive[.]”*. In the current page of

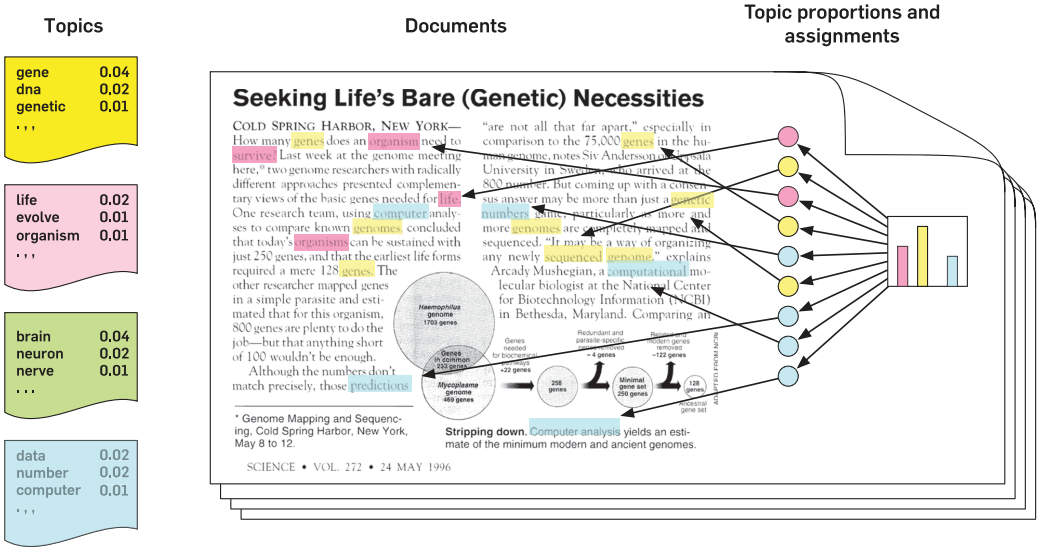


Fig. A.1: “The intuitions behind latent Dirichlet allocation. We assume that some number of ‘topics’, which are distributions over words, exist for the whole collection (far left). each document is assumed to be generated as follows. first choose a distribution over the topics (the histogram at right); then, for each word, choose a topic assignment (the colored coins) and choose the word from the corresponding topic. The topics and topic assignments in this figure are illustrative—they are not fit from real data. [...]” (adapted and modified from Ref. 28, Fig. 1, p. 2(78)).

the article in Figure A.1, different words are highlighted (by hand!) that relate to different topics such as *prediction*, *number* or *computer* that are content-related to generic *data analysis*;

genes, *genomes* etc. are summarized under the topic *genetics*; *organism*, *survive* and *life* belong to the class of *biology*. Conjunctions, connecting words, filler words are excluded as they mostly do not contain information that can be used for topic classification. The intuition behind the association of certain categories of words with a topic it belongs to is supposed to be mimicked by latent Dirichlet approach.

Moving towards a probabilistic interpretation of the aforementioned idea, *topics* are defined *"[...] to be a distribution over a fixed vocabulary[.]"* [48]. Generally, the analyzed data $\boldsymbol{x}$ is assumed to be generated by a *generative process* that depends on a set of latent variables $\boldsymbol{z}$, in which case the data can be interpreted as a joint probability distribution $P(\boldsymbol{x}, \boldsymbol{z})$ of the observed data and the hidden variables of the model. In this picture, the concerted instance of data is then a conditional distribution $P(\boldsymbol{x}|\boldsymbol{z})$ given the latent random variables.

The latent Dirichlet allocation assumes something related to the (multivariate) Dirichlet distribution (which are commonly used as priors in Bayesian statistics) which is an assumption regarding the underlying generative process and therefore constraints the family of distributions of the joint according to

$$P(\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{z}, \boldsymbol{w}) = \prod_{i=1}^{K} P(\beta_i) \prod_{d=1}^{D} P(\theta_d) \left(\prod_{n=1}^{N} P\left(z_{d,n}|\theta_d\right) P(w_{d,n}|\beta_{1:K}, z_{d,n}) \right), \tag{A.1}$$

whereby the $\boldsymbol{\beta}$ vector denotes the different topics with each entry being a distribution of a set of vocabulary. The parameter(s) $\boldsymbol{\theta}$ are the topic proportions of the document and the topic assignments are given by $\boldsymbol{z}$. The observed words associated to document d are $\boldsymbol{w}_d$ with $w_{d,n}$ being the nth word observed in document d. To obtain the related topic(s) of a word (or a family of words), the posterior distribution must be computed $P(\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{z}|\boldsymbol{w}) = P(\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{z}, \boldsymbol{w})/P(\boldsymbol{w})$. As it is often the case for Bayesian statistic models, this problem is computational intractable; to "solve" it, one usually relies on efficient approximation methods to obtain the posterior distribution. The concrete approximation used to obtain to posterior distribution is irrelevant at this point. The interested reader is referred to the original publication in Ref. [28] (Section 2).

To summarize this short thematic excursion, the objective of topic modeling is to analyze hidden structures in extensive text archives for thematic structures. To quote the key insights as stated in Ref. [28]:

- *"[t]opic models are algorithms for discovering the main themes that pervade a large and otherwise unstructured collection of documents. Topic models can organize the collection according to the discovered themes[;]"*
- *"[t]opic modeling algorithms can be applied to massive collections of documents. Recent advances in this field allow us to analyze streaming collections, like you might find from a Web API[;]"*
- *"[t]opic modeling algorithms can be adapted to many kinds of data. among other applications, they have been used to find patterns in genetic data, images, and social networks[;]"*

The last point in the enumeration above is of particular relevance as the following paragraph will reveal a close resemblance between the examination of large bodies of text with respect to hidden structure in terms of themes *and* the distributions of jet-related observables with regard to quark and gluons. This may appear bizarre and non-intuitive at first; however, the analogy will become immediately apparent by appreciating *"[...] that the generative process for producing counts of words in a document is mathematically identical to producing jet observable distributions via [Equation 7.4]"* [29]. Referring to Section 7.2, the different mixtures of jets M_1 and M_2 are constructed such that each one has an underlying, initially unknown fraction of quark- and gluon-initialized jets; the only requirement is that there is one mixture enriched by one category each. The distribution of a certain jet observable (for instance, the number of tracks inside a jet or the width of a jet) can be thought of as being produced in a *generative process* in which quarks and gluon jets are sampled according to their (unknown) fraction in the respective mixture. In light of probabilistic topic modeling, a histogram bin of the

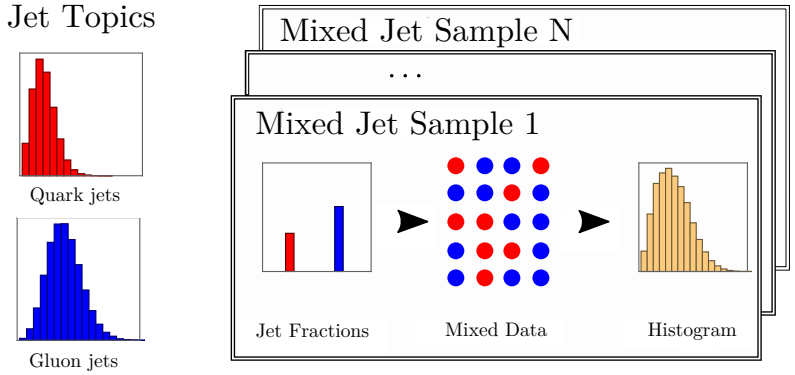


Fig. A.2: "The generation of mixed samples of quark and gluon jets, highlighting the correspondence with topic models. [...]" (adapted and modified from Ref. 29, Fig. 3, p. 3).

jet observable can be associated with a word in a text document because like the words in a text contain information about its topic, the value of the (JSS) observable assigned to a jet holds some information about its flavor. Similarly, the jet observable (as a whole) takes the function of the entire vocabulary of a text. The type of a jet (quark- or gluon-initialized) the topic of a text corpus – for this reason, the extracted distributions of quark and gluon jets according to Equation 7.19 are also referred to as "*jet topics*". The collection of histograms of the jet observable(s) can be regarded as the entire document formed by the total collection of words. Finally, the collection of *all* histograms can be considered as the *total* (text) corpus in the "text-document"-"particle-jet analogy". The correspondence between probabilistic topic modeling and jet-related observables just described is summarized in Table A.1. The analogy

Topic Model	**Jet Distributions**
Word	Histogram bin
Vocabulary	Jet observable
Topic	Type of jet (i.e., "*jet topic*")
Document	Histogram of jet observable(s)
Corpus	Collection of histograms

Table A.1: Analogy between probabilistic topic modeling and the probability distributions of jet observables. (Taken from Ref. 29, Table. 1, p. 4.)

between topic modeling and jet distributions regarding the generative process is conceptually illustrated in Figure A.2 that shows the procedure of "writing" (or sampling) a document in terms of jets, which correspond to the *corpus* mentioned in Table A.1. In this cartoon, the generative process starts from the two topic distributions of the underlying quarks and gluons (see Equation 7.19) and their corresponding fractions in both mixtures (keep in mind: this information is supposed to be extracted in the topic modeling approach). According to the fractions of quarks and gluons in M_1 and M_2, jets are sampled from their respective distributions for which the substructure observables are reconstructed. The set of reconstructed JSS variables is then used to populate histograms for each mixture of the jet sample.

B. Numerical Tables

Partition	Module	Start (Run)	End (Run)	Name
2015				
0	9	276262	284484	LBA10
3	20	276262	284484	EBC21
1	37	283155	283155	LBC38
2016				
1	4	306988	311481	LBC5
0	51	307856	311481	LBA52
2017				
2	2	325713	340453	EBA03
1	62	325713	340453	LBC63
2018				
0	28	351310	353395	LBA29
0	29	350361	353395	LBA30
0	31	353395	364292	LBA32
MC16a				
0	51	284500	284501	LBA52
1	4	284500	284501	LBC5
MC16d				
2	2	300000	300001	EBA3
1	62	300000	300001	LBC63
MC16e				
0	31	310000	310001	LBA32

Table B.1: Summary of disabled tile modules during Run 2 data-taking and in Run 2 simulation.

	Transverse Momentum p_{T} [GeV]				
	[250, 350) ($\times 10^{0}$)	[350, 500) ($\times 10^{-1}$)	[500, 700) ($\times 10^{-2}$)	[700, 1100) ($\times 10^{-3}$)	[1100, 4500) ($\times 10^{-4}$)
PFN $n_{\mathrm{trk}}^{>500\,\mathrm{MeV}}$	Triple-differential multijet cross section $\frac{1}{\sigma}\frac{\mathrm{d}\sigma^{3}}{\mathrm{d}p_{\mathrm{T}}\mathrm{d}\eta_{\mathrm{rel}}\mathrm{d}n_{\mathrm{trk}}}$ ($\times 10^{-6}$)				
[0.0, 5.0)	$64.56\,^{+0.21\,\oplus\,9.28}_{-0.21\,\oplus\,9.29}$ $81.10\,^{+0.26\,\oplus\,14.87}_{-0.26\,\oplus\,14.87}$	$119.81\,^{+0.32\,\oplus\,11.65}_{-0.32\,\oplus\,11.65}$ $155.92\,^{+0.39\,\oplus\,22.01}_{-0.39\,\oplus\,22.01}$	$115.38\,^{+0.17\,\oplus\,11.30}_{-0.17\,\oplus\,11.30}$ $150.79\,^{+0.14\,\oplus\,21.00}_{-0.14\,\oplus\,21.00}$	$75.85\,^{+0.23\,\oplus\,8.57}_{-0.23\,\oplus\,8.58}$ $95.97\,^{+0.20\,\oplus\,14.14}_{-0.20\,\oplus\,14.17}$	$4.11\,^{+0.03\,\oplus\,0.51}_{-0.03\,\oplus\,0.51}$ $4.51\,^{+0.03\,\oplus\,0.76}_{-0.03\,\oplus\,0.76}$
[5.0, 10.0)	$671.33\,^{+0.70\,\oplus\,89.14}_{-0.70\,\oplus\,89.15}$ $785.96\,^{+0.71\,\oplus\,116.48}_{-0.71\,\oplus\,116.49}$	$1364.81\,^{+1.31\,\oplus\,96.27}_{-1.31\,\oplus\,96.31}$ $1690.58\,^{+1.48\,\oplus\,145.34}_{-1.48\,\oplus\,145.38}$	$1506.63\,^{+0.51\,\oplus\,106.47}_{-0.51\,\oplus\,106.50}$ $1907.23\,^{+0.46\,\oplus\,160.96}_{-0.46\,\oplus\,160.99}$	$1148.23\,^{+0.69\,\oplus\,82.88}_{-0.69\,\oplus\,82.91}$ $1417.16\,^{+0.88\,\oplus\,120.68}_{-0.88\,\oplus\,120.96}$	$72.34\,^{+0.14\,\oplus\,6.01}_{-0.14\,\oplus\,6.03}$ $81.01\,^{+0.16\,\oplus\,8.26}_{-0.16\,\oplus\,8.26}$
[10.0, 15.0)	$1099.05\,^{+0.83\,\oplus\,142.23}_{-0.83\,\oplus\,142.24}$ $1115.70\,^{+0.77\,\oplus\,148.23}_{-0.77\,\oplus\,148.24}$	$2345.47\,^{+2.04\,\oplus\,133.73}_{-2.04\,\oplus\,133.80}$ $2533.85\,^{+1.74\,\oplus\,130.51}_{-1.74\,\oplus\,130.57}$	$2789.28\,^{+0.93\,\oplus\,159.80}_{-0.93\,\oplus\,159.85}$ $3177.29\,^{+0.87\,\oplus\,169.66}_{-0.87\,\oplus\,169.71}$	$2324.35\,^{+1.46\,\oplus\,120.50}_{-1.46\,\oplus\,120.53}$ $2682.76\,^{+1.29\,\oplus\,129.59}_{-1.29\,\oplus\,129.77}$	$168.81\,^{+0.36\,\oplus\,10.82}_{-0.36\,\oplus\,10.82}$ $184.70\,^{+0.45\,\oplus\,12.13}_{-0.45\,\oplus\,12.16}$
[15.0, 20.0)	$920.64\,^{+0.85\,\oplus\,118.39}_{-0.85\,\oplus\,118.39}$ $845.27\,^{+0.87\,\oplus\,110.95}_{-0.87\,\oplus\,110.95}$	$2173.57\,^{+1.56\,\oplus\,114.85}_{-1.56\,\oplus\,114.88}$ $2034.99\,^{+1.50\,\oplus\,89.31}_{-1.50\,\oplus\,89.33}$	$2735.40\,^{+0.71\,\oplus\,148.28}_{-0.71\,\oplus\,148.29}$ $2661.49\,^{+0.72\,\oplus\,127.73}_{-0.72\,\oplus\,127.77}$	$2302.18\,^{+1.30\,\oplus\,102.51}_{-1.30\,\oplus\,102.53}$ $2347.15\,^{+1.26\,\oplus\,97.38}_{-1.26\,\oplus\,97.50}$	$169.67\,^{+0.32\,\oplus\,11.24}_{-0.32\,\oplus\,11.23}$ $175.19\,^{+0.29\,\oplus\,9.85}_{-0.29\,\oplus\,9.88}$
[20.0, 25.0)	$480.20\,^{+0.89\,\oplus\,64.29}_{-0.89\,\oplus\,64.28}$ $422.16\,^{+0.55\,\oplus\,61.15}_{-0.55\,\oplus\,61.15}$	$1403.46\,^{+1.57\,\oplus\,90.32}_{-1.57\,\oplus\,90.31}$ $1216.89\,^{+1.66\,\oplus\,81.38}_{-1.66\,\oplus\,81.36}$	$2089.75\,^{+0.61\,\oplus\,123.72}_{-0.61\,\oplus\,123.69}$ $1819.00\,^{+0.69\,\oplus\,103.14}_{-0.69\,\oplus\,103.13}$	$1904.32\,^{+1.21\,\oplus\,85.91}_{-1.21\,\oplus\,85.90}$ $1721.96\,^{+1.01\,\oplus\,76.00}_{-1.01\,\oplus\,76.02}$	$140.09\,^{+0.38\,\oplus\,10.59}_{-0.38\,\oplus\,10.60}$ $135.18\,^{+0.28\,\oplus\,7.61}_{-0.28\,\oplus\,7.62}$
[25.0, 30.0)	$173.03\,^{+0.37\,\oplus\,26.07}_{-0.37\,\oplus\,26.06}$ $149.78\,^{+0.41\,\oplus\,26.55}_{-0.41\,\oplus\,26.55}$	$660.81\,^{+1.06\,\oplus\,69.46}_{-1.06\,\oplus\,69.45}$ $552.80\,^{+0.80\,\oplus\,67.36}_{-0.80\,\oplus\,67.34}$	$1237.54\,^{+0.47\,\oplus\,101.35}_{-0.47\,\oplus\,101.31}$ $1013.82\,^{+0.46\,\oplus\,92.14}_{-0.46\,\oplus\,92.11}$	$1319.26\,^{+1.03\,\oplus\,75.44}_{-1.03\,\oplus\,75.40}$ $1103.96\,^{+0.85\,\oplus\,62.04}_{-0.85\,\oplus\,62.04}$	$105.72\,^{+0.27\,\oplus\,8.90}_{-0.27\,\oplus\,8.90}$ $95.73\,^{+0.26\,\oplus\,5.87}_{-0.26\,\oplus\,5.87}$
[30.0, 35.0)	$47.91\,^{+0.20\,\oplus\,8.49}_{-0.20\,\oplus\,8.49}$ $41.01\,^{+0.22\,\oplus\,8.89}_{-0.22\,\oplus\,8.89}$	$241.83\,^{+0.60\,\oplus\,40.01}_{-0.60\,\oplus\,40.00}$ $200.42\,^{+0.63\,\oplus\,37.74}_{-0.63\,\oplus\,37.73}$	$587.95\,^{+0.38\,\oplus\,73.51}_{-0.38\,\oplus\,73.48}$ $468.32\,^{+0.33\,\oplus\,70.40}_{-0.33\,\oplus\,70.37}$	$772.46\,^{+1.00\,\oplus\,67.61}_{-1.00\,\oplus\,67.56}$ $615.56\,^{+0.73\,\oplus\,49.91}_{-0.73\,\oplus\,49.90}$	$72.55\,^{+0.25\,\oplus\,6.60}_{-0.25\,\oplus\,6.60}$ $62.35\,^{+0.34\,\oplus\,4.48}_{-0.34\,\oplus\,4.48}$
[35.0, 40.0)	$10.84\,^{+0.17\,\oplus\,2.35}_{-0.17\,\oplus\,2.35}$ $9.23\,^{+0.11\,\oplus\,2.50}_{-0.11\,\oplus\,2.50}$	$73.31\,^{+0.31\,\oplus\,16.27}_{-0.31\,\oplus\,16.27}$ $61.06\,^{+0.59\,\oplus\,14.59}_{-0.59\,\oplus\,14.59}$	$233.59\,^{+0.25\,\oplus\,41.27}_{-0.25\,\oplus\,41.26}$ $184.01\,^{+0.26\,\oplus\,39.71}_{-0.26\,\oplus\,39.69}$	$387.93\,^{+0.58\,\oplus\,52.93}_{-0.58\,\oplus\,52.89}$ $300.98\,^{+0.45\,\oplus\,38.67}_{-0.45\,\oplus\,38.64}$	$44.86\,^{+0.23\,\oplus\,4.58}_{-0.23\,\oplus\,4.57}$ $37.17\,^{+0.24\,\oplus\,3.39}_{-0.24\,\oplus\,3.38}$
[40.0, 45.0)	$2.13\,^{+0.07\,\oplus\,0.60}_{-0.07\,\oplus\,0.60}$ $1.71\,^{+0.07\,\oplus\,0.66}_{-0.07\,\oplus\,0.66}$	$19.10\,^{+0.25\,\oplus\,5.07}_{-0.25\,\oplus\,5.07}$ $15.94\,^{+0.29\,\oplus\,4.24}_{-0.29\,\oplus\,4.24}$	$81.17\,^{+0.15\,\oplus\,18.28}_{-0.15\,\oplus\,18.27}$ $63.36\,^{+0.11\,\oplus\,16.71}_{-0.11\,\oplus\,16.71}$	$170.96\,^{+0.31\,\oplus\,32.86}_{-0.31\,\oplus\,32.84}$ $130.29\,^{+0.33\,\oplus\,28.67}_{-0.33\,\oplus\,28.66}$	$24.95\,^{+0.15\,\oplus\,3.10}_{-0.15\,\oplus\,3.09}$ $20.13\,^{+0.14\,\oplus\,2.46}_{-0.14\,\oplus\,2.46}$
[45.0, 50.0)	$0.38\,^{+0.04\,\oplus\,0.14}_{-0.04\,\oplus\,0.14}$ $0.29\,^{+0.03\,\oplus\,0.17}_{-0.03\,\oplus\,0.17}$	$4.62\,^{+0.16\,\oplus\,1.38}_{-0.16\,\oplus\,1.38}$ $3.63\,^{+0.13\,\oplus\,1.03}_{-0.13\,\oplus\,1.03}$	$25.36\,^{+0.09\,\oplus\,6.87}_{-0.09\,\oplus\,6.87}$ $19.63\,^{+0.11\,\oplus\,5.72}_{-0.11\,\oplus\,5.72}$	$67.55\,^{+0.29\,\oplus\,16.93}_{-0.29\,\oplus\,16.92}$ $51.45\,^{+0.21\,\oplus\,17.25}_{-0.21\,\oplus\,17.24}$	$12.67\,^{+0.16\,\oplus\,2.04}_{-0.16\,\oplus\,2.03}$ $10.18\,^{+0.12\,\oplus\,1.70}_{-0.12\,\oplus\,1.70}$
[50.0, 80.0)	$0.01\,^{+0.00\,\oplus\,0.01}_{-0.00\,\oplus\,0.01}$ $0.01\,^{+0.00\,\oplus\,0.01}_{-0.00\,\oplus\,0.01}$	$0.21\,^{+0.02\,\oplus\,0.07}_{-0.02\,\oplus\,0.07}$ $0.16\,^{+0.02\,\oplus\,0.05}_{-0.02\,\oplus\,0.05}$	$1.67\,^{+0.01\,\oplus\,0.53}_{-0.01\,\oplus\,0.53}$ $1.28\,^{+0.02\,\oplus\,0.40}_{-0.02\,\oplus\,0.40}$	$6.41\,^{+0.04\,\oplus\,2.02}_{-0.04\,\oplus\,2.02}$ $4.79\,^{+0.05\,\oplus\,2.09}_{-0.05\,\oplus\,2.09}$	$1.77\,^{+0.02\,\oplus\,0.37}_{-0.02\,\oplus\,0.37}$ $1.35\,^{+0.02\,\oplus\,0.30}_{-0.02\,\oplus\,0.30}$

Table B.2: Numerical values of the unfolded inclusive relative tripe-differential dijet cross section as a function of the track multiplicity n_{trk} including upper and lower uncertainties. The two numbers that are provided for each multiplicity correspond to the relative event rate in central (top) and forward (bottom) rapidity category. Note the different scale factors that have been applied for the respective bins in transverse momentum.

	Transverse Momentum p_T [GeV]				
	[250, 350) ($\times 10^0$)	[350, 500) ($\times 10^{-1}$)	[500, 700) ($\times 10^{-2}$)	[700, 1100) ($\times 10^{-3}$)	[1100, 4500) ($\times 10^{-4}$)
SD multiplicity $n_\mathrm{SD}(k_t > 400\,\mathrm{MeV})$	Triple-differential multijet cross section $\frac{1}{\sigma}\frac{\mathrm{d}\sigma^3}{\mathrm{d}p_\mathrm{T}\mathrm{d}\eta_\mathrm{rel}\mathrm{d}n_\mathrm{SD}}$ ($\times 10^{-6}$)				
[0.0, 2.0)	$62.57\,^{+0.16\,\oplus\,8.53}_{-0.16\,\oplus\,8.53}$ $77.10\,^{+0.18\,\oplus\,12.89}_{-0.18\,\oplus\,12.89}$	$124.28\,^{+0.33\,\oplus\,10.76}_{-0.33\,\oplus\,10.77}$ $159.31\,^{+0.31\,\oplus\,15.50}_{-0.31\,\oplus\,15.50}$	$134.03\,^{+0.15\,\oplus\,9.84}_{-0.15\,\oplus\,9.85}$ $171.68\,^{+0.12\,\oplus\,17.95}_{-0.12\,\oplus\,17.96}$	$99.54\,^{+0.16\,\oplus\,8.13}_{-0.16\,\oplus\,8.15}$ $122.54\,^{+0.19\,\oplus\,13.46}_{-0.19\,\oplus\,13.46}$	$6.09\,^{+0.03\,\oplus\,0.65}_{-0.03\,\oplus\,0.65}$ $6.59\,^{+0.03\,\oplus\,0.76}_{-0.03\,\oplus\,0.76}$
[2.0, 3.0)	$243.15\,^{+0.53\,\oplus\,32.51}_{-0.53\,\oplus\,32.51}$ $288.48\,^{+0.48\,\oplus\,43.15}_{-0.48\,\oplus\,43.14}$	$500.65\,^{+0.96\,\oplus\,38.53}_{-0.96\,\oplus\,38.56}$ $619.08\,^{+0.87\,\oplus\,47.19}_{-0.87\,\oplus\,47.20}$	$572.53\,^{+0.28\,\oplus\,39.18}_{-0.28\,\oplus\,39.23}$ $716.38\,^{+0.37\,\oplus\,56.35}_{-0.37\,\oplus\,56.37}$	$456.45\,^{+0.53\,\oplus\,32.28}_{-0.53\,\oplus\,32.30}$ $551.33\,^{+0.50\,\oplus\,45.82}_{-0.50\,\oplus\,45.85}$	$30.14\,^{+0.09\,\oplus\,2.49}_{-0.09\,\oplus\,2.50}$ $33.21\,^{+0.11\,\oplus\,3.20}_{-0.11\,\oplus\,3.21}$
[3.0, 4.0)	$425.84\,^{+0.78\,\oplus\,56.50}_{-0.78\,\oplus\,56.51}$ $475.03\,^{+0.51\,\oplus\,65.80}_{-0.51\,\oplus\,65.79}$	$889.10\,^{+1.23\,\oplus\,59.76}_{-1.23\,\oplus\,59.79}$ $1053.22\,^{+1.18\,\oplus\,63.91}_{-1.18\,\oplus\,63.94}$	$1051.98\,^{+0.62\,\oplus\,68.13}_{-0.62\,\oplus\,68.17}$ $1275.83\,^{+0.51\,\oplus\,76.35}_{-0.51\,\oplus\,76.39}$	$877.16\,^{+0.65\,\oplus\,57.44}_{-0.65\,\oplus\,57.43}$ $1039.58\,^{+0.80\,\oplus\,66.01}_{-0.80\,\oplus\,66.06}$	$61.96\,^{+0.14\,\oplus\,4.19}_{-0.14\,\oplus\,4.20}$ $67.59\,^{+0.15\,\oplus\,5.56}_{-0.15\,\oplus\,5.56}$
[4.0, 5.0)	$586.34\,^{+0.60\,\oplus\,77.20}_{-0.60\,\oplus\,77.22}$ $611.65\,^{+0.57\,\oplus\,81.58}_{-0.57\,\oplus\,81.59}$	$1254.13\,^{+0.96\,\oplus\,69.74}_{-0.96\,\oplus\,69.78}$ $1382.50\,^{+1.49\,\oplus\,68.05}_{-1.49\,\oplus\,68.10}$	$1521.52\,^{+0.39\,\oplus\,92.55}_{-0.39\,\oplus\,92.60}$ $1741.63\,^{+0.54\,\oplus\,86.58}_{-0.54\,\oplus\,86.63}$	$1299.78\,^{+0.88\,\oplus\,83.60}_{-0.88\,\oplus\,83.59}$ $1488.81\,^{+0.84\,\oplus\,79.75}_{-0.84\,\oplus\,79.79}$	$97.39\,^{+0.25\,\oplus\,6.23}_{-0.25\,\oplus\,6.24}$ $104.43\,^{+0.16\,\oplus\,7.45}_{-0.16\,\oplus\,7.45}$
[5.0, 6.0)	$663.68\,^{+1.01\,\oplus\,86.48}_{-1.01\,\oplus\,86.48}$ $647.19\,^{+0.52\,\oplus\,85.13}_{-0.52\,\oplus\,85.15}$	$1475.38\,^{+1.79\,\oplus\,68.25}_{-1.79\,\oplus\,68.29}$ $1501.46\,^{+1.57\,\oplus\,61.76}_{-1.57\,\oplus\,61.80}$	$1842.14\,^{+0.55\,\oplus\,104.27}_{-0.55\,\oplus\,104.31}$ $1949.24\,^{+0.57\,\oplus\,88.93}_{-0.57\,\oplus\,88.95}$	$1596.38\,^{+1.26\,\oplus\,94.84}_{-1.26\,\oplus\,94.83}$ $1725.26\,^{+0.94\,\oplus\,82.25}_{-0.94\,\oplus\,82.25}$	$123.14\,^{+0.30\,\oplus\,8.63}_{-0.30\,\oplus\,8.63}$ $129.13\,^{+0.25\,\oplus\,7.99}_{-0.25\,\oplus\,7.99}$
[6.0, 7.0)	$612.38\,^{+0.72\,\oplus\,78.85}_{-0.72\,\oplus\,78.85}$ $567.13\,^{+0.61\,\oplus\,74.23}_{-0.61\,\oplus\,74.24}$	$1471.40\,^{+1.17\,\oplus\,65.32}_{-1.17\,\oplus\,65.34}$ $1391.97\,^{+1.04\,\oplus\,55.57}_{-1.04\,\oplus\,55.58}$	$1929.80\,^{+0.66\,\oplus\,102.49}_{-0.66\,\oplus\,102.51}$ $1871.29\,^{+0.50\,\oplus\,82.82}_{-0.50\,\oplus\,82.81}$	$1698.78\,^{+1.29\,\oplus\,84.92}_{-1.29\,\oplus\,84.92}$ $1703.12\,^{+0.77\,\oplus\,71.84}_{-0.77\,\oplus\,71.80}$	$131.50\,^{+0.37\,\oplus\,10.58}_{-0.37\,\oplus\,10.58}$ $132.75\,^{+0.28\,\oplus\,7.27}_{-0.28\,\oplus\,7.28}$
[7.0, 8.0)	$454.19\,^{+0.51\,\oplus\,57.67}_{-0.51\,\oplus\,57.67}$ $408.44\,^{+0.63\,\oplus\,53.96}_{-0.63\,\oplus\,53.96}$	$1233.29\,^{+1.28\,\oplus\,62.96}_{-1.28\,\oplus\,62.97}$ $1102.29\,^{+1.13\,\oplus\,53.46}_{-1.13\,\oplus\,53.46}$	$1758.94\,^{+0.59\,\oplus\,93.14}_{-0.59\,\oplus\,93.15}$ $1574.99\,^{+0.59\,\oplus\,72.56}_{-0.59\,\oplus\,72.56}$	$1606.22\,^{+1.08\,\oplus\,75.08}_{-1.08\,\oplus\,75.13}$ $1482.23\,^{+0.82\,\oplus\,63.86}_{-0.82\,\oplus\,63.78}$	$123.59\,^{+0.28\,\oplus\,10.79}_{-0.28\,\oplus\,10.79}$ $119.26\,^{+0.25\,\oplus\,6.20}_{-0.25\,\oplus\,6.20}$
[8.0, 9.0)	$267.68\,^{+0.44\,\oplus\,33.46}_{-0.44\,\oplus\,33.46}$ $237.89\,^{+0.46\,\oplus\,33.11}_{-0.46\,\oplus\,33.11}$	$847.46\,^{+1.02\,\oplus\,54.16}_{-1.02\,\oplus\,54.15}$ $731.71\,^{+0.84\,\oplus\,49.28}_{-0.84\,\oplus\,49.27}$	$1364.95\,^{+0.51\,\oplus\,84.21}_{-0.51\,\oplus\,84.20}$ $1151.10\,^{+0.53\,\oplus\,66.57}_{-0.53\,\oplus\,66.55}$	$1343.21\,^{+1.17\,\oplus\,75.04}_{-1.17\,\oplus\,75.13}$ $1149.71\,^{+0.99\,\oplus\,63.68}_{-0.99\,\oplus\,63.62}$	$105.46\,^{+0.24\,\oplus\,9.32}_{-0.24\,\oplus\,9.32}$ $96.71\,^{+0.26\,\oplus\,5.22}_{-0.26\,\oplus\,5.22}$
[9.0, 10.0)	$124.44\,^{+0.25\,\oplus\,15.72}_{-0.25\,\oplus\,15.72}$ $109.72\,^{+0.25\,\oplus\,17.16}_{-0.25\,\oplus\,17.16}$	$472.52\,^{+0.82\,\oplus\,37.46}_{-0.82\,\oplus\,37.45}$ $397.92\,^{+0.70\,\oplus\,37.28}_{-0.70\,\oplus\,37.27}$	$881.53\,^{+0.44\,\oplus\,73.11}_{-0.44\,\oplus\,73.08}$ $714.75\,^{+0.44\,\oplus\,62.06}_{-0.44\,\oplus\,62.03}$	$969.12\,^{+0.70\,\oplus\,79.55}_{-0.70\,\oplus\,79.58}$ $782.74\,^{+0.82\,\oplus\,65.11}_{-0.82\,\oplus\,65.09}$	$81.16\,^{+0.27\,\oplus\,7.55}_{-0.27\,\oplus\,7.55}$ $70.60\,^{+0.21\,\oplus\,5.06}_{-0.21\,\oplus\,5.06}$
[10.0, 12.0)	$29.53\,^{+0.11\,\oplus\,4.52}_{-0.11\,\oplus\,4.52}$ $25.70\,^{+0.09\,\oplus\,4.81}_{-0.09\,\oplus\,4.81}$	$143.36\,^{+0.48\,\oplus\,14.47}_{-0.48\,\oplus\,14.46}$ $118.33\,^{+0.39\,\oplus\,15.38}_{-0.39\,\oplus\,15.38}$	$333.59\,^{+0.17\,\oplus\,36.44}_{-0.17\,\oplus\,36.41}$ $263.39\,^{+0.17\,\oplus\,33.57}_{-0.17\,\oplus\,33.55}$	$444.99\,^{+0.58\,\oplus\,53.17}_{-0.58\,\oplus\,53.14}$ $343.50\,^{+0.51\,\oplus\,40.91}_{-0.51\,\oplus\,40.87}$	$44.34\,^{+0.19\,\oplus\,5.71}_{-0.19\,\oplus\,5.70}$ $36.29\,^{+0.18\,\oplus\,4.47}_{-0.18\,\oplus\,4.47}$
[12.0, 15.0)	$1.24\,^{+0.02\,\oplus\,0.30}_{-0.02\,\oplus\,0.30}$ $1.07\,^{+0.02\,\oplus\,0.25}_{-0.02\,\oplus\,0.25}$	$9.18\,^{+0.09\,\oplus\,1.25}_{-0.09\,\oplus\,1.25}$ $7.60\,^{+0.06\,\oplus\,1.41}_{-0.06\,\oplus\,1.41}$	$32.01\,^{+0.06\,\oplus\,4.49}_{-0.06\,\oplus\,4.48}$ $24.83\,^{+0.06\,\oplus\,4.32}_{-0.06\,\oplus\,4.31}$	$61.96\,^{+0.17\,\oplus\,9.50}_{-0.17\,\oplus\,9.49}$ $46.06\,^{+0.16\,\oplus\,7.11}_{-0.16\,\oplus\,7.10}$	$9.03\,^{+0.08\,\oplus\,1.82}_{-0.08\,\oplus\,1.82}$ $7.04\,^{+0.09\,\oplus\,1.45}_{-0.09\,\oplus\,1.45}$
[15.0, 20.0)	$0.00\,^{+0.00\,\oplus\,0.00}_{-0.00\,\oplus\,0.00}$ $0.00\,^{+0.00\,\oplus\,0.00}_{-0.00\,\oplus\,0.00}$	$0.03\,^{+0.00\,\oplus\,0.01}_{-0.00\,\oplus\,0.01}$ $0.02\,^{+0.00\,\oplus\,0.01}_{-0.00\,\oplus\,0.01}$	$0.21\,^{+0.00\,\oplus\,0.04}_{-0.00\,\oplus\,0.04}$ $0.16\,^{+0.00\,\oplus\,0.04}_{-0.00\,\oplus\,0.04}$	$0.78\,^{+0.01\,\oplus\,0.14}_{-0.01\,\oplus\,0.14}$ $0.58\,^{+0.01\,\oplus\,0.11}_{-0.01\,\oplus\,0.11}$	$0.24\,^{+0.01\,\oplus\,0.07}_{-0.01\,\oplus\,0.07}$ $0.18\,^{+0.01\,\oplus\,0.06}_{-0.01\,\oplus\,0.06}$

Table B.3: Numerical values of the unfolded inclusive relative tripe-differential dijet cross section as a function of the SD multiplicity n_SD with $k_t > 400\,\mathrm{MeV}$ including upper and lower uncertainties. The two numbers that are provided for each multiplicity correspond to the relative event rate in central (top) and forward (bottom) rapidity category. Note the different scale factors that have been applied for the respective bins in transverse momentum.

	Transverse Momentum p_{T} [GeV]				
	[250, 350) ($\times 10^{0}$)	[350, 500) ($\times 10^{-1}$)	[500, 700) ($\times 10^{-2}$)	[700, 1100) ($\times 10^{-3}$)	[1100, 4500) ($\times 10^{-4}$)
SD multiplicity $n_{\mathrm{SD}}(z_{\mathrm{cut}}=0.1,\beta=1,\theta_{\mathrm{cut}}=0.002)$	Triple-differential multijet cross section $\frac{1}{\sigma}\frac{\mathrm{d}\sigma^3}{\mathrm{d}p_{\mathrm{T}}\mathrm{d}\eta_{\mathrm{rel}}\mathrm{d}n_{\mathrm{SD}}}$ ($\times 10^{-6}$)				
[0.0, 3.0)	$83.72^{+0.21\oplus 11.52}_{-0.21\oplus 11.52}$ $100.78^{+0.16\oplus 16.19}_{-0.16\oplus 16.20}$	$178.16^{+0.35\oplus 13.36}_{-0.35\oplus 13.37}$ $224.73^{+0.35\oplus 21.82}_{-0.35\oplus 21.84}$	$215.65^{+0.13\oplus 17.04}_{-0.13\oplus 17.06}$ $275.23^{+0.14\oplus 26.60}_{-0.14\oplus 26.63}$	$192.22^{+0.27\oplus 13.44}_{-0.27\oplus 13.47}$ $237.75^{+0.26\oplus 22.22}_{-0.26\oplus 22.24}$	$17.09^{+0.08\oplus 1.39}_{-0.08\oplus 1.39}$ $19.04^{+0.06\oplus 1.69}_{-0.06\oplus 1.69}$
[3.0, 4.0)	$462.95^{+0.51\oplus 60.81}_{-0.51\oplus 60.82}$ $510.09^{+0.68\oplus 72.87}_{-0.68\oplus 72.89}$	$986.20^{+1.55\oplus 61.05}_{-1.55\oplus 61.11}$ $1148.73^{+1.05\oplus 80.97}_{-1.05\oplus 81.10}$	$1202.05^{+0.68\oplus 82.47}_{-0.68\oplus 82.54}$ $1429.70^{+0.45\oplus 102.95}_{-0.45\oplus 103.08}$	$1053.47^{+0.96\oplus 66.26}_{-0.96\oplus 66.33}$ $1236.41^{+0.69\oplus 86.22}_{-0.69\oplus 86.33}$	$86.32^{+0.19\oplus 6.01}_{-0.19\oplus 6.01}$ $94.45^{+0.25\oplus 6.42}_{-0.25\oplus 6.43}$
[4.0, 5.0)	$743.96^{+0.91\oplus 96.29}_{-0.91\oplus 96.30}$ $769.47^{+0.85\oplus 102.21}_{-0.85\oplus 102.23}$	$1635.14^{+1.29\oplus 81.73}_{-1.29\oplus 81.81}$ $1765.09^{+1.47\oplus 93.97}_{-1.47\oplus 94.07}$	$2030.16^{+1.08\oplus 114.45}_{-1.08\oplus 114.53}$ $2246.46^{+0.66\oplus 119.57}_{-0.66\oplus 119.68}$	$1771.15^{+1.26\oplus 99.64}_{-1.26\oplus 99.68}$ $1958.73^{+1.13\oplus 105.91}_{-1.13\oplus 106.02}$	$139.13^{+0.25\oplus 8.70}_{-0.25\oplus 8.71}$ $146.85^{+0.24\oplus 8.13}_{-0.24\oplus 8.13}$
[5.0, 6.0)	$842.59^{+0.78\oplus 109.53}_{-0.78\oplus 109.53}$ $826.59^{+0.97\oplus 106.51}_{-0.97\oplus 106.52}$	$1958.12^{+1.62\oplus 81.68}_{-1.62\oplus 81.72}$ $1968.69^{+1.26\oplus 87.48}_{-1.26\oplus 87.49}$	$2529.44^{+0.79\oplus 112.43}_{-0.79\oplus 112.48}$ $2584.89^{+0.68\oplus 106.08}_{-0.68\oplus 106.12}$	$2222.43^{+0.94\oplus 106.91}_{-0.94\oplus 106.92}$ $2290.68^{+1.24\oplus 102.25}_{-1.24\oplus 102.29}$	$168.32^{+0.19\oplus 10.14}_{-0.19\oplus 10.15}$ $170.47^{+0.29\oplus 8.56}_{-0.29\oplus 8.56}$
[6.0, 7.0)	$694.67^{+0.72\oplus 90.31}_{-0.72\oplus 90.31}$ $655.26^{+0.69\oplus 84.79}_{-0.69\oplus 84.79}$	$1754.87^{+1.35\oplus 69.17}_{-1.35\oplus 69.16}$ $1664.35^{+1.32\oplus 67.95}_{-1.32\oplus 67.94}$	$2416.86^{+0.82\oplus 100.40}_{-0.82\oplus 100.40}$ $2291.57^{+0.69\oplus 90.38}_{-0.69\oplus 90.37}$	$2182.96^{+1.23\oplus 98.07}_{-1.23\oplus 98.05}$ $2088.43^{+1.09\oplus 90.14}_{-1.09\oplus 90.13}$	$163.07^{+0.22\oplus 9.70}_{-0.22\oplus 9.70}$ $157.45^{+0.21\oplus 8.66}_{-0.21\oplus 8.66}$
[7.0, 8.0)	$421.50^{+0.64\oplus 54.77}_{-0.64\oplus 54.77}$ $389.60^{+0.46\oplus 53.09}_{-0.46\oplus 53.08}$	$1186.00^{+1.02\oplus 52.62}_{-1.02\oplus 52.57}$ $1075.39^{+1.04\oplus 54.55}_{-1.04\oplus 54.51}$	$1787.11^{+0.57\oplus 90.97}_{-0.57\oplus 90.93}$ $1594.98^{+0.64\oplus 79.65}_{-0.64\oplus 79.61}$	$1714.33^{+0.84\oplus 86.85}_{-0.84\oplus 86.80}$ $1532.51^{+0.95\oplus 79.95}_{-0.95\oplus 79.88}$	$130.42^{+0.30\oplus 7.85}_{-0.30\oplus 7.85}$ $119.99^{+0.25\oplus 8.84}_{-0.25\oplus 8.84}$
[8.0, 9.0)	$192.70^{+0.31\oplus 25.94}_{-0.31\oplus 25.93}$ $173.38^{+0.37\oplus 26.34}_{-0.37\oplus 26.34}$	$607.35^{+0.73\oplus 37.81}_{-0.73\oplus 37.75}$ $532.13^{+0.88\oplus 41.78}_{-0.88\oplus 41.74}$	$1023.44^{+0.40\oplus 76.53}_{-0.40\oplus 76.45}$ $871.26^{+0.41\oplus 69.16}_{-0.41\oplus 69.10}$	$1071.66^{+0.79\oplus 80.45}_{-0.79\oplus 80.36}$ $904.72^{+0.81\oplus 71.25}_{-0.81\oplus 71.14}$	$86.48^{+0.21\oplus 6.87}_{-0.21\oplus 6.87}$ $76.29^{+0.24\oplus 8.54}_{-0.24\oplus 8.54}$
[9.0, 12.0)	$28.35^{+0.13\oplus 4.26}_{-0.13\oplus 4.26}$ $24.73^{+0.12\oplus 4.42}_{-0.12\oplus 4.41}$	$107.03^{+0.27\oplus 10.61}_{-0.27\oplus 10.59}$ $90.35^{+0.20\oplus 11.16}_{-0.20\oplus 11.14}$	$214.85^{+0.16\oplus 23.75}_{-0.16\oplus 23.71}$ $174.86^{+0.13\oplus 23.77}_{-0.13\oplus 23.74}$	$264.13^{+0.31\oplus 35.67}_{-0.31\oplus 35.62}$ $211.93^{+0.39\oplus 28.35}_{-0.39\oplus 28.30}$	$24.80^{+0.10\oplus 3.91}_{-0.10\oplus 3.91}$ $21.00^{+0.09\oplus 3.91}_{-0.09\oplus 3.91}$
[12.0, 20.0)	$0.07^{+0.00\oplus 0.01}_{-0.00\oplus 0.01}$ $0.05^{+0.00\oplus 0.01}_{-0.00\oplus 0.01}$	$0.39^{+0.01\oplus 0.06}_{-0.01\oplus 0.06}$ $0.30^{+0.00\oplus 0.05}_{-0.00\oplus 0.05}$	$1.19^{+0.01\oplus 0.18}_{-0.01\oplus 0.18}$ $0.90^{+0.00\oplus 0.19}_{-0.00\oplus 0.19}$	$2.20^{+0.01\oplus 0.49}_{-0.01\oplus 0.49}$ $1.67^{+0.01\oplus 0.35}_{-0.01\oplus 0.35}$	$0.32^{+0.01\oplus 0.09}_{-0.01\oplus 0.09}$ $0.25^{+0.01\oplus 0.08}_{-0.01\oplus 0.08}$

Table B.4: Numerical values of the unfolded inclusive relative tripe-differential dijet cross section as a function of the SD multiplicity $n_{\mathrm{SD}}(z_{\mathrm{cut}}=0.1,\beta=1,\theta_{\mathrm{cut}}=0.002)$ including upper and lower uncertainties. The two numbers that are provided for each multiplicity correspond to the relative event rate in central (top) and forward (bottom) rapidity category. Note the different scale factors that have been applied for the respective bins in transverse momentum.

	Transverse Momentum p_{T} [GeV]				
	[250, 350) ($\times 10^{0}$)	[350, 500) ($\times 10^{-1}$)	[500, 700) ($\times 10^{-2}$)	[700, 1100) ($\times 10^{-3}$)	[1100, 4500) ($\times 10^{-4}$)
Track multiplicity z_{PFN}	Triple-differential multijet cross section $\frac{1}{\sigma}\frac{\mathrm{d}\sigma^3}{\mathrm{d}p_{\mathrm{T}}\mathrm{d}\eta_{\mathrm{rel}}\mathrm{d}z_{\mathrm{PFN}}}$ ($\times 10^{-6}$)				
[0.0, 0.04)	$24.30^{+0.14\oplus 5.26}_{-0.14\oplus 5.26}$ $25.21^{+0.13\oplus 4.78}_{-0.13\oplus 4.79}$	$34.27^{+0.23\oplus 4.73}_{-0.23\oplus 4.74}$ $39.33^{+0.17\oplus 6.59}_{-0.17\oplus 6.60}$	$31.47^{+0.10\oplus 4.96}_{-0.10\oplus 4.96}$ $37.14^{+0.08\oplus 5.35}_{-0.08\oplus 5.35}$	$26.05^{+0.12\oplus 4.33}_{-0.12\oplus 4.33}$ $30.82^{+0.10\oplus 4.39}_{-0.10\oplus 4.39}$	$2.84^{+0.04\oplus 0.62}_{-0.04\oplus 0.62}$ $3.15^{+0.02\oplus 0.72}_{-0.02\oplus 0.72}$
[0.04, 0.08)	$35.27^{+0.14\oplus 6.70}_{-0.14\oplus 6.70}$ $38.25^{+0.13\oplus 6.73}_{-0.13\oplus 6.73}$	$54.74^{+0.24\oplus 6.48}_{-0.24\oplus 6.49}$ $62.92^{+0.16\oplus 8.83}_{-0.16\oplus 8.83}$	$55.52^{+0.11\oplus 7.02}_{-0.11\oplus 7.03}$ $65.16^{+0.11\oplus 8.23}_{-0.11\oplus 8.24}$	$52.43^{+0.15\oplus 7.32}_{-0.15\oplus 7.32}$ $60.82^{+0.14\oplus 7.40}_{-0.14\oplus 7.40}$	$6.19^{+0.04\oplus 1.09}_{-0.04\oplus 1.09}$ $6.76^{+0.05\oplus 1.08}_{-0.05\oplus 1.08}$
[0.08, 0.1)	$52.57^{+0.19\oplus 8.41}_{-0.19\oplus 8.41}$ $58.71^{+0.15\oplus 9.45}_{-0.15\oplus 9.46}$	$90.05^{+0.28\oplus 8.79}_{-0.28\oplus 8.79}$ $104.99^{+0.19\oplus 11.46}_{-0.19\oplus 11.47}$	$101.43^{+0.16\oplus 9.73}_{-0.16\oplus 9.73}$ $119.96^{+0.18\oplus 12.23}_{-0.18\oplus 12.23}$	$106.57^{+0.26\oplus 11.59}_{-0.26\oplus 11.59}$ $124.57^{+0.25\oplus 11.75}_{-0.25\oplus 11.75}$	$13.40^{+0.08\oplus 1.69}_{-0.08\oplus 1.69}$ $14.47^{+0.06\oplus 1.51}_{-0.06\oplus 1.51}$
[0.1, 0.15)	$100.80^{+0.24\oplus 13.61}_{-0.24\oplus 13.61}$ $117.36^{+0.18\oplus 17.20}_{-0.18\oplus 17.20}$	$226.67^{+0.47\oplus 17.75}_{-0.47\oplus 17.73}$ $277.51^{+0.36\oplus 22.57}_{-0.36\oplus 22.57}$	$348.31^{+0.28\oplus 24.70}_{-0.28\oplus 24.70}$ $432.49^{+0.21\oplus 31.75}_{-0.21\oplus 31.75}$	$444.14^{+0.45\oplus 35.06}_{-0.45\oplus 35.04}$ $537.82^{+0.48\oplus 35.75}_{-0.48\oplus 35.74}$	$58.06^{+0.14\oplus 4.73}_{-0.14\oplus 4.73}$ $63.72^{+0.20\oplus 4.27}_{-0.20\oplus 4.27}$
[0.15, 0.2)	$249.36^{+0.44\oplus 30.18}_{-0.44\oplus 30.18}$ $303.21^{+0.53\oplus 41.02}_{-0.53\oplus 41.02}$	$730.80^{+1.15\oplus 48.41}_{-1.15\oplus 48.42}$ $928.34^{+1.04\oplus 60.19}_{-1.04\oplus 60.19}$	$1280.15^{+0.51\oplus 75.35}_{-0.51\oplus 75.38}$ $1623.02^{+0.51\oplus 87.08}_{-0.51\oplus 87.07}$	$1527.76^{+1.24\oplus 93.92}_{-1.24\oplus 93.86}$ $1852.45^{+0.89\oplus 91.78}_{-0.89\oplus 91.73}$	$160.02^{+0.36\oplus 9.19}_{-0.36\oplus 9.19}$ $176.15^{+0.33\oplus 8.57}_{-0.33\oplus 8.56}$
[0.2, 0.25)	$293.98^{+0.67\oplus 34.68}_{-0.67\oplus 34.68}$ $347.28^{+0.34\oplus 44.97}_{-0.34\oplus 44.97}$	$806.27^{+1.12\oplus 50.17}_{-1.12\oplus 50.23}$ $984.75^{+0.88\oplus 59.09}_{-0.88\oplus 59.08}$	$1222.64^{+0.58\oplus 69.29}_{-0.58\oplus 69.31}$ $1486.44^{+0.54\oplus 69.46}_{-0.54\oplus 69.45}$	$1233.32^{+0.62\oplus 70.09}_{-0.62\oplus 70.07}$ $1440.80^{+0.86\oplus 64.11}_{-0.86\oplus 64.09}$	$105.67^{+0.20\oplus 5.58}_{-0.20\oplus 5.58}$ $113.77^{+0.26\oplus 5.06}_{-0.26\oplus 5.05}$
[0.25, 0.3)	$251.80^{+0.44\oplus 30.61}_{-0.44\oplus 30.61}$ $281.90^{+0.35\oplus 36.13}_{-0.35\oplus 36.13}$	$637.00^{+0.86\oplus 38.90}_{-0.86\oplus 38.89}$ $738.81^{+0.69\oplus 44.23}_{-0.69\oplus 44.23}$	$903.49^{+0.46\oplus 51.37}_{-0.46\oplus 51.38}$ $1043.20^{+0.43\oplus 46.87}_{-0.43\oplus 46.87}$	$865.89^{+0.57\oplus 48.39}_{-0.57\oplus 48.39}$ $968.25^{+0.60\oplus 42.36}_{-0.60\oplus 42.36}$	$69.99^{+0.13\oplus 3.78}_{-0.13\oplus 3.78}$ $73.59^{+0.11\oplus 3.34}_{-0.11\oplus 3.34}$
[0.3, 0.35)	$230.84^{+0.32\oplus 28.96}_{-0.32\oplus 28.96}$ $247.75^{+0.39\oplus 32.06}_{-0.39\oplus 32.06}$	$561.68^{+0.68\oplus 32.40}_{-0.68\oplus 32.36}$ $621.13^{+0.58\oplus 35.51}_{-0.58\oplus 35.51}$	$766.20^{+0.35\oplus 43.36}_{-0.35\oplus 43.36}$ $846.20^{+0.39\oplus 36.18}_{-0.39\oplus 36.19}$	$708.60^{+0.42\oplus 37.32}_{-0.42\oplus 37.32}$ $762.73^{+0.47\oplus 32.94}_{-0.47\oplus 32.94}$	$54.97^{+0.09\oplus 2.92}_{-0.09\oplus 2.92}$ $56.52^{+0.12\oplus 2.71}_{-0.12\oplus 2.71}$
[0.35, 0.4)	$225.40^{+0.39\oplus 29.13}_{-0.39\oplus 29.13}$ $233.57^{+0.29\oplus 30.92}_{-0.29\oplus 30.92}$	$530.94^{+0.43\oplus 26.93}_{-0.43\oplus 26.91}$ $563.90^{+0.60\oplus 28.66}_{-0.60\oplus 28.67}$	$700.95^{+0.37\oplus 37.70}_{-0.37\oplus 37.70}$ $742.37^{+0.31\oplus 29.88}_{-0.31\oplus 29.89}$	$629.00^{+0.45\oplus 30.37}_{-0.45\oplus 30.37}$ $652.34^{+0.37\oplus 27.98}_{-0.37\oplus 27.99}$	$47.19^{+0.10\oplus 2.43}_{-0.10\oplus 2.43}$ $47.38^{+0.10\oplus 2.44}_{-0.10\oplus 2.44}$
[0.4, 0.5)	$232.42^{+0.31\oplus 31.07}_{-0.31\oplus 31.07}$ $228.79^{+0.27\oplus 31.03}_{-0.27\oplus 31.03}$	$525.34^{+0.51\oplus 23.05}_{-0.51\oplus 23.05}$ $527.85^{+0.57\oplus 24.78}_{-0.57\oplus 24.79}$	$671.84^{+0.27\oplus 33.50}_{-0.27\oplus 33.50}$ $669.60^{+0.25\oplus 26.39}_{-0.25\oplus 26.39}$	$585.38^{+0.41\oplus 27.26}_{-0.41\oplus 27.27}$ $573.69^{+0.35\oplus 25.12}_{-0.35\oplus 25.13}$	$42.19^{+0.11\oplus 2.21}_{-0.11\oplus 2.21}$ $40.70^{+0.10\oplus 2.26}_{-0.10\oplus 2.26}$
[0.5, 0.6)	$270.05^{+0.23\oplus 37.01}_{-0.23\oplus 37.02}$ $250.10^{+0.28\oplus 34.28}_{-0.28\oplus 34.29}$	$591.25^{+0.53\oplus 24.28}_{-0.53\oplus 24.30}$ $550.96^{+0.51\oplus 26.89}_{-0.51\oplus 26.90}$	$735.44^{+0.23\oplus 35.61}_{-0.23\oplus 35.61}$ $676.22^{+0.26\oplus 27.63}_{-0.26\oplus 27.64}$	$623.30^{+0.43\oplus 30.21}_{-0.43\oplus 30.21}$ $563.92^{+0.35\oplus 25.96}_{-0.35\oplus 25.97}$	$42.81^{+0.10\oplus 2.51}_{-0.10\oplus 2.51}$ $38.69^{+0.10\oplus 2.33}_{-0.10\oplus 2.33}$
[0.6, 0.65)	$323.21^{+0.39\oplus 44.54}_{-0.39\oplus 44.55}$ $289.72^{+0.34\oplus 39.22}_{-0.34\oplus 39.23}$	$708.46^{+0.63\oplus 30.73}_{-0.63\oplus 30.76}$ $627.39^{+0.52\oplus 33.86}_{-0.52\oplus 33.88}$	$871.62^{+0.40\oplus 44.42}_{-0.40\oplus 44.42}$ $752.96^{+0.35\oplus 33.24}_{-0.35\oplus 33.25}$	$722.92^{+0.48\oplus 38.00}_{-0.48\oplus 38.00}$ $613.65^{+0.48\oplus 30.02}_{-0.48\oplus 30.03}$	$47.44^{+0.11\oplus 3.29}_{-0.11\oplus 3.29}$ $40.56^{+0.10\oplus 2.62}_{-0.10\oplus 2.62}$
[0.65, 0.7)	$367.24^{+0.44\oplus 49.59}_{-0.44\oplus 49.60}$ $323.51^{+0.47\oplus 42.70}_{-0.47\oplus 42.70}$	$825.72^{+0.87\oplus 40.06}_{-0.87\oplus 40.07}$ $708.61^{+0.75\oplus 41.61}_{-0.75\oplus 41.63}$	$1019.43^{+0.37\oplus 57.03}_{-0.37\oplus 57.03}$ $842.90^{+0.36\oplus 41.26}_{-0.36\oplus 41.26}$	$829.11^{+0.69\oplus 46.39}_{-0.69\oplus 46.38}$ $669.70^{+0.60\oplus 34.60}_{-0.60\oplus 34.61}$	$51.27^{+0.13\oplus 3.99}_{-0.13\oplus 3.99}$ $41.97^{+0.12\oplus 2.86}_{-0.12\oplus 2.86}$
[0.7, 0.75)	$385.30^{+0.62\oplus 50.54}_{-0.62\oplus 50.54}$ $333.63^{+0.56\oplus 44.28}_{-0.56\oplus 44.29}$	$918.19^{+1.20\oplus 49.02}_{-1.20\oplus 49.02}$ $764.76^{+1.12\oplus 47.72}_{-1.12\oplus 47.72}$	$1152.39^{+0.41\oplus 70.21}_{-0.41\oplus 70.20}$ $914.04^{+0.39\oplus 51.82}_{-0.39\oplus 51.82}$	$911.79^{+0.79\oplus 51.04}_{-0.79\oplus 51.03}$ $702.74^{+0.67\oplus 38.69}_{-0.67\oplus 38.69}$	$51.84^{+0.15\oplus 4.10}_{-0.15\oplus 4.10}$ $40.52^{+0.19\oplus 2.92}_{-0.19\oplus 2.92}$
[0.75, 0.8)	$285.10^{+0.56\oplus 37.43}_{-0.56\oplus 37.44}$ $246.32^{+0.53\oplus 34.75}_{-0.53\oplus 34.75}$	$758.46^{+0.84\oplus 44.03}_{-0.84\oplus 44.03}$ $621.36^{+0.94\oplus 41.56}_{-0.94\oplus 41.56}$	$999.00^{+0.65\oplus 65.37}_{-0.65\oplus 65.36}$ $771.11^{+0.48\oplus 54.94}_{-0.48\oplus 54.93}$	$771.24^{+0.62\oplus 41.98}_{-0.62\oplus 41.97}$ $573.27^{+0.74\oplus 34.24}_{-0.74\oplus 34.23}$	$39.30^{+0.15\oplus 2.99}_{-0.15\oplus 2.99}$ $29.75^{+0.19\oplus 2.30}_{-0.19\oplus 2.30}$
[0.8, 0.85)	$113.37^{+0.33\oplus 15.49}_{-0.33\oplus 15.50}$ $97.87^{+0.39\oplus 14.87}_{-0.39\oplus 14.87}$	$326.50^{+0.83\oplus 20.67}_{-0.83\oplus 20.67}$ $267.58^{+0.59\oplus 18.68}_{-0.59\oplus 18.69}$	$435.93^{+0.38\oplus 30.23}_{-0.38\oplus 30.23}$ $334.76^{+0.37\oplus 29.89}_{-0.37\oplus 29.89}$	$323.48^{+0.49\oplus 17.89}_{-0.49\oplus 17.88}$ $236.93^{+0.47\oplus 15.92}_{-0.47\oplus 15.93}$	$14.62^{+0.08\oplus 1.10}_{-0.08\oplus 1.10}$ $11.02^{+0.09\oplus 0.96}_{-0.09\oplus 0.96}$
[0.85, 0.9)	$19.28^{+0.14\oplus 2.89}_{-0.14\oplus 2.89}$ $17.14^{+0.14\oplus 2.80}_{-0.14\oplus 2.80}$	$50.37^{+0.26\oplus 3.81}_{-0.26\oplus 3.81}$ $43.00^{+0.30\oplus 3.08}_{-0.30\oplus 3.09}$	$64.41^{+0.13\oplus 4.83}_{-0.13\oplus 4.83}$ $51.68^{+0.11\oplus 5.32}_{-0.11\oplus 5.32}$	$47.54^{+0.18\oplus 3.02}_{-0.18\oplus 3.02}$ $36.44^{+0.21\oplus 3.12}_{-0.21\oplus 3.12}$	$2.30^{+0.03\oplus 0.19}_{-0.03\oplus 0.19}$ $1.86^{+0.03\oplus 0.19}_{-0.03\oplus 0.19}$
[0.9, 0.93)	$5.28^{+0.04\oplus 0.94}_{-0.04\oplus 0.94}$ $4.82^{+0.03\oplus 0.88}_{-0.03\oplus 0.88}$	$10.21^{+0.07\oplus 1.08}_{-0.07\oplus 1.08}$ $9.45^{+0.06\oplus 0.86}_{-0.06\oplus 0.86}$	$12.07^{+0.04\oplus 1.10}_{-0.04\oplus 1.10}$ $10.84^{+0.03\oplus 1.28}_{-0.03\oplus 1.28}$	$9.49^{+0.04\oplus 0.91}_{-0.04\oplus 0.91}$ $8.41^{+0.04\oplus 1.02}_{-0.04\oplus 1.02}$	$0.59^{+0.01\oplus 0.06}_{-0.01\oplus 0.06}$ $0.54^{+0.01\oplus 0.07}_{-0.01\oplus 0.07}$
[0.93, 0.96)	$4.35^{+0.05\oplus 0.95}_{-0.05\oplus 0.95}$ $3.77^{+0.03\oplus 0.79}_{-0.03\oplus 0.79}$	$7.47^{+0.07\oplus 1.12}_{-0.07\oplus 1.12}$ $6.68^{+0.05\oplus 0.97}_{-0.05\oplus 0.97}$	$8.20^{+0.03\oplus 1.00}_{-0.03\oplus 1.00}$ $7.61^{+0.03\oplus 1.11}_{-0.03\oplus 1.11}$	$6.26^{+0.04\oplus 1.02}_{-0.04\oplus 1.02}$ $5.81^{+0.04\oplus 0.98}_{-0.04\oplus 0.98}$	$0.40^{+0.01\oplus 0.06}_{-0.01\oplus 0.06}$ $0.35^{+0.01\oplus 0.06}_{-0.01\oplus 0.06}$
[0.96, 1.0)	$4.04^{+0.05\oplus 1.07}_{-0.05\oplus 1.07}$ $3.47^{+0.05\oplus 0.84}_{-0.05\oplus 0.84}$	$6.28^{+0.09\oplus 1.24}_{-0.09\oplus 1.24}$ $5.84^{+0.06\oplus 1.31}_{-0.06\oplus 1.31}$	$6.21^{+0.04\oplus 1.00}_{-0.04\oplus 1.00}$ $5.89^{+0.03\oplus 1.10}_{-0.03\oplus 1.10}$	$4.53^{+0.05\oplus 1.14}_{-0.05\oplus 1.14}$ $4.29^{+0.04\oplus 0.93}_{-0.04\oplus 0.93}$	$0.27^{+0.01\oplus 0.06}_{-0.01\oplus 0.06}$ $0.24^{+0.01\oplus 0.05}_{-0.01\oplus 0.05}$

Table B.5: Numerical values of the unfolded inclusive relative tripe-differential dijet cross section as a function of the NN-score as predicted by the PFN including upper and lower uncertainties. The two numbers that are provided for each score interval correspond to the relative event rate in central (top) and forward (bottom) rapidity category. Note the different scale factors that have been applied for the respective bins in transverse momentum.

	Transverse Momentum p_T [GeV]				
	[250, 350) ($\times 10^0$)	[350, 500) ($\times 10^{-1}$)	[500, 700) ($\times 10^{-2}$)	[700, 1100) ($\times 10^{-3}$)	[1100, 4500) ($\times 10^{-4}$)
EFN z_EFN	Triple-differential multijet cross section $\frac{1}{\sigma}\frac{\mathrm{d}\sigma^3}{\mathrm{d}p_\mathrm{T}\mathrm{d}\eta_\mathrm{rel}\mathrm{d}z_\mathrm{EFN}}$ ($\times 10^{-6}$)				
[0.0, 0.04)	$0.26^{+0.02\oplus 0.10}_{-0.02\oplus 0.10}$ $0.28^{+0.01\oplus 0.15}_{-0.01\oplus 0.15}$	$0.34^{+0.02\oplus 0.09}_{-0.02\oplus 0.09}$ $0.38^{+0.01\oplus 0.20}_{-0.01\oplus 0.20}$	$0.30^{+0.01\oplus 0.07}_{-0.01\oplus 0.07}$ $0.36^{+0.01\oplus 0.08}_{-0.01\oplus 0.08}$	$0.18^{+0.01\oplus 0.03}_{-0.01\oplus 0.03}$ $0.25^{+0.01\oplus 0.11}_{-0.01\oplus 0.11}$	$0.01^{+0.00\oplus 0.00}_{-0.00\oplus 0.00}$ $0.02^{+0.00\oplus 0.01}_{-0.00\oplus 0.01}$
[0.04, 0.08)	$0.81^{+0.03\oplus 0.24}_{-0.03\oplus 0.24}$ $0.89^{+0.02\oplus 0.36}_{-0.02\oplus 0.36}$	$1.10^{+0.02\oplus 0.21}_{-0.02\oplus 0.21}$ $1.34^{+0.02\oplus 0.52}_{-0.02\oplus 0.52}$	$0.91^{+0.02\oplus 0.19}_{-0.02\oplus 0.19}$ $1.16^{+0.01\oplus 0.21}_{-0.01\oplus 0.21}$	$0.64^{+0.01\oplus 0.11}_{-0.01\oplus 0.11}$ $0.74^{+0.01\oplus 0.21}_{-0.01\oplus 0.21}$	$0.04^{+0.00\oplus 0.01}_{-0.00\oplus 0.01}$ $0.05^{+0.00\oplus 0.02}_{-0.00\oplus 0.02}$
[0.08, 0.1)	$1.97^{+0.03\oplus 0.45}_{-0.03\oplus 0.45}$ $2.42^{+0.03\oplus 0.69}_{-0.03\oplus 0.69}$	$3.13^{+0.06\oplus 0.43}_{-0.06\oplus 0.43}$ $3.77^{+0.04\oplus 0.86}_{-0.04\oplus 0.86}$	$2.59^{+0.03\oplus 0.44}_{-0.03\oplus 0.44}$ $3.15^{+0.03\oplus 0.43}_{-0.03\oplus 0.43}$	$1.84^{+0.03\oplus 0.25}_{-0.03\oplus 0.25}$ $2.11^{+0.02\oplus 0.36}_{-0.02\oplus 0.36}$	$0.13^{+0.00\oplus 0.02}_{-0.00\oplus 0.02}$ $0.13^{+0.00\oplus 0.02}_{-0.00\oplus 0.02}$
[0.1, 0.15)	$19.71^{+0.10\oplus 3.36}_{-0.10\oplus 3.36}$ $24.23^{+0.09\oplus 4.51}_{-0.09\oplus 4.51}$	$54.30^{+0.18\oplus 5.11}_{-0.18\oplus 5.11}$ $70.25^{+0.22\oplus 7.33}_{-0.22\oplus 7.33}$	$89.92^{+0.15\oplus 9.47}_{-0.15\oplus 9.44}$ $116.79^{+0.11\oplus 10.05}_{-0.11\oplus 10.04}$	$109.08^{+0.19\oplus 9.65}_{-0.19\oplus 9.66}$ $135.56^{+0.24\oplus 12.62}_{-0.24\oplus 12.62}$	$13.31^{+0.07\oplus 1.04}_{-0.07\oplus 1.04}$ $14.59^{+0.05\oplus 1.31}_{-0.05\oplus 1.31}$
[0.15, 0.2)	$268.14^{+0.54\oplus 35.66}_{-0.54\oplus 35.67}$ $335.76^{+0.45\oplus 46.38}_{-0.45\oplus 46.38}$	$823.27^{+0.94\oplus 54.42}_{-0.94\oplus 54.37}$ $1062.43^{+1.32\oplus 57.70}_{-1.32\oplus 57.65}$	$1439.36^{+0.66\oplus 87.45}_{-0.66\oplus 87.35}$ $1847.82^{+0.54\oplus 102.49}_{-0.54\oplus 102.45}$	$1737.54^{+1.22\oplus 92.81}_{-1.22\oplus 92.81}$ $2131.46^{+1.65\oplus 114.96}_{-1.65\oplus 114.87}$	$196.60^{+0.51\oplus 9.65}_{-0.51\oplus 9.65}$ $216.47^{+0.37\oplus 11.05}_{-0.37\oplus 11.05}$
[0.2, 0.25)	$311.85^{+0.51\oplus 38.30}_{-0.51\oplus 38.30}$ $369.24^{+0.45\oplus 46.97}_{-0.45\oplus 46.97}$	$815.93^{+0.78\oplus 44.67}_{-0.78\oplus 44.62}$ $993.35^{+0.79\oplus 47.74}_{-0.79\oplus 47.72}$	$1191.16^{+0.50\oplus 59.84}_{-0.50\oplus 59.97}$ $1442.24^{+0.54\oplus 67.56}_{-0.54\oplus 67.53}$	$1195.98^{+0.83\oplus 48.82}_{-0.83\oplus 48.82}$ $1397.17^{+1.06\oplus 60.16}_{-1.06\oplus 60.11}$	$106.75^{+0.22\oplus 4.87}_{-0.22\oplus 4.87}$ $115.33^{+0.18\oplus 5.56}_{-0.18\oplus 5.56}$
[0.25, 0.3)	$257.12^{+0.47\oplus 32.22}_{-0.47\oplus 32.22}$ $290.24^{+0.40\oplus 36.84}_{-0.40\oplus 36.84}$	$634.47^{+0.83\oplus 34.54}_{-0.83\oplus 34.52}$ $739.53^{+0.85\oplus 40.65}_{-0.85\oplus 40.63}$	$891.58^{+0.38\oplus 47.74}_{-0.38\oplus 47.78}$ $1036.60^{+0.50\oplus 48.50}_{-0.50\oplus 48.49}$	$872.21^{+0.57\oplus 35.84}_{-0.57\oplus 35.84}$ $983.42^{+0.75\oplus 46.97}_{-0.75\oplus 46.97}$	$74.93^{+0.14\oplus 3.70}_{-0.14\oplus 3.70}$ $79.27^{+0.18\oplus 4.49}_{-0.18\oplus 4.49}$
[0.3, 0.35)	$224.98^{+0.34\oplus 28.97}_{-0.34\oplus 28.97}$ $243.69^{+0.28\oplus 31.14}_{-0.28\oplus 31.14}$	$545.89^{+0.60\oplus 30.38}_{-0.60\oplus 30.38}$ $610.15^{+0.71\oplus 37.06}_{-0.71\oplus 37.02}$	$758.44^{+0.29\oplus 41.96}_{-0.29\oplus 41.96}$ $848.72^{+0.36\oplus 39.32}_{-0.36\oplus 39.32}$	$735.51^{+0.64\oplus 34.26}_{-0.64\oplus 34.26}$ $799.90^{+0.63\oplus 46.58}_{-0.63\oplus 46.59}$	$61.49^{+0.16\oplus 3.19}_{-0.16\oplus 3.19}$ $63.62^{+0.12\oplus 4.29}_{-0.12\oplus 4.29}$
[0.35, 0.4)	$213.18^{+0.37\oplus 27.62}_{-0.37\oplus 27.62}$ $223.30^{+0.27\oplus 28.66}_{-0.27\oplus 28.66}$	$510.11^{+0.71\oplus 27.25}_{-0.71\oplus 27.25}$ $549.08^{+0.48\oplus 33.10}_{-0.48\oplus 33.07}$	$705.37^{+0.37\oplus 38.49}_{-0.37\oplus 38.49}$ $758.66^{+0.27\oplus 33.95}_{-0.27\oplus 33.95}$	$677.50^{+0.46\oplus 33.62}_{-0.46\oplus 33.62}$ $711.11^{+0.53\oplus 45.63}_{-0.53\oplus 45.63}$	$55.32^{+0.19\oplus 2.95}_{-0.19\oplus 2.95}$ $55.73^{+0.13\oplus 4.88}_{-0.13\oplus 4.88}$
[0.4, 0.5)	$218.58^{+0.25\oplus 28.11}_{-0.25\oplus 28.11}$ $218.27^{+0.21\oplus 28.25}_{-0.21\oplus 28.25}$	$517.39^{+0.66\oplus 25.31}_{-0.66\oplus 25.31}$ $527.16^{+0.47\oplus 29.76}_{-0.47\oplus 29.76}$	$706.70^{+0.30\oplus 37.85}_{-0.30\oplus 37.85}$ $715.74^{+0.28\oplus 31.52}_{-0.28\oplus 31.52}$	$663.47^{+0.48\oplus 34.23}_{-0.48\oplus 34.23}$ $659.62^{+0.54\oplus 42.95}_{-0.54\oplus 42.95}$	$51.96^{+0.12\oplus 2.93}_{-0.12\oplus 2.93}$ $49.99^{+0.13\oplus 5.27}_{-0.13\oplus 5.27}$
[0.5, 0.6)	$263.26^{+0.30\oplus 33.61}_{-0.30\oplus 33.61}$ $249.39^{+0.29\oplus 32.92}_{-0.29\oplus 32.92}$	$614.28^{+0.57\oplus 28.70}_{-0.57\oplus 28.70}$ $583.47^{+0.53\oplus 32.28}_{-0.53\oplus 32.30}$	$819.41^{+0.33\oplus 43.88}_{-0.33\oplus 43.88}$ $762.51^{+0.26\oplus 34.62}_{-0.26\oplus 34.62}$	$732.72^{+0.52\oplus 39.77}_{-0.52\oplus 39.77}$ $668.92^{+0.53\oplus 44.04}_{-0.53\oplus 44.04}$	$51.85^{+0.14\oplus 3.20}_{-0.14\oplus 3.20}$ $46.86^{+0.14\oplus 5.29}_{-0.14\oplus 5.29}$
[0.6, 0.65)	$322.44^{+0.50\oplus 41.07}_{-0.50\oplus 41.07}$ $293.90^{+0.38\oplus 39.74}_{-0.38\oplus 39.75}$	$739.07^{+0.78\oplus 36.19}_{-0.78\oplus 36.20}$ $667.73^{+0.60\oplus 37.91}_{-0.60\oplus 37.92}$	$957.64^{+0.39\oplus 52.06}_{-0.39\oplus 52.06}$ $838.33^{+0.39\oplus 41.44}_{-0.39\oplus 41.44}$	$815.55^{+0.70\oplus 45.74}_{-0.70\oplus 45.74}$ $695.75^{+0.49\oplus 47.21}_{-0.49\oplus 47.21}$	$51.92^{+0.11\oplus 3.50}_{-0.11\oplus 3.50}$ $44.35^{+0.14\oplus 7.49}_{-0.14\oplus 7.49}$
[0.65, 0.7)	$378.22^{+0.51\oplus 48.26}_{-0.51\oplus 48.26}$ $337.54^{+0.43\oplus 46.05}_{-0.43\oplus 46.05}$	$862.07^{+0.97\oplus 46.19}_{-0.97\oplus 46.20}$ $752.86^{+0.63\oplus 51.22}_{-0.63\oplus 50.33}$	$1091.86^{+0.55\oplus 60.40}_{-0.55\oplus 60.40}$ $916.28^{+0.35\oplus 50.93}_{-0.35\oplus 50.93}$	$890.69^{+0.67\oplus 50.94}_{-0.67\oplus 50.94}$ $724.61^{+0.57\oplus 55.47}_{-0.57\oplus 55.47}$	$52.06^{+0.18\oplus 3.73}_{-0.18\oplus 3.73}$ $42.59^{+0.16\oplus 17.57}_{-0.16\oplus 17.57}$
[0.7, 0.75)	$435.35^{+0.54\oplus 56.09}_{-0.54\oplus 56.10}$ $381.31^{+0.46\oplus 51.47}_{-0.46\oplus 51.47}$	$992.17^{+1.24\oplus 57.48}_{-1.24\oplus 57.49}$ $838.52^{+0.97\oplus 57.66}_{-0.97\oplus 54.54}$	$1220.57^{+0.53\oplus 61.27}_{-0.53\oplus 61.27}$ $982.21^{+0.42\oplus 61.47}_{-0.42\oplus 61.47}$	$941.09^{+0.94\oplus 53.60}_{-0.94\oplus 53.59}$ $732.18^{+0.50\oplus 85.78}_{-0.50\oplus 85.78}$	$49.30^{+0.16\oplus 3.76}_{-0.16\oplus 3.76}$ $38.55^{+0.16\oplus 33.72}_{-0.16\oplus 33.72}$
[0.75, 0.8)	$386.83^{+0.71\oplus 50.26}_{-0.71\oplus 50.26}$ $335.71^{+0.62\oplus 45.35}_{-0.62\oplus 45.35}$	$890.29^{+1.42\oplus 55.77}_{-1.42\oplus 55.78}$ $734.01^{+1.11\oplus 37.12}_{-1.11\oplus 36.47}$	$1064.57^{+0.43\oplus 47.00}_{-0.43\oplus 46.99}$ $826.78^{+0.31\oplus 60.12}_{-0.31\oplus 60.12}$	$768.91^{+0.76\oplus 42.22}_{-0.76\oplus 42.22}$ $571.59^{+0.70\oplus 130.99}_{-0.70\oplus 130.99}$	$35.04^{+0.13\oplus 2.90}_{-0.13\oplus 2.90}$ $26.18^{+0.11\oplus 34.81}_{-0.11\oplus 34.81}$
[0.8, 0.85)	$166.65^{+0.34\oplus 22.79}_{-0.34\oplus 22.79}$ $143.13^{+0.39\oplus 19.35}_{-0.39\oplus 19.35}$	$375.18^{+0.88\oplus 44.34}_{-0.88\oplus 44.34}$ $305.28^{+0.71\oplus 20.07}_{-0.71\oplus 21.49}$	$424.24^{+0.32\oplus 21.78}_{-0.32\oplus 21.82}$ $321.29^{+0.25\oplus 45.17}_{-0.25\oplus 45.17}$	$277.83^{+0.49\oplus 19.48}_{-0.49\oplus 19.47}$ $199.03^{+0.47\oplus 78.79}_{-0.47\oplus 78.79}$	$10.21^{+0.09\oplus 0.95}_{-0.09\oplus 0.95}$ $7.53^{+0.06\oplus 11.74}_{-0.06\oplus 11.74}$
[0.85, 0.9)	$5.57^{+0.05\oplus 1.62}_{-0.05\oplus 1.62}$ $4.73^{+0.03\oplus 0.73}_{-0.03\oplus 0.73}$	$10.66^{+0.11\oplus 3.94}_{-0.11\oplus 3.94}$ $8.55^{+0.07\oplus 0.87}_{-0.07\oplus 0.86}$	$10.18^{+0.04\oplus 1.27}_{-0.04\oplus 1.27}$ $7.45^{+0.03\oplus 3.15}_{-0.03\oplus 3.15}$	$5.57^{+0.03\oplus 0.84}_{-0.03\oplus 0.84}$ $3.91^{+0.03\oplus 2.17}_{-0.03\oplus 2.17}$	$0.15^{+0.00\oplus 0.03}_{-0.00\oplus 0.03}$ $0.13^{+0.00\oplus 0.21}_{-0.00\oplus 0.21}$
[0.9, 0.93)	$0.19^{+0.01\oplus 0.17}_{-0.01\oplus 0.17}$ $0.17^{+0.01\oplus 0.05}_{-0.01\oplus 0.05}$	$0.35^{+0.01\oplus 0.27}_{-0.01\oplus 0.27}$ $0.30^{+0.01\oplus 0.08}_{-0.01\oplus 0.08}$	$0.38^{+0.01\oplus 0.16}_{-0.01\oplus 0.16}$ $0.29^{+0.01\oplus 0.25}_{-0.01\oplus 0.25}$	$0.24^{+0.01\oplus 0.07}_{-0.01\oplus 0.07}$ $0.19^{+0.01\oplus 0.14}_{-0.01\oplus 0.14}$	$0.01^{+0.00\oplus 0.01}_{-0.00\oplus 0.01}$ $0.01^{+0.00\oplus 0.01}_{-0.00\oplus 0.01}$
[0.93, 0.96)	$0.04^{+0.00\oplus 0.08}_{-0.00\oplus 0.08}$ $0.03^{+0.00\oplus 0.02}_{-0.00\oplus 0.02}$	$0.08^{+0.00\oplus 0.09}_{-0.00\oplus 0.09}$ $0.07^{+0.00\oplus 0.04}_{-0.00\oplus 0.04}$	$0.08^{+0.00\oplus 0.06}_{-0.00\oplus 0.06}$ $0.06^{+0.00\oplus 0.07}_{-0.00\oplus 0.07}$	$0.05^{+0.00\oplus 0.03}_{-0.00\oplus 0.03}$ $0.04^{+0.00\oplus 0.04}_{-0.00\oplus 0.04}$	$0.00^{+0.00\oplus 0.00}_{-0.00\oplus 0.00}$ $0.00^{+0.00\oplus 0.00}_{-0.00\oplus 0.00}$
[0.96, 1.0)	$0.00^{+0.00\oplus 0.01}_{-0.00\oplus 0.01}$ $0.00^{+0.00\oplus 0.00}_{-0.00\oplus 0.00}$	$0.01^{+0.00\oplus 0.01}_{-0.00\oplus 0.01}$ $0.00^{+0.00\oplus 0.00}_{-0.00\oplus 0.00}$	$0.01^{+0.00\oplus 0.01}_{-0.00\oplus 0.01}$ $0.00^{+0.00\oplus 0.01}_{-0.00\oplus 0.01}$	$0.00^{+0.00\oplus 0.00}_{-0.00\oplus 0.00}$ $0.00^{+0.00\oplus 0.00}_{-0.00\oplus 0.00}$	$0.00^{+0.00\oplus 0.00}_{-0.00\oplus 0.00}$ $0.00^{+0.00\oplus 0.00}_{-0.00\oplus 0.00}$

Table B.6: Numerical values of the unfolded inclusive relative tripe-differential dijet cross section as a function of the NN-score as predicted by the EFN including upper and lower uncertainties. The two numbers that are provided for each score interval correspond to the relative event rate in central (top) and forward (bottom) rapidity category. Note the different scale factors that have been applied for the respective bins in transverse momentum.

Bibliography

[1] C. Crego. [Online]. Available: http://picture-poems.com/week5/fallstorm.html

[2] M.-J. Nederhof, "Dispute of a man with his ba," *Online*, nov 2006, transliteration and translation for Dialogue of a Man and his Ba, Papyrus Berlin 3024, following the transcription and facsimile in Goedicke (1970). An older transcription can be found in Faulkner (1956) and Barta (1969).

[3] M. Zaheer, S. Kottur, S. Ravanbakhsh, B. Poczos, R. Salakhutdinov, and A. Smola, "Deep sets," 2018.

[4] "Multijet simulation for 13 TeV ATLAS Analyses," CERN, Geneva, Tech. Rep., Apr 2019, all figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2019-017. [Online]. Available: https://cds.cern.ch/record/2672252

[5] S. Höche, "Introduction to parton-shower event generators," 2014.

[6] R. K. Ellis, W. J. Stirling, and B. R. Webber, *QCD and Collider Physics*, ser. Cambridge Monographs on Particle Physics, Nuclear Physics and Cosmology. Cambridge University Press, 1996.

[7] H. Kvedalen, "Study of radiation effects in cots semiconductors for use in high energy physics experiments," Master's thesis, Department of Informatics, University of Oslo, 2002.

[8] E. Ward, "LHC and HL-LHC timeline for ATLAS website," Dec 2018, general Photo. [Online]. Available: http://cds.cern.ch/record/2652466

[9] J. Pequenao, "Computer generated image of the whole ATLAS detector," Mar 2008. [Online]. Available: https://cds.cern.ch/record/1095924

[10] I. Dawson and V. Hedberg, "Radiation in the USA15 cavern in ATLAS," CERN, Geneva, Tech. Rep., 2004. [Online]. Available: http://cds.cern.ch/record/734133

[11] O. S. Brüning, P. Collier, P. Lebrun, S. Myers, R. Ostojic, J. Poole, and P. Proudlock, *LHC Design Report*, ser. CERN Yellow Reports: Monographs. Geneva: CERN, 2004. [Online]. Available: https://cds.cern.ch/record/815187

[12] *ATLAS inner detector: Technical Design Report, 1*, ser. Technical design report. ATLAS. Geneva: CERN, 1997. [Online]. Available: https://cds.cern.ch/record/331063

[13] A. Paoloni and R. Santonico, "Resistive plate chambers for the atlas level-1 muon trigger," 01 2022.

[14] G. e. a. Aad, "The ATLAS Experiment at the CERN Large Hadron Collider," *JINST*, vol. 3, p. S08003. 437 p, 2008, also published by CERN Geneva in 2010. [Online]. Available: https://cds.cern.ch/record/1129811

[15] A. Airapetian *et al.*, "ATLAS calorimeter performance Technical Design Report," 12 1996.

[16] "ATLAS liquid argon calorimeter: Technical design report," 12 1996.

[17] "Technical Design Report for the Phase-II Upgrade of the ATLAS Muon Spectrometer," CERN, Geneva, Tech. Rep., Sep 2017. [Online]. Available: https://cds.cern.ch/record/2285580

[18] A. Colaboration, *ATLAS detector and physics performance: Technical Design Report, 1*, ser. Technical design report. ATLAS. Geneva: CERN, 1999. [Online]. Available: https://cds.cern.ch/record/391176

[19] *ATLAS magnet system: Technical Design Report, 1*, ser. Technical design report. ATLAS. Geneva: CERN, 1997. [Online]. Available: https://cds.cern.ch/record/338080

[20] G. Aad, B. Abbott, J. Abdallah, O. Abdinov, R. Aben, M. Abolins, O. S. AbouZeid, H. Abramowicz, H. Abreu, and et al., "Topological cell clustering in the atlas calorimeters and its performance in lhc run 1," *The European Physical Journal C*, vol. 77, no. 7, Jul 2017. [Online]. Available: http://dx.doi.org/10.1140/epjc/s10052-017-5004-5

[21] M. Aaboud, G. Aad, B. Abbott, J. Abdallah, O. Abdinov, B. Abeloos, S. Abidi, O. AbouZeid, N. Abraham, H. Abramowicz, and et al., "Jet energy scale measurements and their systematic uncertainties in proton-proton collisions at s=13 tev with the atlas detector," *Physical Review D*, vol. 96, no. 7, Oct 2017. [Online]. Available: http://dx.doi.org/10.1103/PhysRevD.96.072002

[22] B. Povh, K. Rith, C. Scholz, and F. Zetsche, *Teilchen und Kerne. Eine Einfuehrung in die physikalischen Konzepte*, 6th ed. Springer-Verlag GmbH, 2004.

[23] J. R. Andersen *et al.*, "Les Houches 2015: Physics at TeV Colliders Standard Model Working Group Report," in *9th Les Houches Workshop on Physics at TeV Colliders*, 5 2016.

[24] A. J. Larkoski, J. Thaler, and W. J. Waalewijn, "Gaining (Mutual) Information about Quark/Gluon Discrimination," *JHEP*, vol. 11, p. 129, 2014.

[25] The CMS Collaboration, "The CMS experiment at the CERN LHC," *JINST*, vol. 3, p. S08004, 2008. [Online]. Available: http://dx.doi.org/10.1088/1748-0221/3/08/S08004

[26] C. Frye, A. J. Larkoski, J. Thaler, and K. Zhou, "Casimir Meets Poisson: Improved Quark/Gluon Discrimination with Counting Observables," *JHEP*, vol. 09, p. 083, 2017.

[27] P. T. Komiske, E. M. Metodiev, and J. Thaler, "Energy flow networks: deep sets for particle jets," *Journal of High Energy Physics*, vol. 2019, no. 1, jan 2019. [Online]. Available: https://doi.org/10.1007%2Fjhep01%282019%29121

[28] D. M. Blei, "Probabilistic topic models," *Commun. ACM*, vol. 55, no. 4, p. 77–84, apr 2012. [Online]. Available: https://doi.org/10.1145/2133806.2133826

[29] E. M. Metodiev and J. Thaler, "Jet Topics: Disentangling Quarks and Gluons at Colliders," *Phys. Rev. Lett.*, vol. 120, no. 24, p. 241602, 2018.

[30] V. Khachatryan *et al.*, "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," *Eur. Phys. J. C*, vol. 75, no. 6, p. 288, 2015.

[31] ——, "Precise determination of the mass of the Higgs boson and tests of compatibility of its couplings with the standard model predictions using proton collisions at 7 and 8 TeV," *Eur. Phys. J. C*, vol. 75, no. 5, p. 212, 2015.

[32] S. Chatrchyan *et al.*, "Measurement of the Ratio of the Inclusive 3-Jet Cross Section to the Inclusive 2-Jet Cross Section in pp Collisions at $\sqrt{s} = 7$ TeV and First Determination of the Strong Coupling Constant in the TeV Range," *Eur. Phys. J. C*, vol. 73, no. 10, p. 2604, 2013.

[33] ——, "Determination of the Top-Quark Pole Mass and Strong Coupling Constant from the $t\bar{t}$ Production Cross Section in pp Collisions at $\sqrt{s} = 7$ TeV," *Phys. Lett. B*, vol. 728, pp. 496–517, 2014, [Erratum: Phys.Lett.B 738, 526–528 (2014)].

[34] V. Andreev *et al.*, "Measurement of Jet Production Cross Sections in Deep-inelastic *ep* Scattering at HERA," *Eur. Phys. J. C*, vol. 77, no. 4, p. 215, 2017, [Erratum: Eur.Phys.J.C 81, 739 (2021)].

[35] V. E. Aushev, I. M. Kadenko, D. I. Lontkovskyi, and I. V. Makarenko, "Photoproduction of jets and Determination of α_s at a lepton-Proton collider HERA," in *4th International Conference Current Problems in Nuclear Physics and Atomic Energy*, 2012, pp. 601–607.

[36] V. M. Abazov *et al.*, "Determination of the strong coupling constant from the inclusive jet cross section in $p\bar{p}$ collisions at sqrt(s)=1.96 TeV," *Phys. Rev. D*, vol. 80, p. 111107, 2009.

[37] G. Aad *et al.*, "Measurement of the production cross section for a Higgs boson in association with a vector boson in the $H \to WW^* \to \ell\nu\ell\nu$ channel in pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector," *Phys. Lett. B*, vol. 798, p. 134949, 2019.

[38] D. d'Enterria and P. Z. Skands, Eds., *Proceedings, High-Precision α_s Measurements from LHC to FCC-ee: Geneva, Switzerland, October 2-13, 2015.* Geneva: CERN, 12 2015.

[39] J. Pumplin, D. R. Stump, J. Huston, H. L. Lai, P. M. Nadolsky, and W. K. Tung, "New generation of parton distributions with uncertainties from global QCD analysis," *JHEP*, vol. 07, p. 012, 2002.

[40] M. Aaboud *et al.*, "Jet reconstruction and performance using particle flow with the ATLAS Detector," *Eur. Phys. J. C*, vol. 77, no. 7, p. 466, 2017.

[41] F. Spano, "Unfolding in particle physics: a window on solving inverse problems," *EPJ Web Conf.*, vol. 55, p. 03002, 2013.

[42] P. T. Komiske, E. M. Metodiev, and J. Thaler, "An operational definition of quark and gluon jets," *JHEP*, vol. 11, p. 059, 2018.

[43] M. Aaboud *et al.*, "Study of the material of the ATLAS inner detector for Run 2 of the LHC," *JINST*, vol. 12, no. 12, p. P12009, 2017.

[44] ——, "Performance of the ATLAS Track Reconstruction Algorithms in Dense Environments in LHC Run 2," *Eur. Phys. J. C*, vol. 77, no. 10, p. 673, 2017.

[45] G. Aad *et al.*, "Jet energy scale and resolution measured in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector," *Eur. Phys. J. C*, vol. 81, no. 8, p. 689, 2021.

[46] M. LeBlanc, B. Nachman, and C. Sauer, "Going off topics to demix quark and gluon jets in α_S extractions," 6 2022.

[47] K. a. K. a. M. Eidelman, S. andHayes *et al.*, "Review of Particle Physics," *Physics Letters B*, vol. 592, pp. 1+, 2004. [Online]. Available: http://pdg.lbl.gov

[48] D. M. Blei, "Introduction to probabilistic topic models," in *In Communications of the ACM*, 2011.

[49] C. Sauer, "Towards a data-driven simulation of qcd radiation with generative models utilizing machine learning methods," Ph.D. dissertation, Universität Heidelberg, 2010.

[50] M. W. Grunewald, "Precision tests of the standard model," *PoS*, vol. HEP2005, p. 306, 2006.

[51] S. L. Glashow, "Partial Symmetries of Weak Interactions," *Nucl. Phys.*, vol. 22, pp. 579–588, 1961.

[52] A. Salam, "Weak and Electromagnetic Interactions," *Conf. Proc.*, vol. C680519, pp. 367–377, 1968.

[53] S. Weinberg, "A model of leptons," *Phys. Rev. Lett.*, vol. 19, pp. 1264–1266, Nov 1967. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.19.1264

[54] G. Aad *et al.*, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC," *Phys.Lett.*, vol. B716, pp. 1–29, 2012.

[55] J. R. Ellis, "Limits of the standard model," in *PSI Zuoz Summer School on Exploring the Limits of the Standard Model Zuoz, Engadin, Switzerland, August 18-24, 2002*, 2002. [Online]. Available: http://weblib.cern.ch/abstract?CERN-TH-2002-320

[56] J. C. Maxwell, "A dynamical theory of the electromagnetic field," *Philosophical Transactions of the Royal Society of London*, vol. 155, pp. 459–513, 1865.

[57] A. Einstein, "Die Grundlage der allgemeinen Relativitätstheorie," *Annalen der Physik*, vol. 354, pp. 769–822, 1916.

[58] I. Newton, A. Motte, and J. Machin, *The Mathematical Principles of Natural Philosophy*, ser. The Mathematical Principles of Natural Philosophy. B. Motte, 1729, no. Bd. 1. [Online]. Available: https://books.google.de/books?id=Tm0FAAAAQAAJ

[59] B. Bothe, "Künstliche erregung von kern-γ-strahlen," *Z. Phys.*, vol. 66, p. 289, 1930.

[60] E. Fermi, "Tentativo di una teoria dell'emissione dei raggi beta," *Ric. Sci.*, vol. 4, pp. 491–495, 1933.

[61] H. Kramers, *Die grundlagen der quantentheorie: Quantentheorie des elektrons und der strahlung*, ser. Die grundlagen der quantentheorie: Quantentheorie des elektrons und der strahlung. Akademische verlagsgesellschaft m.b.h., 1938, no. Bd. 2. [Online]. Available: https://books.google.de/books?id=iA_QAAAAMAAJ

[62] M. Gell-Mann, "The eightfold way: A theory of strong interaction symmetry," 3 1961.

[63] ——, "A Schematic Model of Baryons and Mesons," *Phys. Lett.*, vol. 8, pp. 214–215, 1964.

[64] G. Zweig, "An SU(3) model for strong interaction symmetry and its breaking. Version 2," in *DEVELOPMENTS IN THE QUARK THEORY OF HADRONS. VOL. 1. 1964 - 1978*, D. Lichtenberg and S. P. Rosen, Eds., 1964, pp. 22–101.

[65] C.-N. Yang and R. L. Mills, "Conservation of Isotopic Spin and Isotopic Gauge Invariance," *Phys. Rev.*, vol. 96, pp. 191–195, 1954, [,150(1954)].

[66] J. S. Schwinger, "A Theory of the Fundamental Interactions," *Annals Phys.*, vol. 2, pp. 407–434, 1957.

[67] A. Salam and J. C. Ward, "Electromagnetic and weak interactions," *Phys. Lett.*, vol. 13, pp. 168–171, 1964.

[68] D. Hanneke, S. Fogwell, and G. Gabrielse, "New measurement of the electron magnetic moment and the fine structure constant," *Physical Review Letters*, vol. 100, no. 12, 3 2008.

[69] P. W. Higgs, "Broken Symmetries and the Masses of Gauge Bosons," *Phys. Rev. Lett.*, vol. 13, pp. 508–509, 1964, [,160(1964)].

[70] F. Englert and R. Brout, "Broken Symmetry and the Mass of Gauge Vector Mesons," *Phys. Rev. Lett.*, vol. 13, pp. 321–323, 1964, [,157(1964)].

[71] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, "Global Conservation Laws and Massless Particles," *Phys. Rev. Lett.*, vol. 13, pp. 585–587, 1964, [,162(1964)].

[72] F. J. Hasert *et al.*, "Observation of Neutrino Like Interactions Without Muon Or Electron in the Gargamelle Neutrino Experiment," *Phys. Lett.*, vol. B46, pp. 138–140, 1973, [,5.15(1973)].

[73] F. J. Hasert, H. Faissner, W. Krenz, J. Von Krogh, and D. Lanske, "Search for elastic muon neutrino electron scattering," *Phys. Lett. B*, vol. 46, pp. 121–124, 1973. [Online]. Available: http://cds.cern.ch/record/243640

[74] P. M. Watkins, "DISCOVERY OF THE W AND Z BOSONS," *Contemp. Phys.*, vol. 27, pp. 291–324, 1986.

[75] C. Berger *et al.*, "Jet Analysis of the Υ (9.46) Decay Into Charged Hadrons," *Phys. Lett.*, vol. 82B, pp. 449–455, 1979.

[76] D. P. Barber *et al.*, "Discovery of Three Jet Events and a Test of Quantum Chromodynamics at PETRA Energies," *Phys. Rev. Lett.*, vol. 43, p. 830, 1979.

[77] R. Brandelik *et al.*, "Evidence for Planar Events in e+ e- Annihilation at High-Energies," *Phys. Lett.*, vol. 86B, pp. 243–249, 1979.

[78] G. Goldhaber, "Discovery of charm," vol. 300, 11 1984.

[79] C. Campagnari and M. Franklin, "The Discovery of the top quark," *Rev. Mod. Phys.*, vol. 69, pp. 137–212, 1997.

[80] S. Weinberg, "The Making of the standard model," *Eur. Phys. J.*, vol. C34, pp. 5–13, 2004, [,99(2005)].

[81] B. L. G. Bakker, A. I. Veselov, and M. A. Zubkov, "Internal structure of discretized Weinberg-Salam model," *Phys. Lett.*, vol. B583, pp. 379–382, 2004.

[82] D. J. Gross and F. Wilczek, "Ultraviolet behavior of non-abelian gauge theories," *Phys. Rev. Lett.*, vol. 30, pp. 1343–1346, Jun 1973. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.30.1343

[83] N. Foundation, "The nobel prize in physics," 2004. [Online]. Available: https://www.nobelprize.org/prizes/physics/2004/summary/

[84] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjostrand, "Parton Fragmentation and String Dynamics," *Phys. Rept.*, vol. 97, pp. 31–145, 1983.

[85] N. Foundation, "The nobel prize in physics," 1965. [Online]. Available: https://www.nobelprize.org/prizes/physics/1965/summary/

[86] A. M. Jaffe and E. Witten, "Quantum Yang-Mills theory," 2000.

[87] V. Bornyakov *et al.*, "Heavy quark potential in lattice QCD at finite temperature," in *Quark confinement and the hadron spectrum. Proceedings, 5th International Conference, Gargnano, Italy, September 10-14, 2002*, 2003, pp. 294–296.

[88] B. Andersson, "THE LUND STRING MODEL," in *7th European Symposium on Antiproton Interactions: From LEAR to the Collider and Beyond Durham, England, July 9-13, 1984*, 1986, pp. 447–462.

[89] Y. L. Dokshitzer, "Calculation of the Structure Functions for Deep Inelastic Scattering and e+ e- Annihilation by Perturbation Theory in Quantum Chromodynamics." *Sov. Phys. JETP*, vol. 46, pp. 641–653, 1977, [Zh. Eksp. Teor. Fiz.73,1216(1977)].

[90] V. N. Gribov and L. N. Lipatov, "Deep inelastic e p scattering in perturbation theory," *Sov. J. Nucl. Phys.*, vol. 15, pp. 438–450, 1972, [Yad. Fiz.15,781(1972)].

[91] G. Altarelli and G. Parisi, "Asymptotic Freedom in Parton Language," *Nucl. Phys.*, vol. B126, pp. 298–318, 1977.

[92] E. W. N. Glover and D. A. Kosower, "Recombination methods for jets in p anti-p collisions," *Phys. Lett.*, vol. B367, pp. 369–376, 1996.

[93] J. E. Huth *et al.*, "Toward a standardization of jet definitions," in *1990 DPF Summer Study on High-energy Physics: Research Directions for the Decade (Snowmass 90) Snowmass, Colorado, June 25-July 13, 1990*, 1990, pp. 0134–136. [Online]. Available: http://lss.fnal.gov/cgi-bin/find_paper.pl?conf-90-249

[94] F. t. Abe, "Topology of three-jet events in $\overline{p}p$ collisions at $\sqrt{s} = 1.8$ tev," *Phys. Rev. D*, vol. 45, pp. 1448–1458, Mar 1992. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevD.45.1448

[95] M. H. Seymour, "Jet shapes in hadron collisions: Higher orders, resummation and hadronization," *Nucl. Phys.*, vol. B513, pp. 269–300, 1998.

[96] N. Metropolis and S. Ulam, "The Monte Carlo method," vol. 44, no. 247, pp. 335–341, Sep. 1949.

[97] R. D. Richtmyer, S. Ulam, and J. von Neumann, "Statistical methods in neutron diffusion," Technical Report LAMS-551, Apr. 1947, republished in typeset form in [((?))].

[98] W. Feller, "The fundamental limit theorems in probability," *Bull. Amer. Math. Soc.*, vol. 51, no. 11, pp. 800–832, 11 1945.

[99] N. Metropolis and W. Aspray, "Oral history interview with Nicholas Metropolis," Audio recording, 1987, the Charles Babbage Institute.

[100] S. Hartmann, "James t. cushing, philosophical concepts in physics. the historical relation between philosophy and scientific theories," *Erkenntnis*, vol. 52, pp. 133–137, 01 2000.

[101] S. Agostinelli *et al.*, "GEANT4: A Simulation toolkit," *Nucl. Instrum. Meth.*, vol. A506, pp. 250–303, 2003.

[102] B. Webber, "Fragmentation and Hadronization," *Int. J. Mod. Phys. AS*, vol. 15S1, pp. 577–606. 31 p, Jan 2000. [Online]. Available: https://cds.cern.ch/record/419784

[103] K. A. Johnson, "The bag model of quark confinement," *Scientific American*, vol. 241, no. 1, pp. 112–121, 1979. [Online]. Available: http://www.jstor.org/stable/24965246

[104] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjöstrand, "Parton fragmentation and string dynamics," *Physics Reports*, vol. 97, no. 2, pp. 31–145, 1983. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0370157383900807

[105] A. Buckley, J. Butterworth, S. Gieseke, D. Grellscheid, S. Hoeche, H. Hoeth, F. Krauss, L. Lonnblad, E. Nurse, P. Richardson, S. Schumann, M. Seymour, T. Sjostrand, P. Skands, and B. Webber, "Mcnet-11-01 general-purpose event generators for lhc physics," *Physics Reports*, vol. 504, 01 2011.

[106] D. Amati and G. Veneziano, "Preconfinement as a Property of Perturbative QCD," *Phys. Lett. B*, vol. 83, pp. 87–92, 1979.

[107] M. H. Seymour and M. Marx, "Monte Carlo Event Generators," in *69th Scottish Universities Summer School in Physics: LHC Physics*, 4 2013, pp. 287–319.

[108] G. Arnison *et al.*, "Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at $\sqrt{s} = 540$ GeV," *Phys. Lett. B*, vol. 122, pp. 103–116, 1983.

[109] C. H. Albright, "y and v distributions for neutral current reactions of the weinberg-type," *Nuclear Physics B*, vol. 70, no. 3, pp. 486–500, 1974. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0550321374904441

[110] *Proceedings of the ECFA-CERN Workshop : Large Hadron Collider in the LEP tunnel: Lausanne, Switzerland 21 - 27 Mar 1984. ECFA-CERN Workshop on Large Hadron Collider in the LEP tunnel*, CERN. Geneva: CERN, 1984. [Online]. Available: https://cds.cern.ch/record/154938

[111] "ATLAS: Letter of intent for a general purpose p p experiment at the large hadron collider at CERN," 10 1992.

[112] *ATLAS: technical proposal for a general-purpose pp experiment at the Large Hadron Collider at CERN*, ser. LHC technical proposal. Geneva: CERN, 1994. [Online]. Available: https://cds.cern.ch/record/290968

[113] S. W. Hawking, "Particle Creation by Black Holes," *Commun. Math. Phys.*, vol. 43, pp. 199–220, 1975, [Erratum: Commun.Math.Phys. 46, 206 (1976)].

[114] *ATLAS liquid-argon calorimeter: Technical Design Report*, ser. Technical design report. ATLAS. Geneva: CERN, 1996. [Online]. Available: https://cds.cern.ch/record/331061

[115] F. Cavallari, "Performance of calorimeters at the LHC," *Journal of Physics: Conference Series*, vol. 293, p. 012001, apr 2011. [Online]. Available: https://doi.org/10.1088/1742-6596/293/1/012001

[116] A. Outreach, "ATLAS Fact Sheet : To raise awareness of the ATLAS detector and collaboration on the LHC," 2010. [Online]. Available: https://cds.cern.ch/record/1457044

[117] T. A. Collaboration, "The ATLAS experiment at the CERN large hadron collider," *Journal of Instrumentation*, vol. 3, no. 08, pp. S08 003–S08 003, aug 2008. [Online]. Available: https://doi.org/10.1088/1748-0221/3/08/s08003

[118] O. Fedin, "Reconstruction and identification of photons and electrons with the ATLAS detector," *Nucl. Instrum. Methods Phys. Res., A*, vol. 623, pp. 306–308. 3 p, Apr 2009, tIPP09 conference proceedings. [Online]. Available: https://cds.cern.ch/record/1171936

[119] L. Sonnenschein, "Analytical solution of ttbar dilepton equations," 2006.

[120] S. Rettie, "Muon identification and performance in the ATLAS experiment," CERN, Geneva, Tech. Rep., Jun 2018. [Online]. Available: https://cds.cern.ch/record/2626330

[121] C. Patrignani *et al.*, "Review of Particle Physics," *Chin. Phys. C*, vol. 40, no. 10, p. 100001, 2016.

[122] R. Iguchi, "The Phase-I Trigger Readout Electronics Upgrade of the ATLAS Liquid Argon Calorimeters," *PoS*, vol. EPS-HEP2019, p. 202. 5 p, 2020. [Online]. Available: https://cds.cern.ch/record/2760981

[123] P. L. and, "Jet measurements in ATLAS," *Journal of Physics: Conference Series*, vol. 323, p. 012002, nov 2011. [Online]. Available: https://doi.org/10.1088/1742-6596/323/1/012002

[124] G. Aad, B. Abbott, J. Abdallah, O. Abdinov, R. Aben, M. Abolins, O. S. AbouZeid, H. Abramowicz, H. Abreu, and et al., "Topological cell clustering in the atlas calorimeters and its performance in lhc run 1," *The European Physical Journal C*, vol. 77, no. 7, Jul 2017. [Online]. Available: http://dx.doi.org/10.1140/epjc/s10052-017-5004-5

[125] G. F. Sterman and S. Weinberg, "Jets from Quantum Chromodynamics," *Phys. Rev. Lett.*, vol. 39, p. 1436, 1977.

[126] M. H. Seymour, "Jets in hadron collisions," in *Deep inelastic scattering. Proceedings, 8th International Workshop, DIS 2000, Liverpool, UK, April 25-30, 2000*, 2000, pp. 27–41, [,21(2000)].

[127] R. Atkin, "Review of jet reconstruction algorithms," *J. Phys. Conf. Ser.*, vol. 645, no. 1, p. 012008, 2015.

[128] G. P. Salam and G. Soyez, "A Practical Seedless Infrared-Safe Cone jet algorithm," *JHEP*, vol. 05, p. 086, 2007.

[129] W. Bartel, L. Becker, R. Felst, D. Haidt, G. Knies, H. Krehbiel, P. Laurikainen, N. Magnussen, R. Meinke, B. Naroska, J. Olsson, D. Schmidt, G. Dietrich, T. Greenshaw, J. Hagemann, G. Heinzelmann, H. Kado, C. Kleinwort, M. Kuhlen, and S. Yamada, "Experimental studies on multijet production ine+e− annihilation at petra energies," *Zeitschrift für Physik C*, vol. 33, pp. 23–31, 03 1986.

[130] S. Catani, Y. L. Dokshitzer, M. H. Seymour, and B. R. Webber, "Longitudinally invariant K_t clustering algorithms for hadron hadron collisions," *Nucl. Phys.*, vol. B406, pp. 187–224, 1993.

[131] S. D. Ellis and D. E. Soper, "Successive combination jet algorithm for hadron collisions," *Phys. Rev.*, vol. D48, pp. 3160–3166, 1993.

[132] M. Wobisch and T. Wengler, "Hadronization corrections to jet cross-sections in deep inelastic scattering," in *Monte Carlo generators for HERA physics. Proceedings, Workshop, Hamburg, Germany, 1998-1999*, 1998, pp. 270–279.

[133] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-k_t jet clustering algorithm," *JHEP*, vol. 04, p. 063, 2008.

[134] ——, "FastJet user manual," *Eur. Phys. J. C*, vol. 72, no. arXiv:1111.6097. CERN-PH-TH-2011-297, p. 1896. 69 p, Nov 2011, comments: 69 pages. FastJet 3 is available from http://fastjet.fr/. [Online]. Available: https://cds.cern.ch/record/1402449

[135] G. P. Salam, "Towards Jetography," *Eur. Phys. J.*, vol. C67, pp. 637–686, 2010.

[136] M. Cacciari, G. P. Salam, and G. Soyez, "The catchment area of jets," *Journal of High Energy Physics*, vol. 2008, no. 04, pp. 005–005, apr 2008. [Online]. Available: https://doi.org/10.1088/1126-6708/2008/04/005

[137] V. Coco, P.-A. Delsart, J. Rojo-Chacon, G. Soyez, and C. Sander, "Jets and jet algorithms," in *HERA and the LHC: 4th Workshop on the Implications of HERA for LHC Physics*, 3 2009, pp. 182–204.

[138] Rationality., *Oxford English Dictionary.* Oxford University Press, 1989. [Online]. Available: https://www.oed.com/view/Entry/53038082?redirectedFrom=in+situ#eid

[139] G. Aad, B. Abbott, J. Abdallah, A. A. Abdelalim, A. Abdesselam, O. Abdinov, B. Abi, M. Abolins, H. Abramowicz, and et al., "Jet energy measurement with the atlas detector in proton-proton collisions at $\sqrt{s} = 7$ TeV," *The European Physical Journal C*, vol. 73, no. 3, Mar 2013. [Online]. Available: http://dx.doi.org/10.1140/epjc/s10052-013-2304-2

[140] S. Schmitt, "Data unfolding methods in high energy physics," *EPJ Web of Conferences*, vol. 137, p. 11008, 2017. [Online]. Available: http://dx.doi.org/10.1051/epjconf/201713711008

[141] G. D'Agostini, "A multidimensional unfolding method based on Bayes' Theorem," DESY, Hamburg, Tech. Rep., Jun 1994. [Online]. Available: http://cds.cern.ch/record/265717

[142] T. S. Kuhn, *The structure of scientific revolutions.* Chicago: University of Chicago Press, 1970.

[143] L. Galvani, "De viribus electricitatis in motu musculari: commentarius." 1791.

[144] J. Bernstein, *Untersuchungen über den Erregungsvorgang im Nerven- und Muskelsysteme.* Carl Winter's Universitätsbuchhandlung, Heidelberg, 1871.

[145] S. Ramón y Cajal, "Revista trimestral de histología normal y patológica, año 1, n. 1," 1888.

[146] W. von Waldeyer-Hartz, *Ueber einige neuere Forschungen im Gebiete der Anatomie des Centralnervensystems*, ser. Wochenschrift, Deutsche Medicinische 1891, No. 44 u. ff. Georg Thieme, 1891.

[147] W. Mcculloch and W. Pitts, "A logical calculus of ideas immanent in nervous activity," *Bulletin of Mathematical Biophysics*, vol. 5, pp. 127–147, 1943.

[148] D. O. Hebb, *The organization of behavior: A neuropsychological theory.* New York: Wiley, Jun. 1949.

[149] F. Rosenblatt, *Principles of neurodynamics: perceptrons and the theory of brain mechanisms*, ser. Report (Cornell Aeronautical Laboratory). Spartan Books, 1962. [Online]. Available: https://books.google.de/books?id=7FhRAAAAMAAJ

[150] M. Minsky and S. Papert, *Perceptrons: An Introduction to Computational Geometry.* Cambridge, MA, USA: MIT Press, 1969.

[151] J. J. Hopfield, "Neurocomputing: Foundations of research," J. A. Anderson and E. Rosenfeld, Eds. Cambridge, MA, USA: MIT Press, 1988, ch. Neural Networks and Physical Systems with Emergent Collective Computational Abilities, pp. 457–464. [Online]. Available: http://dl.acm.org/citation.cfm?id=65669.104422

[152] J. J. Hopfield and D. W. Tank, ""neural" computation of decisions in optimization problems," *Biol. Cybern.*, vol. 52, no. 3, pp. 141–152, Jul. 1985. [Online]. Available: http://dx.doi.org/10.1007/BF00339943

[153] S. Linnainmaa, "The representation of the cumulative rounding error of an algorithm as a Taylor expansion of the local rounding errors," Master's thesis, Univ. Helsinki, 1970.

[154] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, pp. 533–, Oct. 1986. [Online]. Available: http://dx.doi.org/10.1038/323533a0

[155] G. Cybenko, "Approximation by superpositions of a sigmoidal function," 1989.

[156] K. Hornik, "Approximation capabilities of multilayer feedforward networks." *Neural Networks*, vol. 4, no. 2, pp. 251–257, 1991. [Online]. Available: http://dblp.uni-trier.de/db/journals/nn/nn4.html#Hornik91

[157] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997. [Online]. Available: http://dx.doi.org/10.1162/neco.1997.9.8.1735

[158] *A Training Algorithm for Optimal Margin Classifiers*, ser. COLT '92. New York, NY, USA: ACM, 1992. [Online]. Available: http://doi.acm.org/10.1145/130385.130401

[159] C. J. C. H. Watkins and P. Dayan, "Technical note q-learning," *Machine Learning*, vol. 8, pp. 279–292, 1992. [Online]. Available: https://doi.org/10.1007/BF00992698

[160] Y. Li, "Deep reinforcement learning: An overview," *CoRR*, vol. abs/1701.07274, 2017. [Online]. Available: http://arxiv.org/abs/1701.07274

[161] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in Neural Information Processing Systems 27*, Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2014, pp. 2672–2680. [Online]. Available: http://papers.nips.cc/paper/5423-generative-adversarial-nets.pdf

[162] D. H. Ballard, "Modular learning in neural networks," in *Proceedings of the Sixth National Conference on Artificial Intelligence - Volume 1*, ser. AAAI'87. AAAI Press, 1987, pp. 279–284. [Online]. Available: http://dl.acm.org/citation.cfm?id=1863696.1863746

[163] D. Kriesel, *A Brief Introduction to Neural Networks*, 2007. [Online]. Available: availableathttp://www.dkriesel.com

[164] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proceedings of the 27th International Conference on International Conference on Machine Learning*, ser. ICML'10. USA: Omnipress, 2010, pp. 807–814. [Online]. Available: http://dl.acm.org/citation.cfm?id=3104322.3104425

[165] A. L. Maas, "Rectifier nonlinearities improve neural network acoustic models," 2013.

[166] C. Nwankpa, W. Ijomah, A. Gachagan, and S. Marshall, "Activation functions: Comparison of trends in practice and research for deep learning," *CoRR*, vol. abs/1811.03378, 2018. [Online]. Available: http://arxiv.org/abs/1811.03378

[167] B. Xu, N. Wang, T. Chen, and M. Li, "Empirical evaluation of rectified activations in convolutional network," *CoRR*, vol. abs/1505.00853, 2015. [Online]. Available: http://arxiv.org/abs/1505.00853

[168] O. Vinyals, S. Bengio, and M. Kudlur, "Order matters: Sequence to sequence for sets," *CoRR*, vol. abs/1511.06391, 2016.

[169] P. Gras, S. Höche, D. Kar, A. Larkoski, L. Lönnblad, S. Plätzer, A. Siódmok, P. Skands, G. Soyez, and J. Thaler, "Systematics of quark/gluon tagging," *Journal of High Energy Physics (Online)*, vol. 2017, no. 7, 7 2017.

[170] W. Deming, *The New Economics: For Industry, Government, Education*, ser. Mit Press. MIT Press, 2000. [Online]. Available: https://books.google.de/books?id=RnsCXffehcEC

[171] A. J. Larkoski, G. P. Salam, and J. Thaler, "Energy Correlation Functions for Jet Substructure," *JHEP*, vol. 06, p. 108, 2013.

[172] N. Garelli, "Performance of the ATLAS Detector in Run-2," *EPJ Web Conf.*, vol. 164, p. 01021. 10 p, 2017. [Online]. Available: https://cds.cern.ch/record/2310929

[173] "Welcome to PYTHIA," https://pythia.org/, accessed: 2022-10-20.

[174] T. Sjöstrand, S. Mrenna, and P. Skands, "A brief introduction to PYTHIA 8.1," *Computer Physics Communications*, vol. 178, no. 11, pp. 852–867, jun 2008. [Online]. Available: https://doi.org/10.1016%2Fj.cpc.2008.01.036

[175] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands, "An introduction to PYTHIA 8.2," *Computer Physics Communications*, vol. 191, pp. 159–177, jun 2015. [Online]. Available: https://doi.org/10.1016%2Fj.cpc.2015.01.024

[176] T. Sjöstrand, S. Mrenna, and P. Skands, "PYTHIA 6.4 physics and manual," *Journal of High Energy Physics*, vol. 2006, no. 05, pp. 026–026, may 2006. [Online]. Available: https://doi.org/10.1088%2F1126-6708%2F2006%2F05%2F026

[177] R. D. Ball, V. Bertone, S. Carrazza, C. S. Deans, L. D. Debbio, S. Forte, A. Guffanti, N. P. Hartland, J. I. Latorre, J. Rojo, and M. Ubiali, "Parton distributions with LHC data," *Nuclear Physics B*, vol. 867, no. 2, pp. 244–289, feb 2013. [Online]. Available: https://doi.org/10.1016%2Fj.nuclphysb.2012.10.003

[178] "ATLAS Pythia 8 tunes to 7 TeV data," CERN, Geneva, Tech. Rep., Nov 2014, all figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2014-021. [Online]. Available: https://cds.cern.ch/record/1966419

[179] C. Chen, M. Cao, X. Hu, and S. Peng, "Key person aided re-identification in partially ordered pedestrian set," *CoRR*, vol. abs/1805.10017, 2018. [Online]. Available: http://arxiv.org/abs/1805.10017

[180] P. Nason, "A New method for combining NLO QCD with shower Monte Carlo algorithms," *JHEP*, vol. 11, p. 040, 2004.

[181] S. Frixione, P. Nason, and C. Oleari, "Matching NLO QCD computations with Parton Shower simulations: the POWHEG method," *JHEP*, vol. 11, p. 070, 2007.

[182] J. Alwall *et al.*, "A Standard format for Les Houches event files," *Comput. Phys. Commun.*, vol. 176, pp. 300–304, 2007.

[183] T. Gleisberg, S. Höche, F. Krauss, M. Schönherr, S. Schumann, F. Siegert, and J. Winter, "Event generation with SHERPA 1.1," *Journal of High Energy Physics*, vol. 2009, no. 02, pp. 007–007, feb 2009. [Online]. Available: https://doi.org/10.1088/1126-6708/2009/02/007

[184] "Sherpa Homepage," https://sherpa-team.gitlab.io/, accessed: 2022-10-20.

[185] S. Schumann and F. Krauss, "A Parton shower algorithm based on Catani-Seymour dipole factorisation," *JHEP*, vol. 03, p. 038, 2008.

[186] S. Dulat, T.-J. Hou, J. Gao, M. Guzzi, J. Huston, P. Nadolsky, J. Pumplin, C. Schmidt, D. Stump, and C. P. Yuan, "New parton distribution functions from a global analysis of quantum chromodynamics," *Phys. Rev. D*, vol. 93, no. 3, p. 033006, 2016.

[187] H.-L. Lai, M. Guzzi, J. Huston, Z. Li, P. M. Nadolsky, J. Pumplin, and C. P. Yuan, "New parton distributions for collider physics," *Phys. Rev. D*, vol. 82, p. 074024, 2010.

[188] J.-C. Winter, F. Krauss, and G. Soff, "A Modified cluster hadronization model," *Eur. Phys. J. C*, vol. 36, pp. 381–395, 2004.

[189] T. Sjostrand, S. Mrenna, and P. Z. Skands, "PYTHIA 6.4 Physics and Manual," *JHEP*, vol. 05, p. 026, 2006.

[190] "The Herwig Event Generator," https://herwig.hepforge.org/, accessed: 2022-10-20.

[191] M. Bahr *et al.*, "Herwig++ Physics and Manual," *Eur. Phys. J.*, vol. C58, pp. 639–707, 2008.

[192] J. Bellm *et al.*, "Herwig 7.0/Herwig++ 3.0 release note," *Eur. Phys. J.*, vol. C76, no. 4, p. 196, 2016.

[193] ——, "Herwig 7.1 Release Note," 5 2017.

[194] L. A. Harland-Lang, A. D. Martin, P. Motylinski, and R. S. Thorne, "Parton distributions in the LHC era: MMHT 2014 PDFs," *Eur. Phys. J. C*, vol. 75, no. 5, p. 204, 2015.

[195] S. Platzer and S. Gieseke, "Dipole Showers and Automated NLO Matching in Herwig++," *Eur. Phys. J. C*, vol. 72, p. 2187, 2012.

[196] S. Frixione and B. R. Webber, "Matching NLO QCD computations and parton shower simulations," *JHEP*, vol. 06, p. 029, 2002.

[197] "Early Inner Detector Tracking Performance in the 2015 data at $\sqrt{s} = 13$ TeV," CERN, Geneva, Tech. Rep., Dec 2015, all figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2015-051. [Online]. Available: http://cds.cern.ch/record/2110140

[198] P. Zyla *et al.*, "Review of Particle Physics," *PTEP*, vol. 2020, no. 8, p. 083C01, 2020.

[199] R. Brun, F. Rademakers, P. Canal, A. Naumann, O. Couet, L. Moneta, V. Vassilev, S. Linev, D. Piparo, G. GANIS, B. Bellenot, E. Guiraud, G. Amadio, wverkerke, P. Mato, TimurP, M. Tadel, wlav, E. Tejedor, J. Blomer, A. Gheata, S. Hageboeck, S. Roiser, marsupial, S. Wunsch, O. Shadura, A. Bose, CristinaCristescu, X. Valls, and R. Isemann, "root-project/root: v6.18/02," Aug. 2019. [Online]. Available: https://doi.org/10.5281/zenodo.3895860

[200] A. Collaboration, "Athena," Apr. 2019. [Online]. Available: https://doi.org/10.5281/zenodo.2641997

[201] "Selection of jets produced in 13TeV proton-proton collisions with the ATLAS detector," CERN, Geneva, Tech. Rep., Jul 2015, all figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2015-029. [Online]. Available: https://cds.cern.ch/record/2037702

[202] G. Aad *et al.*, "Properties of jet fragmentation using charged particles measured with the ATLAS detector in pp collisions at $\sqrt{s} = 13$ TeV," *Phys. Rev. D*, vol. 100, no. 5, p. 052011, 2019.

[203] "Early Inner Detector Tracking Performance in the 2015 data at $\sqrt{s} = 13$ TeV," 2015.

[204] R. D. Ball, V. Bertone, F. Cerutti, L. Del Debbio, S. Forte, A. Guffanti, J. I. Latorre, J. Rojo, and M. Ubiali, "Impact of Heavy Quark Masses on Parton Distributions and LHC Phenomenology," *Nucl. Phys. B*, vol. 849, pp. 296–363, 2011.

[205] "NNPDF," http://nnpdf.mi.infn.it/, accessed: 2022-10-21.

[206] P. M. Nadolsky, H.-L. Lai, Q.-H. Cao, J. Huston, J. Pumplin, D. Stump, W.-K. Tung, and C. P. Yuan, "Implications of CTEQ global analysis for collider observables," *Phys. Rev. D*, vol. 78, p. 013004, 2008.

[207] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt, "Parton distributions for the LHC," *Eur. Phys. J. C*, vol. 63, pp. 189–285, 2009.

[208] G. Aad *et al.*, "Measurement of the Lund Jet Plane Using Charged Particles in 13 TeV Proton-Proton Collisions with the ATLAS Detector," *Phys. Rev. Lett.*, vol. 124, no. 22, p. 222002, 2020.

[209] S. D. Ellis, C. K. Vermilion, and J. R. Walsh, "Recombination Algorithms and Jet Substructure: Pruning as a Tool for Heavy Particle Searches," *Phys. Rev. D*, vol. 81, p. 094023, 2010.

[210] ——, "Techniques for improved heavy particle searches with jet substructure," *Phys. Rev. D*, vol. 80, p. 051501, 2009.

[211] D. Krohn, J. Thaler, and L.-T. Wang, "Jet Trimming," *JHEP*, vol. 02, p. 084, 2010.

[212] J. M. Butterworth, A. R. Davison, M. Rubin, and G. P. Salam, "Jet substructure as a new Higgs search channel at the LHC," *Phys. Rev. Lett.*, vol. 100, p. 242001, 2008.

[213] M. Dasgupta and G. P. Salam, "Resummation of nonglobal QCD observables," *Phys. Lett. B*, vol. 512, pp. 323–330, 2001.

[214] A. J. Larkoski, S. Marzani, G. Soyez, and J. Thaler, "Soft Drop," *JHEP*, vol. 05, p. 146, 2014.

[215] B. Andersson, G. Gustafson, L. Lonnblad, and U. Pettersson, "Coherence Effects in Deep Inelastic Scattering," *Z. Phys. C*, vol. 43, p. 625, 1989.

[216] J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lemaître, A. Mertens, and M. Selvaggi, "DELPHES 3, A modular framework for fast simulation of a generic collider experiment," *JHEP*, vol. 02, p. 057, 2014.

[217] M. Selvaggi, "DELPHES 3: A modular framework for fast-simulation of generic collider experiments," *J. Phys. Conf. Ser.*, vol. 523, p. 012033, 2014.

[218] A. Mertens, "New features in Delphes 3," *J. Phys. Conf. Ser.*, vol. 608, no. 1, p. 012045, 2015.

[219] L. Heinrich, M. Feickert, and E. Rodrigues, "pylhe: v0.3.0." [Online]. Available: https://github.com/scikit-hep/pylhe

[220] S. Ovyn, X. Rouby, and V. Lemaitre, "DELPHES, a framework for fast simulation of a generic collider experiment," 3 2009.

[221] A. J. Larkoski, G. P. Salam, and J. Thaler, "Energy correlation functions for jet substructure," *Journal of High Energy Physics*, vol. 2013, no. 6, jun 2013. [Online]. Available: https://doi.org/10.1007%2Fjhep06%282013%29108

[222] J. Thaler and K. V. Tilburg, "Identifying boosted objects with n-subjettiness," *Journal of High Energy Physics*, vol. 2011, no. 3, mar 2011. [Online]. Available: https://doi.org/10.1007%2Fjhep03%282011%29015

[223] ——, "Maximizing boosted top identification by minimizing n-subjettiness," *Journal of High Energy Physics*, vol. 2012, no. 2, feb 2012. [Online]. Available: https://doi.org/10.1007%2Fjhep02%282012%29093

[224] F. Dreyer, G. Salam, and G. Soyez, "The lund jet plane," 07 2018.

[225] O. Vinyals, M. Fortunato, and N. Jaitly, "Pointer networks," 2015. [Online]. Available: https://arxiv.org/abs/1506.03134

[226] P. T. Komiske, E. M. Metodiev, and J. Thaler, "Energy Flow Networks: Deep Sets for Particle Jets," *JHEP*, vol. 01, p. 121, 2019.

[227] F. Chollet *et al.* (2015) Keras. [Online]. Available: https://github.com/fchollet/keras

[228] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015, software available from tensorflow.org. [Online]. Available: https://www.tensorflow.org/

[229] "EnergyFlow Python Package," https://energyflow.network/, accessed: 2022-04-03.

[230] O. R. developers, "Onnx runtime," https://onnxruntime.ai/, 2021, version: x.y.z.

[231] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, cite arxiv:1412.6980Comment: Published as a conference paper at the 3rd International Conference for Learning Representations, San Diego, 2015. [Online]. Available: http://arxiv.org/abs/1412.6980

[232] E. Pennisi, "Seeking life's bare (genetic) necessities," *Science*, vol. 272, no. 5265, pp. 1098–1099, 1996. [Online]. Available: https://www.science.org/doi/abs/10.1126/science.272.5265.1098