Precision measurement of the inclusive b jet production in proton-proton collisions with the CMS experiment at the LHC at $\sqrt{s} = 13$ TeV

Dissertation
zur Erlangung des Doktorgrades
an der Fakultät für Mathematik, Informatik und Naturwissenschaften
Fachbereich Physik
der Universität Hamburg

vorgelegt von

Patrick L.S. Connor

aus Brüssel, Belgien

Hamburg
2018

Zusammenfassung

Der doppel-differentiezellen Querschnitt für inclusive b-Jetproduktion in Proton-Proton Kollisionen sowie den Verhältnis von b-Jets zu inclusiver Jetproduktion werden als Funktion des Transversalimpulses p_T und der absoluten Rapidität $|y|$ gemessen. Die Daten wurden in 2016 bei einer Schwerpunktenergie von 13 TeV und einer integrierten Luminosität von $35.2\,\mathrm{fb}^{-1}$ mit dem CMS Detektor am LHC gesammelt. Die Jets müssen im Bereich $p_T > 74\,\mathrm{GeV}$ und $|y| < 2.4$ liegen; die b-Jets müssen ein B-Hadron enthalten. Die Messung hat ausreichende Statistik bis $p_T \sim \mathcal{O}(\mathrm{TeV})$. Anspruchsvolle Entfaltungsmethoden werden für die Bestimmung des Signals genutzt. Es wird gezeigt, dass Matrixelementberechnungen in NLO zusammen mit der Simulation des Underlying-Events die Messungen sehr gut beschreiben.

Abstract

A measurement of the double differential cross section for inclusive b jet production in proton-proton collisions as well as fraction of b jets in the inclusive jet production is presented as a function of the transverse momentum p_T and the absolute rapidity $|y|$. The data samples were collected in the CMS experiment at LHC during 2016 and correspond to an integrated luminosity of $35.2\,\mathrm{fb}^{-1}$ at a centre-of-mass energy of 13 TeV. The jets are selected with $p_T > 74\,\mathrm{GeV}$ and $|y| < 2.4$; the b jets must contain a B hadron. The measurement has significant statistics up to $p_T \sim \mathcal{O}(\mathrm{TeV})$. Advanced methods of unfolding are performed to extract the signal. It is found that fixed-order calculations with underlying event describe the measurement well.

Eidesstattliche Versicherung

Hiermit versichere ich an Eides statt, die vorliegende Dissertationsschrift selbst verfasst und keine anderen als die angegebenen Hilfsmittel und Quellen benutzt zu haben. Die eingereichte schriftliche Fassung entspricht der auf dem elektronischen Speichermedium. Die Dissertation wurde in der vorgelegten oder einer ähnlichen Form nicht schon einmal in einem früheren Promotionsverfahren angenommen oder als ungenügend beurteilt.

Hamburg, den 1. Februar 2018,

Patrick L.S. Connor

Gutachter der Dissertation:	Prof. Dr. Johannes Haller Priv.-Doz. Dr. Hannes Jung
Zusammensetzung der Prüfungskommission:	Prof. Dr. Dieter Horns Prof. Dr. Caren Hagner
Vorsitzender der Prüfungskommission:	Prof. Dr. Sven-Olaf Moch
Datum der Disputation:	28. März 2018
Vorsitzender Fach-Promotionsausschusses Physik:	Prof. Dr. Wolfgang Hansen
Leiter des Fachbereichs Physik:	Prof. Dr. Michael Potthoff
Dekan der Fakultät MIN:	Prof. Dr. Heinrich Graener

« La pensée
ne doit jamais
se soumettre,
ni à un dogme,
ni à un parti,
ni à une passion,
ni à un intérêt,
ni à une idée préconçue,
ni à quoi que ce soit,
si ce n'est
aux **faits** eux-mêmes,
parce que,
pour elle,
se soumettre,
ce serait cesser d'être. »

Henri Poincaré

Speech for the 75th anniversary
of the *Université libre de Bruxelles*

21 November 1909

给王群

Preamble

More than any other field of science, the complexity of the theory and the size of the experiments in Particle Physics require a wide knowledge of many different topics as well as large collaborations to face the unprecedented amount of data to acquire and to analyse. The training of the physicist comprises maths, fundamental physics, detector physics, modelling, data reduction and programming. Today's collaborations gather up to a few thousands of analysts and engineers per experiment; fifty years were needed between the theoretical prediction of the so-called Higgs boson and its experimental discovery; around two hundred papers were submitted when, in 2016, a new resonance was believed to appear around 600 GeV in the invariant mass spectrum of the diphoton production. One sometimes evokes Particle Physics as the physics of the extremes; I think this does not only have to do with the quantum and relativistic scales.

From October 2014 to March 2018, I had the chance to participate to various aspects of research in Particle Physics under the supervision of Hannes Jung. Initially, following up on the main topic of my master thesis, I was supposed to work on the Drell-Yan process and associated jet production, and started some Monte Carlo investigations. But the work really first started with contributions to the measurement of the Minimum Bias cross section at LHC Run-II. In parallel, I started my service tasks in the group dedicated to the alignment of the tracker system at CMS, taking advantage of the strong involvement of the DESY group in this area; I continued this activity throughout the whole duration of the doctorate. After the minimum bias analysis, we decided together with Hannes Jung to change the main topic of the thesis to b jet measurements, more appropriate with respect to the other activities in the group since the departure of a member of our group with whom I was supposed to work. Initially, the new project consisted in the investigation of the associated jet production to $b\bar{b}$ production, starting with "some quick analysis on the inclusive b jet production". It turned out, however, that the b jet production at the TeV scale was non-trivial, and became the main topic of the thesis; this will be the object of the present memoir. Finally, in parallel of this work, I redeveloped from scratch the `TMDplotter`, an on-line facility for plotting parton densities; I also participated as a teaching assistant the exercise sessions of the MC lectures by Hannes Jung, in Hamburg and in Antwerp; and I took an active part in the organisation of internal paper reading sessions at DESY.

The present thesis is organised in three parts. The first part covers a general introduction: historical, theoretical and experimental; running across this different aspects, it aims at offering a general understanding of the stakes in Particle Physics, together with a motivation of the study of these famous b jets. Here, the aim is not to substitute to manuals, but to draw up the main lines of Particle Physics in a more complete way than a list of references, and to introduce some concepts and formulae — maybe a bit more — that will be useful in the following. Benefiting from the concepts introduced in the first part,

and following a logic of going from the top to the bottom, the second part will treat of the analysis of data at CMS from the LHC taken in 2016; it will be more technical and specific, representing the bulk of the work that I have performed along this doctorate in the topic of b jets. Questions regarding various calibrations, modern investigations in b tagging and advanced unfolding techniques will be addressed there; the literature about these not always being easy to approach, the detail in which they are described in this thesis was also reached in the hope and the wish of being useful to future students. In addition to these two parts, a third part, much shorter, will be dedicated to discussions about our current knowledge about b jets, and prospects about future measurements and predictions; finally, appendices will close off this thesis, complementing previous chapters or describe additional activities performed during this *Promotion* at DESY: conferences and service work, especially in tracker alignment.

Although completely transparent through this thesis, another aspect of great impact in modern Particle Physics is the ever increasing need in computing skills, essential to treat large amount of data in an efficient way and to collaborate with other physicists. It is not a coincidence if the modern World Wide Web was developed at CERN in the context of Particle Physics in 1991. Advanced programming skills are nowadays essential to conduct an analysis successfully, and since schools and universities still rarely offer their proper teaching, it important to stress that a signifiant part of the time spent before and at DESY had to be dedicated to their self-learning, especially to that of `C++`.

Lastly, before getting down to the business, I would make a mistake if I would not mention here a non-scientific but nonetheless non-negligible aspect of working in a large collaboration such as the ones in Particle Physics. Indeed, since it gathers physicists from all around the world, Particle Physics is also made of social interactions, together with its cultural shocks; the social aspect of the work in collaboration has three entangled consequences. The first consequence is that politics is a significant component of research in Particle Physics, sometimes more driving choices than scientific arguments; as much as I have been able to do so, I have tried to avoid politics and attempted to follow the philosophy of my *Alma Mater*, the *Université libre de Bruxelles*, very well summarised in Poincaré's quotation given in the first pages [1]. The second one is that this thesis was to be written in the self-proclaimed international language, English, although by a Belgian graduating in Germany, having therefore *a priori* few to do with this language (leaving aside the Irish name...); on this point, I would like to mention that I chose to use the European spelling rather than the American, which, as a corollary of the previous point, is a political choice. The third and last point is that the confrontation of cultures implies the confrontation of ideas and approaches, not only in science but also in daily life; in my case, it has first implied living in Germany and learning its language. This has for sure influenced me, and if I had been to write this thesis in Belgium, it would likely look quite different. But more than anything else — and I will finish the preamble there —, it has also meant meeting my fiancée, Wang Qun; and despite the great interest that I have had for the topic of the present volume, our encounter is certainly the most important thing that Particle Physics will have ever granted to me.

1. Translation to English: "Thinking must never submit itself neither to a dogma, nor to a political party, nor to a passion, nor to an interest, nor to a preconceived idea, nor to anything except facts themselves, because for it to submit would mean cease existing."

Contents

Abstract i

Eidesstattliche Versicherung ii

Quotation & Dedication iii

Preamble v

Contents 1

I Premisses 9

1 Introduction to High Energy Physics 11
1.1 Introduction . . . 11
1.1.1 Units . . . 11
1.1.2 Fundamental interactions . . . 12
1.1.3 Particles . . . 12
1.2 History . . . 14
1.2.1 Atomism (10^{-10} m) . . . 14
1.2.2 Nucleus (10^{-13} m) . . . 14
1.2.3 Nucleus structure (10^{-15} m) . . . 14
1.2.4 Nucleon substructure (10^{-18} m) . . . 15
1.3 Experimentation . . . 16
1.3.1 Sources . . . 16
1.3.2 Observables . . . 16
1.4 Challenges . . . 17
In this thesis . . . 17
1.A Dark matter and dark energy . . . 17
1.A.1 Dark matter . . . 18
1.A.2 Dark energy . . . 18
References . . . 18

2 Elements of theory 21
2.1 Introduction . . . 21
2.2 The Standard Model of High Energy Physics . . . 22
2.2.1 Definition . . . 22
2.2.1.1 Quantum field theory . . . 23
2.2.1.2 Local Gauge invariance . . . 23
2.2.1.3 Renormalisation . . . 25

2.2.1.4 Parameters . . . 25
2.2.2 Lagrangian . . . 26
2.2.2.1 Electroweak sector . . . 27
2.2.2.2 Quantum Chromodynamics sector . . . 28
2.2.2.3 Yukawa and Higgs sectors . . . 28
2.2.3 Running couplings . . . 28
2.2.4 Calculations with Feynman diagrams in the perturbative regime . . . 29
2.2.4.1 Example . . . 32
2.2.4.2 Vocabulary . . . 33
2.2.5 Current limits of calculations within the SM . . . 34
2.2.5.1 Non-perturbative QCD . . . 35
2.2.5.2 Complexity of perturbative calculations . . . 35
2.2.5.3 High-multiplicity final states . . . 35
2.3 Phenomenology . . . 36
2.3.1 Factorisation . . . 36
2.3.2 Objects in the initial and final states . . . 38
2.3.2.1 Hadrons . . . 38
2.3.2.2 Jets . . . 39
2.3.3 Overview of a proton-proton scattering at LHC . . . 41
2.3.3.1 Hard process . . . 42
2.3.3.2 Parton Shower . . . 44
2.3.3.3 Multi-Parton Interactions and Beam-Beam Remnants . . . 46
2.3.3.4 Hadronisation . . . 50
2.3.3.5 Stable particles . . . 51
2.3.4 More on evolution . . . 51
2.3.4.1 BFKL evolution . . . 52
2.3.4.2 CCFM evolution . . . 53
Conclusion . . . 53
2.A Historical perspective . . . 54
2.B Variables and coordinates . . . 58
2.B.1 Coordinate system . . . 58
2.B.2 Mandelstam variables . . . 59
2.B.2.1 Definition . . . 59
2.B.3 Channels. . . . 59
2.B.4 Properties . . . 60
References . . . 60

3 The Large Hadron Collider and the Compact Muon Solenoid 67
3.1 Large Hadron Collider . . . 67
3.1.1 Overview . . . 69
3.1.2 Principles of acceleration . . . 71
3.1.3 Collisions . . . 74
3.2 Compact Muon Solenoid . . . 75
3.2.1 General principles of detection and reconstruction . . . 79
3.2.2 Application at CMS . . . 80
3.2.2.1 Tracking and vertexing . . . 82
3.2.2.2 Calorimetry . . . 85
3.2.2.3 Muon detection . . . 88
3.2.2.4 Global event reconstruction . . . 89

3.2.2.5 Trigger and storage systems . . . 89
Summary . . . 90
References . . . 90

4 Monte Carlo techniques and Physics Generators 93
4.1 Introduction . . . 93
4.1.1 Reminder on probability theory . . . 93
4.1.2 Simple Monte Carlo integration . . . 94
4.1.3 Importance sampling . . . 95
4.2 Physics generators . . . 96
4.2.1 Fixed-order parton-level calculations . . . 96
4.2.2 Event generators . . . 97
4.2.2.1 General-purpose event generators . . . 97
4.2.2.2 Tuning . . . 98
4.3 Detector simulation . . . 99
Conclusion . . . 100
4.A Further methods of MC integration . . . 100
References . . . 101

5 Discovery, overview and motivation of beauty physics 105
5.1 Discovery and first measurements . . . 105
5.1.1 The E288 experiment at Fermilab . . . 105
5.1.2 ISR . . . 106
5.1.3 DORIS . . . 107
5.1.4 CESR . . . 108
5.2 Further investigations . . . 108
5.3 Physics at hadron colliders with b's . . . 109
5.3.1 Heavy-flavour jets . . . 111
5.3.1.1 Definition . . . 111
5.3.1.2 Tagging . . . 111
5.3.2 Mechanism of production of **b**'s with QCD processes . . . 113
5.3.3 **b** as a test for QCD . . . 115
5.3.3.1 Evolution . . . 116
5.3.3.2 Flavour democracy . . . 116
5.3.3.3 Phenomenology . . . 117
5.3.4 **b** as a probe for other processes . . . 118
5.3.4.1 Top physics . . . 118
5.3.4.2 Higgs physics . . . 118
5.3.4.3 Searches . . . 119
Conclusion . . . 120
References . . . 120

II Physics analysis 125

6 General strategy and outline of the analysis 127
6.1 Monte Carlo studies . . . 127
6.1.1 Contributions from QCD processes . . . 127
6.1.2 Other processes . . . 130
6.2 Previous measurements of the inclusive b jet production at the LHC . . . 130

6.2.1 ATLAS at 7 TeV . . . 130
6.2.2 CMS at 7 TeV . . . 130
6.3 Strategy . . . 131
6.3.1 Phase space definition and selection . . . 132
Convention . . . 134
6.3.2 Samples . . . 134
6.3.2.1 Data . . . 135
6.3.2.2 Simulation . . . 136
6.3.3 Analysis at detector level . . . 136
6.3.4 Unfolding to particle level . . . 137
6.3.5 Theoretical predictions . . . 140
References . . . 140

7 Analysis at detector level **143**
7.1 Trigger . . . 146
7.1.1 Trigger strategy . . . 146
7.1.1.1 Determination of the trigger efficiency . . . 147
7.1.1.2 Final subdivision of the phase space . . . 148
7.1.1.3 Run-by-run stability . . . 149
7.1.2 Effective luminosity and average pre-scales . . . 149
7.2 Pile-up . . . 154
7.2.1 General considerations on the pile-up . . . 154
7.2.1.1 Measurement of pile-up in data . . . 155
7.2.1.2 Simulation of the pile-up . . . 155
7.2.2 Removal of overweighted events in simulation . . . 156
7.2.3 Procedure of pile-up reweighting . . . 156
7.3 Jets . . . 160
7.3.1 Jet Energy Scale Correction . . . 160
7.3.2 Jet Energy Resolution . . . 161
7.3.2.1 JER smearing . . . 161
7.3.2.2 Differential resolution and binning . . . 164
7.4 Tagging . . . 166
7.4.1 Taggers at CMS . . . 166
7.4.1.1 Combined-Secondary-Vertex (`CSVv2`) . . . 166
7.4.1.2 Jet Probability . . . 167
7.4.1.3 combined-Multi-Variate-Analysis (`cMVAv2`) . . . 167
7.4.2 Performance of the taggers . . . 168
7.4.2.1 Calibration . . . 168
7.4.2.2 Performance . . . 170
Conclusions . . . 171
7.A Missing Transverse Energy . . . 175
7.A.1 Detector studies . . . 175
7.A.1.1 MET variables . . . 175
7.A.1.2 Spectrum of reconstructed transverse momentum . . . 177
7.A.2 Generator studies . . . 177
7.B Details of trigger efficiency . . . 180
7.C Additional control plots about jets . . . 186
7.D More on tagging . . . 187
7.D.1 Comparison of the working points . . . 187

7.D.2 Comparison of the taggers . . . 187
7.E Jet constituents . . . 188
7.E.1 Jet ID . . . 188
7.E.2 Jet constituents in bins of rapidity . . . 188
References . . . 189

8 Analysis at particle level **203**
8.1 Purity . . . 203
8.1.1 Origin of the disagreement . . . 204
8.1.2 Method of the determination of the purity . . . 204
8.1.3 Results . . . 206
8.2 Unfolding . . . 208
8.2.1 Response of the detector . . . 208
8.2.1.1 Response matrix and probability matrix . . . 208
8.2.1.2 Matching . . . 209
8.2.1.3 Miss and fake jets . . . 211
8.2.1.4 Description of the migrations . . . 211
8.2.1.5 Uncertainties . . . 213
8.2.1.6 Model reweighting . . . 213
8.2.2 Procedure . . . 213
8.2.2.1 Origin of the instabilities . . . 214
8.2.2.2 Tikhonov regularisation . . . 214
8.2.2.3 Regularisation matrix . . . 215
8.2.2.4 Regularisation parameter . . . 216
8.2.2.5 Closure test . . . 216
8.2.2.6 Results in data . . . 218
8.2.3 D'Agostini unfolding . . . 219
8.2.3.1 Principles . . . 219
8.2.3.2 Results . . . 224
Conclusions . . . 224
8.A Details about fit of purity . . . 226
8.A.1 Templates . . . 226
8.A.2 Determination of the purity in the CSVv2-tagged region . . . 226
8.A.3 Determination of the purity in the non-CSVv2-tagged region . . . 226
8.B Details about unfolding procedure . . . 238
8.B.1 Control plots for Tikhonov regularisation . . . 238
8.B.2 Treatment of statistical uncertainties . . . 238
8.B.3 Additional checks . . . 239
8.B.3.1 Backfolding . . . 239
8.B.3.2 χ^2 of agreement . . . 245
8.B.3.3 χ^2 of change . . . 245
8.B.3.4 Bottom Line Test . . . 245
References . . . 247

9 Results **249**
9.1 Comparison to LO predictions . . . 249
9.2 Comparison to NLO predictions . . . 251
9.2.1 Theoretical predictions . . . 251
9.2.2 Comparison . . . 251
9.2.3 Contributions to the prediction . . . 260

9.A List of tune parameters . . . 260
References . . . 261

III Conclusions **271**

10 Summary, Conclusions and Perspectives **273**
10.1 Summary . . . 273
10.2 Perspectives . . . 273
10.2.1 New measurements with Run-II data . . . 274
10.2.1.1 Cross section for $b\bar{b}$ production . . . 274
10.2.1.2 Inclusive c jet cross section . . . 274
10.2.2 Techniques of measurements . . . 274
10.2.3 Prospects at longer term . . . 275
References . . . 275

Appendices **275**

A Tracker Alignment **279**
A.1 Introduction . . . 279
A.1.1 Tracker alignment at CMS . . . 280
A.1.2 Track-based approach . . . 282
A.1.2.1 MillePede-II . . . 282
A.1.2.2 HipPy . . . 283
A.1.3 Samples . . . 283
A.1.4 Weak modes . . . 283
A.1.4.1 Types . . . 284
A.1.4.2 Fix with tracks of various topologies . . . 284
A.1.5 Time variations . . . 285
A.1.5.1 Types and reasons for variations . . . 285
A.1.5.2 Strategy . . . 285
A.2 2016 Legacy Alignment . . . 286
A.2.1 Configuration . . . 286
A.2.2 Performance . . . 286
A.2.2.1 Distribution of the medians of the residuals . . . 288
A.2.2.2 Primary-vertex validation . . . 291
A.2.2.3 $Z \to \mu\mu$ validation . . . 292
A.3 Summary & Conclusions . . . 294
References . . . 294

B TMD Plotter **297**
B.1 Introduction . . . 299
B.2 `TMDlib` . . . 299
B.3 `TMDplotter` . . . 299
B.4 Application . . . 300
B.5 Summary . . . 300
References . . . 300

C Latest jet measurements at CMS **303**

C.1 Introduction . . . 305
C.2 Inclusive jet analyses at 8 and 13 TeV . . . 305
C.3 Multijet analysis at 8 TeV . . . 306
C.4 Dijet triple differential cross section at 8 TeV . . . 306
C.5 Azimuthal correlations at 8 and 13 TeV . . . 307
C.6 Conclusions . . . 307
References . . . 307

List of Figures **309**

List of Tables **313**

Acronyms **315**

Index **321**

Acknowledgements **325**

Part I

Premisses

1

Introduction to High Energy Physics

The object of High Energy Physics is to study the constituents of matter and their interactions at quantic and relativistic scales, conditions only reachable with very high energies. At these scales, matter and interactions are both described in terms of *particles*.

In this chapter, we give a general introduction to High-Energy Physics (HEP): the context, history, achievements and challenges of this field of research.

1.1 Introduction

1.1.1 Units

Being both relativistic and quantic, the scope of HEP is reflected in terms of units [1].

1. Electric charges are counted in units of the *elementary charge*:

$$e = 1.602{,}176{,}6208(98) \times 10^{-19}\,\mathrm{C} \tag{1.1}$$

2. Terms of the dynamical equations are of order of the *reduced Planck constant*[1]:

$$\hbar = \frac{h}{2\pi} = 1.054{,}571{,}800(13) \times 10^{-34}\,\mathrm{Js} \tag{1.2}$$

3. Velocities are measured in units of the speed of light[2]:

$$c = 299{,}792{,}458\,\mathrm{m/s} \tag{1.3}$$

Units of energy are *electron-Volt*, or multiples:

$$1\,\mathrm{GeV} = 1.602{,}176{,}6208(98) \times 10^{-10}\,\mathrm{J} \tag{1.4}$$

1 eV corresponds to the kinetic energy that an electron gains when accelerated from rest in an electric potential of 1 V. Momentum and mass are respectively measured in GeV/c and GeV/c^2; in practice however, the speed of light c is rarely used. The $\hbar$ constant is

1. $h = 6.626\ 070\ 040(81) \times 10^{-34}\,\mathrm{Js}$
2. The symbol c stands for *celerity*, an outmoded synonymous of velocity.

always omitted, but if need be, it is straightforward to get it back into the formula using dimensional analysis.

« *High Energy* » refers to the relation that can be derived using dimensional analysis between an energy E to a wavelength λ:

$$E = 2\pi\frac{\hbar c}{\lambda} \quad \text{or more simply} \quad E = \frac{2\pi}{\lambda} \tag{1.5}$$

i.e. probing high (low) energy scales implies probing low (high) distance scales.

1.1.2 Fundamental interactions

Nowadays, nature is understood in terms of four *fundamental interactions*:

gravitation Gravitation describes interactions between objects due to their masses. Typical systems purely based on this force are solar systems, usually at large distance scales. Gravitation is by far the weakest interaction among the four; however, since masses are always positive, it is only cumulative and becomes therefore the dominant interactions from scales of $\mathcal{O}(1\,\mathrm{m})$.

electromagnetism Electromagnetism describes interactions between objects carrying an electric charge (so far, no magnetic charge has been observed). It is indeed the dominant force at scales from $\mathcal{O}(10^{-10}\,\mathrm{m})$ to $\mathcal{O}(10^{-3}\,\mathrm{m})$. Typical structures holding via the electromagnetic interaction are atoms and molecules; it also explains γ decay, as well as gaseous, liquid and solid states of matter.

strong nuclear interaction The strong (nuclear) interaction describes interactions of components and subcomponents of the atomic nucleus. "Nuclear" refers to the scales at which they take place, below $\mathcal{O}(10^{-14})$; the only macroscopic manifestation is α decay (emission of a nucleus of helium). In general, the study of the strong interaction requires very specific set-ups, as will be the case in this thesis.

weak nuclear interaction Finally, the weak (nuclear) interaction takes place only at small distance scales, similarly to the strong interaction. It is responsible for β decay (emission of an electron), which is crucial in the nucleosynthesis of stars, but no analogous system like planetary systems, molecules or nuclei may be found for holding only thanks to the weak interaction. This is related to the fact that similarly to the strong interaction, it takes only place at scales of the nucleus, but is much weaker than the strong interaction. However, unlike the strong interaction, it does not only affect the constituents of the nucleus but all particles of matter.

At the scale of experimentation of HEP, gravitation is too weak to produce any measurable effect; moreover, it is extremely difficult to formalise a quantum theory of gravitation. Nowadays, only the three other interactions are physically experimented in HEP and mathematically described within the so-called Standard Model (SM). The SM will be the object of the next chapter.

1.1.3 Particles

In general, any object that can be regarded as *pointlike* can be called a *particle*:

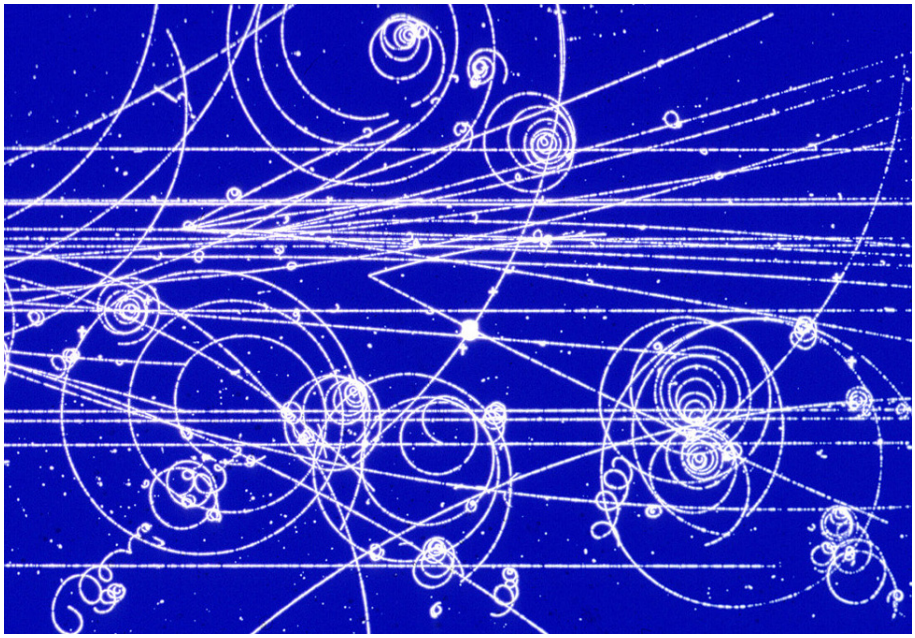


FIGURE 1.1 – "This image from 1960 is of real particle tracks formed in CERN's first liquid hydrogen bubble chamber to be used in experiments. It was a tiny detector by today's standards at only 32 cm in diameter. Negatively charged pions with an energy of 16 GeV enter from the left. One of them interacts with a proton in the liquid hydrogen and creates sprays of new particles, including a neutral particle (a lambda) that decays to produce the "V" of two charged particle tracks at the centre. Lower-energy charged particles produced in the interactions spiral in the magnetic field of the chamber. The invention of bubble chambers in 1952 revolutionized the field of particle physics, allowing real particle tracks to be seen and photographed, after releasing the pressure that had kept a liquid above its normal boiling point." [2]

— galaxies in the universe (cosmology),

— stars in a galaxy or planets in a solar system (astrophysics),

— molecules in a medium (statistical physics)

— atoms in a molecule (chemistry),

— nucleus in an atom or nucleons in a nucleus (nuclear physics),

— and partons in a proton (particle physics).

But properly said, Particle Physics concerns the *fundamental* and *composite* particles, *i.e.* the tiniest components and the sets made of these. It is in this sense that *particles* shall here be meant.

Essentially, these particles have two peculiar behaviours:

— Most of them are unstable and decay in a very short time (at most a small fraction of a second — see Fig. 1.1). Normal matter only consists of atoms made of stables particles: the protons, neutrons and electrons.

— When two particles collide *violently*, they may *produce* other particles.

The study of HEP consists in trying to understand these two behaviours. In particular, in this thesis, we are going to study an unstable particle, the b quark, which can be produced by colliding protons.

1.2 History

In this section, we recall some key steps in the history of the discovery of the fundamental constituents of matter, following both chronology and decreasing distance scale, as shown in Tab. 1.1.

energy scale	distance scale	object
keV	10^{-10} m	atom
MeV	10^{-13} m	nucleus
GeV	10^{-16} m	proton
TeV	10^{-19} m	parton

TABLE 1.1 – A few key figures relating the energy scale and the involved type of object. Note that 1 TeV represents approximately the kinetic energy carried by a domestic fly.

1.2.1 Atomism (10^{-10} m)

The modern theory of atomism started with the publication of the book *Les Atomes* [3] in 1913 by Jean PERRIN, where thirteen different, compatible measurements of the *Avogadro number* were presented:

$$N_A \approx 6 \cdot 10^{23}\,\mathrm{mol}^{-1} \tag{1.6}$$

This number is the typical number of atoms to be found in a few centimeters. Matter is not a *continuum* but is made of small corpuscles.

1.2.2 Nucleus (10^{-13} m)

The existence of the electron was admitted but no such oppositely charged particle was found. The atom was thought to be a diffuse, positive body in which the electrons, negative, would shelter (*Thomson model* or *plum-pudding model*). Between 1908 and 1913, the *golden-foil* experiment [4] by Sir Ernest RUTHERFORD, Hans GEIGER and Ernest MARSDEN highlighted the existence of a charged pointlike object — the *nucleus* — in atoms. The experiment (shown in Fig. 1.2) consists in bombing a golden foil gets with alpha rays, which can be found in naturally radioactive sources (like 238-uranium); according to the plum-pudding model, the radiation should have gone through the foil; however, they observed some alpha rays coming back. RUTHERFORD said: "It was almost as incredible as if you fired a 15-inch shell at a piece of tissue paper and it came back and hit you!" This was the first sign for the existence of a heavy, charged nucleus. The size of the nucleus was at most of the order of 10^{-13} m, since the energy of a natural source of alpha rays is around 5 MeV. This new model of the atom made of a small nucleus surrounded by an electronic cloud is called *Rutherford atom*.

1.2.3 Nucleus structure (10^{-15} m)

From the 1950s at SLAC in the U.S. and later in the 1960s at DESY, the nucleus was probed with particle beams with energy scale of the order of 100 MeV – 1 GeV, in order to probe smaller scales. Similarly to the golden-foil experiment, a nuclear target was bombed with a beam of electrons:

$$e + N \longrightarrow e + X \tag{1.7}$$

where e stands for *electron*, N for *nucleus* and X for some additional production. The nucleus itself was found to have a structure, make of pointlike *nucleons* (either protons or neutrons), arranged in a similar way to the structure of the electrons in the atom.

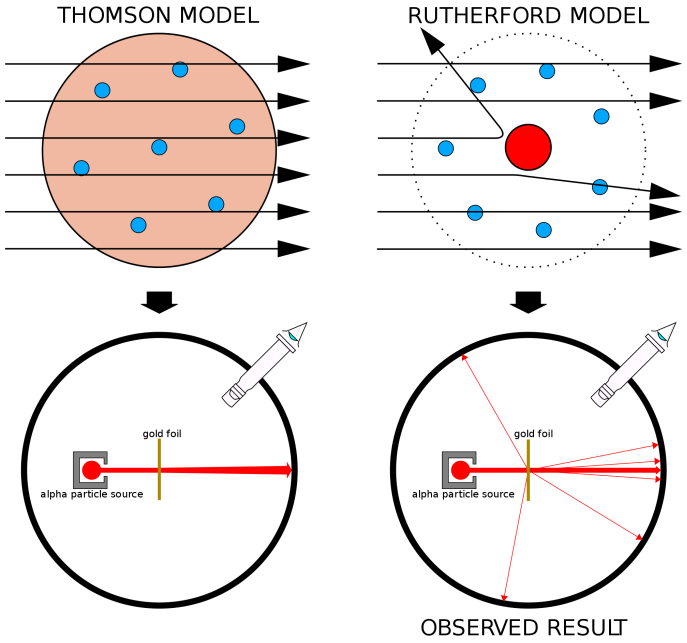


FIGURE 1.2 – "A simple diagram illustrating the Geiger-Marsden experiment. The left column shows the scattering pattern that the experimenters expected to see, given the plum pudding model of the atom. The right column shows the actual results, along with Rutherford's new planetary model." [5]

1.2.4 Nucleon substructure (10^{-18} m)

At a still lower distance scale, *i.e.* with higher-energy beams, protons and neutrons also were found not to be pointlike, but with a substructure surprisingly different to the atoms'. At that time, electrons were scattered on protons:

$$e + p \longrightarrow e + X \tag{1.8}$$

The results could be interpreted in two complementary ways:

1. the study the kinematics of the outgoing electron e led to the *parton model*, imagined by Richard FEYNMAN;
2. and independently the study of the symmetries of the hadronic production X led to the *quark* model, imagined by Murray GELL'MANN.

In the former, the *scattering* effects on the proton suggested pointlike subcomponents to exist, called *partons*, coherently moving without interacting with one another. In the latter, the existence of different subparticles, called *quarks*, could explain some *symmetries* among different types of particles. The two models were based on different observations and described the proton differently.

In the 1990s, at the HERA ep collider, at $E \sim 100\,\text{GeV} \rightarrow 10^{-18}\,\text{m}$, the H1 and ZEUS collaborations measured the content of the proton in terms of partons [7].

Since then, despite active searches, no new substructure has been found. However, six different *flavours* of quarks have been found. This thesis is dedicated to the study of one of them: the b quark.

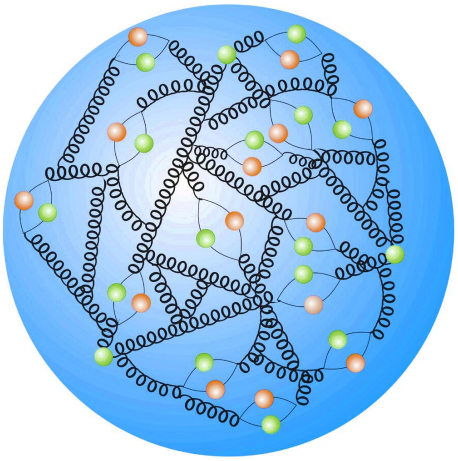


FIGURE 1.3 – On this artistic view, the proton is made of *quarks* (green), *antiquarks* (orange) and *gluons* (curly lines). There are three more quarks than antiquarks, called *valence quarks* (one can be found in the top, one in the right hand side, one in the bottom left). Quarks and antiquarks radiate gluons, themselves either radiating other gluons or decaying into a quark and an antiquark. [6]

1.3 Experimentation

We now discuss the experimental possibilities in HEP: first how to achieve the right conditions of experimentation, and what can exactly be measured.

1.3.1 Sources

Experimentation in HEP is limited in two aspects:

— Since most particles have a very short lifetime, one needs to find or set up sources of particles.

— Since particles have a very small size, very specific detectors needs to be set up.

One may distinguish three types of sources:

radioactive elements These can be found in nature or synthesised. This was how RUTHERFORD *et al.* first studied the atomic nucleus. Detection of particles are also performed.

cosmic rays Stars radiate particles, which scatter on molecules in the atmosphere. The *cosmic microwave background* also gives a picture of the universe when atoms were not yet bound together.

scattering experiments Accelerating particles and making them collide is another way to produce particles. In Part II, we are going to analyse data coming from a scattering experiment.

1.3.2 Observables

One distinguishes two observables: the *decay length* and the *cross section*. Any other fundamental parameter, *e.g.* the mass of the particles, is then extracted from the comparison of prediction and measurement.

Decay length. The *decay length* is the first observable that was measured in particle physics. All unstable particles, fundamental as well as composite, have a different lifetime.

Cross section. The second main measurable quantity in HEP is the *cross section*. Classically, the cross section is the overlapping area of the projections of the target in the transverse plane and the projectile particle. A first generalisation was performed when studying diffraction in optics, the cross section being then defined in terms of intensities rather than areas [8]. The concept was further extended in particle physics: particles having no clear borders, the cross section cannot be properly defined as a physical area; eventually, it is interpreted as a *rate* of scattering. Techniques to compute cross sections in HEP will be discussed in Chap. 2. In this thesis, we are going to measure a cross section.

1.4 Challenges

Despite the remarkable precision achieved in HEP experiments, many questions remained unanswered; for instance:

— gravitation is not described;

— the asymmetry observed between matter and antimatter in the universe is not explained;

— evidence for dark matter and dark energy abound in the universe (more in App. 1.A);

— the mathematical structure of the SM is unexplained, as well as its nineteen input parameters (see Chap. 2);

— calculations from the SM are not always analytically feasible, resulting in difficulties to produce predictions (see Chap. 2).

In this thesis

We present the measurement of the cross section of the inclusive b jet production in proton-proton collisions with the CMS experiment. The goal of this analysis is to test our knowledge at the *TeV scale*.

The first part is dedicated to present the context of the measurement. First, some elements of theory are given in Chap. 2 in order to discuss the current understanding of HEP and of proton-proton collisions; the notions of cross section and jet will be more rigorously detailed. This will be followed by a description of the CMS experiment, our experimental set-up, in Chap. 3. In Chap. 4, the Monte Carlo techniques, abundantly used for calculations in HEP, are discussed, and some models used in the second part are already discussed. A review of b physics closes the first part in Chap. 5.

The second part is dedicated to the measurement itself. First the strategy of the analysis will be described in Chap. 6. Then Chaps. 7-8 contain the analysis itself. The comparison of the measurement to predictions is presented in Chap. 9.

Finally, a third part is composed of prospects and of various appendices.

1.A Dark matter and dark energy

As we already stressed, gravitation is too weak to compete with other interactions, and is therefore not described in HEP. But the existence of dark matter and dark energy is a strong motivation for searches Beyond the SM (BSM).

1.A.1 Dark matter

Several observations suggest more matter to be in the universe than the radiated light may let it believe, *i.e.* some type of matter that does not interact electromagnetically and that cannot be found on earth; this unknown matter is called *dark matter*.

Historically, the main technique to detect dark matter has been to compute the difference between the luminous mass and the dynamical mass. This can be done at different scales:

— star clusters [9],

— galaxy clusters [10]

— and galaxy dynamics [11].

More recently, *gravitational lenses* even allowed to map dark matter in the universe [12]; in addition, anisotropies in the *Cosmic Microwave Background* may be partly explained by the presence of dark matter [13, 14]. This list is not exhaustive, but these observations are pointing to an important missing piece of modern HEP.

1.A.2 Dark energy

The *cosmological constant* is necessary to explain the observed expansion of the universe with the theory of general relativity [15]. Since it can be understood as a contribution to the energy, it is called *dark energy*. It would account for around two thirds of the content of the universe [16]; on the other hand, unlike baryonic and dark matter, it would fill the entire universe quite uniformly. Its nature is totally unknown; and its density is too low to be detected in experiments as of today.

References

[1] C. Patrignani et al. "Review of Particle Physics". In: *Chin. Phys.* C40.10 (2016), p. 100001. DOI: 10.1088/1674-1137/40/10/100001.

[2] "The decay of a lambda particle in the 32 cm hydrogen bubble chamber". 1960. URL: https://cds.cern.ch/record/39474.

[3] Jean Perrin. *Les Atomes*. Librairie Félix Alcan, 1930.

[4] Sir Ernest Rutherford F.R.S. "LVII. The structure of the atom". In: *Philosophical Magazine* 27.159 (1914), pp. 488–498. DOI: 10.1080/14786440308635117. eprint: http://dx.doi.org/10.1080/14786440308635117. URL: http://dx.doi.org/10.1080/14786440308635117.

[5] Kurzon. *Geiger-Marsden experiment expectation and result*. Accessed: 2017-09-10. URL: https://upload.wikimedia.org/wikipedia/commons/f/f9/Geiger-Marsden_experiment_expectation_and_result.svg.

[6] DESY. *A picture of the proton*. URL: http://www.desy.de.

[7] H. Abramowicz et al. "Combination of measurements of inclusive deep inelastic $e^{\pm}p$ scattering cross sections and QCD analysis of HERA data". In: *Eur. Phys. J.* C75.12 (2015), p. 580. DOI: 10.1140/epjc/s10052-015-3710-4. arXiv: 1506.06042 [hep-ex].

[8] Vincenzo Barone and Enrico Predazzi. *High-Energy Particle Diffraction*. Vol. v.565. Texts and Monographs in Physics. Berlin Heidelberg: Springer-Verlag, 2002. ISBN: 3540421076. URL: http://www-spires.fnal.gov/spires/find/books/www?cl=QC794.6.C6B37::2002.

[9] J. C. Kapteyn. "First Attempt at a Theory of the Arrangement and Motion of the Sidereal System". In: *Astrophys. J.* 55 (1922), pp. 302–328. DOI: 10.1086/142670.

[10] F. Zwicky. "Die Rotverschiebung von extragalaktischen Nebeln". In: *Helv. Phys. Acta* 6 (1933). [Gen. Rel. Grav.41,207(2009)], pp. 110–127. DOI: 10.1007/s10714-008-0707-4.

[11] Vera C. Rubin and W. Kent Ford Jr. "Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions". In: *Astrophys. J.* 159 (1970), pp. 379–403. DOI: 10.1086/150317.

[12] Priyamvada Natarajan et al. “Mapping substructure in the HST Frontier Fields cluster lenses and in cosmological simulations”. In: *Mon. Not. Roy. Astron. Soc.* 468.2 (2017), pp. 1962–1980. DOI: 10.1093/mnras/stw3385. arXiv: 1702.04348 [astro-ph.GA].

[13] P. A. R. Ade et al. “Planck 2015 results. XIII. Cosmological parameters”. In: *Astron. Astrophys.* 594 (2016), A13. DOI: 10.1051/0004-6361/201525830. arXiv: 1502.01589 [astro-ph.CO].

[14] Wayne Hu. *Intermediate Guide to the Acoustic Peaks and Polarization*. http://background.uchicago.edu/~whu/intermediate/intermediate.html. Accessed: 2017-10-21.

[15] Dragan Huterer and Michael S. Turner. “Prospects for probing the dark energy via supernova distance measurements”. In: *Phys. Rev.* D60 (1999), p. 081301. DOI: 10.1103/PhysRevD.60.081301. arXiv: astro-ph/9808133 [astro-ph].

[16] Joshua Frieman, Michael Turner, and Dragan Huterer. “Dark Energy and the Accelerating Universe”. In: *Ann. Rev. Astron. Astrophys.* 46 (2008), pp. 385–432. DOI: 10.1146/annurev.astro.46.060407.145243. arXiv: 0803.0982 [astro-ph].

2

Elements of theory

"Hypotheses non fingo."

— Isaac Newton [1]

In this chapter, some ideas of the theoretical aspects of modern HEP are presented, with a special emphasis on the topics underlying the measurement presented in Part II: first the Standard Model (SM), with the emphasis on Quantum Chromodynamics (QCD); then additional phenomenological models used in the treatment of proton-proton collisions are introduced, in particular the *evolution equations*.

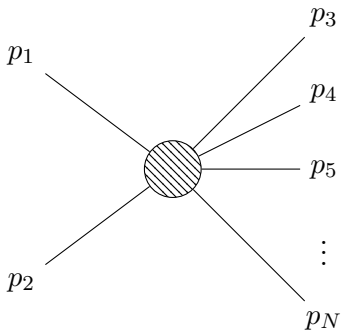
The approach given here does not correspond to the historical discoveries: as an alternative of the two first sections, the development of fundamental physics from the foundations to early days may be found in appendix of this chapter. The physics of the b quark will be the object of Chap. 5.

2.1 Introduction

The two main observables in HEP were already mentioned in Sect. 1.3.2: the *decay length* and the *cross section*. In both cases, phenomena with different numbers of particles in the initial and final states have to be accounted for; the *amplitude of transition* $\mathcal{M}$, or Matrix Element (ME), is written as follows [2]:

$$\mathcal{M} = \langle \text{final state} | \text{ interaction Hamiltonian } | \text{initial state} \rangle \tag{2.1}$$

The interaction Hamiltonian can be complicated — this will be discussed later in the chapter. From now on, the discussion is restricted to the computation of the cross section of any process $1+2 \to 3+4+\ldots+N$, which can be represented by the following diagram:



where the blob represents the interaction. The cross section can be deduced from the amplitude of transition by the following formula:

$$\sigma = \frac{S}{4\sqrt{(p_1 p_2)^2 - (m_1 m_2)^2}} \int_{\substack{\text{phase} \\ \text{space}}} \cdots \int \overline{|\mathcal{M}|^2} (2\pi)^4 \delta^4 \left(p_1 + p_2 - \sum_{j=3}^{N} p_j \right) \prod_{j=3}^{N} 2\pi \delta \left(p_j^2 - m_j^2 \right) \Theta(E_j) \frac{\mathrm{d}^4 p_j}{(2\pi)^4} \tag{2.2}$$

where

— the $p_i = (E_i, \mathbf{p}_i)$ stand for the four-momenta;

— the line over the squared ME stands for the summation (average) over the spins and over the colours in the initial (final) state if relevant,

— S is a combinatorial factor to avoid double counting when particles in the final state are identical;

— the integral runs over the phase space of all particles in the final state.

— the first Dirac delta function ensures that energy and momentum are conserved between initial and final states;

— the second Dirac delta function fixes the four-momenta of the outgoing particles to satisfy the mass condition, *i.e.* the outgoing particles have to be *on the mass shell*;

— the Heaviside function ensures that all particles in the final state have positive energy.

Among others, the SM provides tools to compute the ME: this will be the topic of Sect. 2.2. In some cases however, such as at Large Hadron Collider (LHC), the possibilities of calculation within the SM will be limited, and one resorts to phenomenological models: this will be the topic of Sect. 2.3.

2.2 The Standard Model of High Energy Physics

The SM consists of a whole, self-consistent, auto-sufficient body of theory that aims at accounting for all high-energy phenomena that are unambivalently founded theoretically and strongly attested by several experiments.

In this section, we draw a portrait of the SM [3, 4]. First, a(n attempt of) definition of the SM is given (Sect. 2.2.1). Then, its Lagrangian formulation is outlined (Sect. 2.2.2), the couplings of the interactions are discussed (Sect. 2.2.3), and computations techniques with Feynman diagrams are introduced (Sect. 2.2.4). At the end of the section, the current difficulties of computation techniques within the SM are discussed (Sect. 2.2.5).

2.2.1 Definition

The SM is a *renormalisable* (Sect. 2.2.1.3) *relativistic quantum field theory* (Sect. 2.2.1.1) based on *local Gauge invariance* (Sect. 2.2.1.2) [5] with *19 input parameters* (Sect. 2.2.1.4), describing the electromagnetic, weak and strong interactions (already introduced in Chap. 1).

2.2.1.1 Quantum field theory

As a consequence from the marriage of quantic and relativistic physics, the number of particles cannot be conserved; requiring a fixed number of particles in a relativistic system would violate causality [4]. Therefore the fundamental objects of the SM are not particles but *quantum fields* [6]. Particles — what is eventually seen in a detector — are considered as degrees of excitation of these fields.

The quantum fields existing in the SM are summarised in Tab. 2.1a; fields may be classified according to different properties:

spin	Fields in the SM can have spin 0 (scalar), $\frac{1}{2}$ (spinor) or 1 (vector). Spin-$\frac{1}{2}$ (spin-1) fields are usually considered as the fields of matter (interaction); they are listed in Tab. 2.1b (Tab. 2.1c). The BEH boson [1], or simply Higgs boson, is the only spinless field.
transformation under boost	Vector fields (spin-1), spinor fields (spin-$\frac{1}{2}$) and scalar fields (spin-0) have different transformations laws under relativistic transformations [7, 8].
mass	The mass plays also a rôle in the relation between the components of a spinor or vector field. For instance, neutrinos are massless and have only two components (Weyl spinor), while electrons are massive and have four components (Dirac spinor).

Fields have units of energy density.

2.2.1.2 Local Gauge invariance

Before describing it directly in the SM, it may be worthwhile to retrace the principle of Gauge invariance from earlier theories, in classical electrodynamics and in quantum mechanics.

In classical electrodynamics. Gauge invariance has already existed in classical electromagnetism; the electric and magnetic fields can be deduced from a scalar and a vectorial potentials ϕ and $\mathbf{A}$: $\mathbf{E} = -\nabla\phi - \frac{\partial \mathbf{A}}{\partial t}$, $\mathbf{B} = \nabla \times \mathbf{A}$. Any *Gauge transformation* Λ leads to the same evolution: $\phi \longrightarrow \phi + \frac{\partial \Lambda}{\partial t}$, $\mathbf{A} \longrightarrow \mathbf{A} + \nabla\Lambda$. In four-vectorial notation, the electromagnetic field is written $A = (\phi, \mathbf{A})$, hence the notation for the vector field for the photon. The Gauge invariance is a *local symmetry*, which means that the transformation is non-constant and can smoothly vary in space time.

In quantum mechanics. Gauge invariance also applies in non-relativistic quantum mechanics, and is equivalent to a redefinition of the phase [2] of the wave function [9]; in this case, one can observe that this symmetry corresponds to the group $U(1)$. Several properties may be derived consequently to the property of Gauge invariance, such as the charge conservation and the conservation of the amplitude of probability in quantum mechanics.

In the Standard Model. In the SM, the Gauge invariance is generalised to more complex groups like $SU(2)$ ($SU(3)$) for weak (strong) interactions [5, 8, 10]; all interactions may be derived as consequences of Gauge symmetries. The locality of interaction between

1. Named after the three physicists Brout, Englert and Higgs.
2. The word *Gauge* is indeed an old word for "phase".

spin	mass	type of field	d.o.f.	particles	diagram
0	yes	scalar	1	BEH scalar boson	
$\frac{1}{2}$	no	Weyl spinor	2	neutrinos	
	yes	Dirac spinor	4	charged leptons	
	yes	Dirac spinor	4	quarks	
1	no	massless vector	2	photon	
	yes	massive vector	3	weak bosons	
	no	massless vector	2	gluons	

(A) Fields in the Standard Model.

type	particles	electromagnetism	weak forces	strong forces
charged leptons	e, μ, τ	$Q = \pm 1$	yes	white
neutral leptons	ν_e, ν_μ, ν_τ	$Q = 0$		
up quarks	u, c, t	$Q = \frac{2}{3}$	yes	RGB
down quarks	d, d, b	$Q = -\frac{1}{3}$		

(B) Particles of matter are all half-spin particles. Particles have been ordered by increasing mass. Six *flavours* exists for the quarks. Except for the neutrinos that are neutral, every elementary particle of matter has a matching antiparticle that has the same properties but an opposed electric charge, contained in the same field.

Gauge boson	symbol	interaction	effective range
photon	γ	electromagnetism	infinite
neutral weak boson	Z^0	weak interaction	nuclear scale
charged weak bosons	$W^\pm$		
gluons	g	colour interaction	nuclear scale

(C) Particles of interaction are all 1-spin particles. The eight gluons being perfectly symmetric in the colour space, they cannot be distinguished from one another and are referred altogether only once.

TABLE 2.1 – In the SM, the fields are classified according to their symmetries: the spin, the mass and the charge.

two fields can be derived consequently to the locality of Gauge invariance; interactions at different places of space-time would indeed violate causality [4]. The vector fields, carrying the interaction in the SM, are thus said to be *Gauge fields*. Moreover, the Gauge symmetry is said to be internal, because it is a symmetry in the space of the charge (*e.g.* $U(1)$ for the electric charge, $SU(3)$ for the colour charge). In general, group theory (and Lie algebras) play a very important rôle in the description of symmetries of the SM [11, 5]; here, we only mention some properties. The cases of Quantum Electrodynamics (QED) and of Quantum Chromodynamics (QCD) are taken, the former for its simplicity, the second for its relevance in this thesis:

QED Electrically charged particles ψ interact by exchanges of photons A. The dynamics are invariant under the following transformation:

$$\text{electron} \quad \psi \to e^{ig\alpha}\psi \tag{2.3}$$

$$\text{photon} \quad A_\mu \to A_\mu - \frac{1}{g}\partial_\mu \alpha \tag{2.4}$$

The $U(1)$ Gauge invariance implies several important properties of QED: the conservation of the charge, the null mass of the photon, or the absence of

self-interactions of the photon. The electric charge can take two values: $\pm e$, exactly like in classical electromagnetism.

QCD Coloured particles ψ interact by exchanges of gluons A. One counts three colours (anti-colours) for quarks (anti-quarks) [3] and eight colours for gluons. The strong interaction relying on $SU(3)$, the Gauge transformation takes a more complicated form than in QED:

$$\text{quark} \quad \psi \to e^{ig_s \frac{\lambda^a}{2} \theta^a} \psi \tag{2.5}$$

$$\text{gluon} \quad A^a_\mu \to A^a_\mu - \frac{1}{g_s} \partial_\mu \theta^a + f^{abc} \frac{\lambda^b}{2} A^c_\mu \tag{2.6}$$

with the Gell'mann matrices λ^a ($a = 1, \ldots, 8$) and f^{abc} the structure constants of $SU(3)$. The existence of eight colour states for the gluon is directly related to the structure of $SU(3)$ [2]. Unlike QED, this additional term in QCD will be responsible for self-interactions of the gluon field.

The different symmetries corresponding to the charges of the three fundamental interactions in the SM are described by different *unitary groups*, with *group structure* $U(1) \times SU(2) \times SU(3)$. The reason for this structure is an open question in HEP.

2.2.1.3 Renormalisation

Most Quantum Field Theories (QFTs) do not lead to finite amplitudes of transitions. Only a limited number of interactions can be considered without rendering the theory ill-defined, with unresolvable *ultraviolet divergences*. Fortunately, there exist QFTs on which a procedure of *renormalisation* can be applied at the cost of

— introducing a (non-physical) *renormalisation scale*,

— renormalising the field,

— and redefining the couplings and the masses.

The SM is indeed renormalisable [12, 2, 13]. Some aspects related to renormalisation issues will be discussed later on.

2.2.1.4 Parameters

The SM requires 19 parameters, summarised in Tab. 2.2, which are not constrained by the theory and need to be determined experimentally:

— masses of the charged leptons,

— masses of the quarks,

— fundamental parameters,

— CKM matrix [4],

— extra parameters.

While the three charged leptons are massive, the neutrinos are massless in the SM. Experimentally though, it has been measured that the neutrinos are also massive, but since their masses are very small and since some doubts on their exact field properties still remain, their masses are not included in the SM yet.

3. The term of colour is therefore taken in analogy to the three primary colours and explain the *chromo* in Quantum Chromodynamics.

4. Named after the three physicists Cabibbo, Kobayashi and Maskawa.

	description	symbol	value
charged lepton masses	electron	m_e	511 keV
	muon	m_μ	105.7 MeV
	tauon	m_τ	1.78 GeV
quark masses	up	m_u	~ 2 MeV
	down	m_d	~ 4.5 MeV
	strange	m_s	~ 87 MeV
	charm	m_c	~ 1.3 GeV
	bottom	m_b	~ 4.2 GeV
	top	m_t	~ 173 GeV
fundamental constants	fine structure constant	α	$\frac{1}{128.957 \pm 0.020}$
	Weinberg angle	$\sin^2 \theta_W$	0.23116 ± 0.00012
	strong coupling	α_S	0.1184 ± 0.0007
CKM matrix	12-mixing angle	θ_{12}	13.1°
	23-mixing angle	θ_{23}	2.4°
	13-mixing angle	θ_{31}	0.2°
	CP violation phase	δ	0.995
extra parameters	QCD vacuum angle	Θ_{QCD}	~ 0
	vacuum expectation value	v	246 GeV
	Higgs mass	m_H	125.09 ± 0.24 GeV

TABLE 2.2 – The nineteen parameters of the SM have been measured or constrained experimentally. The mass of the leptons, the mixing angles, the CP violation phase and the extra parameters are absolute, whereas the mass of the quarks and the fundamental parameters, which can be obtained from the Gauge couplings, depend on the renormalisation scheme (especially, the fundamental parameters are here given at the scale $Q^2 = M_Z^2$). The values from [15].

The mass of the quarks is affected by the renormalisation. The present values are given approximately to give an order, where the scale is close to the mass. Heavy Flavour (HF) stands for quarks with mass above 1 GeV, *i.e.* charm, bottom and top.

The CKM matrix is involved in the mixing of the quark flavours, describing the weak decays of massive quarks into lighter quarks of different charge. The phase is involved in the *(weak) CP violation* [14] (not discussed here).

The fundamental parameters describe the intrinsic strength of an interaction, in contrast to the (electric, weak, or strong) charges that describe the behaviour of the particle according to this interaction. The fundamental parameters can be directly related to the Gauge couplings g, which appear explicitly in the Gauge transformation. Their values also depend on the renormalisation scale; in particular, the evolution of their values will be discussed in more detail in Sect. 2.2.3.

The Θ_{QCD} parameter is related to the *strong CP violation*, unobserved but predicted by the SM (not discussed here).

Finally, the *vacuum expectation value* and the *Higgs mass* are related to Higgs physics (not discussed in this thesis).

2.2.2 Lagrangian

We describe the mathematical expression of the SM to compute predictions.

The commonest form in which the SM is expressed consists in a Lagrangian density $\mathcal{L}_{\text{SM}}$. In principle, one can apply the Principle of Least Action to deduce the Equations of Motion (EOM):

$$S = \int \mathcal{L}\left(\phi, \partial_\mu \phi\right) \, \mathrm{d}^4 x \quad \longrightarrow \quad \text{EOM} \equiv \delta S = 0 \tag{2.7}$$

The amplitude of transition can then be determined from the action ($\mathcal{M} \sim \exp(iS)$). In practice however, because of the difficulties to solve them, the EOMs are rarely directly used in analyses such as the one presented in this thesis. But an interpretation may be read directly from each term of the Lagrangian, and generic rules how to compute cross sections have been invented by Richard Feynman; before going to this topic (Sect. 2.2.4), the Lagrangian is further described.

The Lagrangian contains a large number of terms (varying according to the representation), and can be divided in certain sectors:

$$\mathcal{L}_{\text{SM}} = \mathcal{L}_{\text{EW}} + \mathcal{L}_{\text{QCD}} + \mathcal{L}_{\text{Higgs}} + \mathcal{L}_{\text{Yukawa}} \tag{2.8}$$

2.2.2.1 Electroweak sector

The electromagnetism and weak interactions are entangled in the same sector, known as Electroweak (EW) [3]. The involved fermions are *leptons* (either charged leptons or neutrinos) and *quarks*, interacting electromagnetically by exchanges of photons or weakly by exchanges of $W^\pm$ or Z^0 bosons. The photon has infinite range and is massless, while the weak bosons exist only at the scale of the nucleus and are massive. Each field possesses excitation states corresponding to particles and to antiparticles, according to the electric charge. In addition, fermions exist in three generations, as shown in Tab. 2.1b:

charged leptons	electron (e), muon (μ), tauon (τ)
neutral leptons	partner neutrinos (ν_e, ν_μ, ν_τ)
up-type quarks	up (u), charm (c), top (t)
down-type quarks	down (d), strange (s), bottom (b)

The discussion is now restricted to electromagnetism, *i.e.* to QED, leaving out the weak interaction. The commonest representation of the Lagrangian of QED reads as follows:

$$\mathcal{L}_{\text{QED}} = \sum_{\text{fermions}} \bar{\psi}_f \left(i \not{D} - m\right) \psi_f - \frac{1}{4} F_{\mu\nu} F^{\mu\nu} \tag{2.9}$$

where the electromagnetic field (*i.e.* the photon) is hidden both in the *covariant derivative* D (the slash only indicating how the components of the spinor are combined) and in the *electromagnetic tensor* F:

$$D_\mu = \partial_\mu - iA_\mu \tag{2.10}$$
$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \tag{2.11}$$

After reorganising the terms, one can distinguish three types of terms:

mass terms	order two in the fields (*e.g.* $-\frac{1}{2} A_\mu A^\mu$)
interaction terms	order three or four in the fields (*e.g.* $e^2 \bar{\psi} \gamma^\mu \psi A_\mu$)
kinematic terms	terms involving derivatives (*e.g.* $i \bar{\psi} \gamma^\mu \partial_\mu \psi$)

2.2.2.2 Quantum Chromodynamics sector

The Quantum Chromodynamics describes the strong nuclear interactions. Quarks and antiquarks interact by exchanges of gluons [16]. The $SU(3)$ symmetry implies that quark (antiquarks) exist in three colour (anti-colour) states, and gluons in eight; the symmetry is however perfect and the states cannot be distinguished. The range of this interaction is that of the nuclear scale. QCD is not directly sensitive to the flavour; the Lagrangian of QCD is invariant with respect to the flavour, since the different quarks have different masses, the *flavour democracy*, or *flavour blindness*, only takes place at energies where all the quark masses are negligible.

The Lagrangian of QCD is similar to the Lagrangian [5] of QED (Eq. 2.9):

$$\mathcal{L}_{\text{QCD}} = \sum_{\text{quarks}} \bar{\psi}_q^i \left(i \not{D}^{ij} - m\delta^{ij} \right) \psi_q^j - \frac{1}{4} G_{\mu\nu}^a G^{a\mu\nu} \tag{2.12}$$

where the strong field (*i.e.* the gluon) is hidden both in D and G:

$$D_\mu^{ij} = \delta_{ij}\partial_\mu - i g_s \frac{\lambda_a^{ij}}{2} B_\mu^a \tag{2.13}$$

$$G_{\mu\nu}^a = \partial_\mu B_\nu^a - \partial_\nu B_\mu^a + g_s f^{abc} B_\mu^b B_\nu^c \tag{2.14}$$

In comparison with the Lagrangian of QED in Eq. 2.9, an additional Latin index runs on the colour ($i, j = 1, 2, 3$ for quarks, $a, b, c = 1, \ldots, 8$ for gluons). Similar terms are found, with additional interaction terms for the gluon field with itself.

2.2.2.3 Yukawa and Higgs sectors

The Higgs boson is the most recently discovered particle in the SM [17, 18]. It is involved in a complex procedure of symmetry breaking of the electroweak sector, allowing the presence of mass terms in the Lagrangian [19, 20]. It allows to recover mass terms for the fermions (Yukawa sector) and for the weak bosons (Higgs sector) without violating the Gauge symmetry in the EW sector [6]. The Higgs sector also describes the self-interactions of the Higgs boson when interacting with itself.

2.2.3 Running couplings

The QFT of an interaction is characterised by a coupling α, directly related to the constant g in the Gauge transformation (*e.g.* in Eqs. 2.3 and 2.5). Gauge couplings have already been briefly mentioned while discussing the nineteen input parameters of the SM (Sect. 2.2.1.4). For instance, the electromagnetic coupling is the *fine structure constant*.

Gauge couplings are crucial for they determine the perturbative or non-perturbative nature of the energy scale. Different Gauge structures are related to different behaviours of the respective couplings; especially, while the QED coupling *decreases* from small to large distances, the QCD coupling *increases*. This property is related to the self-interactions of the gluon field.

In QED and QCD, this coupling is said to be a *running coupling*, since it depends on the energy scale Q. In a first approximation, QED and QCD are scale invariant, which

5. The term related to Θ_{QCD} is here neglected, since there is no experimental evidence for it.
6. Unlike the electroweak sector, the QCD sector involves vectors; mass terms are therefore not forbidden by the Gauge symmetry in QCD.

means that the interaction will be the same at all scale. The *scale violation* is only logarithmic [21], but when different orders of magnitudes are involved, the variations of the couplings cannot be neglected.

In addition, an important difference between the couplings of QED and QCD have qualitatively opposite behaviours:

QED The coupling α, hence the effect of the electric charge, decreases with larger and larger distances, the effect of the electric charge becoming smaller and smaller. However, at very high energies, *i.e.* at very small distances, the coupling should become closer to unity; the validity of the perturbative regime is expected to break down. While in atomic physics, the fine structure constant [7] is close to 1/137, in HEP at LHC, at around 100 GeV, it is close to 1/127; the non-perturbative regime is still far from the accessible phase space.

QCD The behaviour is opposite: the coupling α_S increases at small energies, *i.e* at large distances; this is called the *asymptotic freedom*. At small energies, the QCD is in the *non-perturbative regime*. In QCD, only the ME at high energy can potentially be calculated; hadrons can rarely be described with perturbative equations. One can write the strong running coupling as follows [22] [8]:

$$\alpha_S(Q^2) \propto \frac{1}{\ln\left(\frac{Q^2}{\Lambda^2_{\text{QCD}}}\right)} \tag{2.15}$$

where $\Lambda_{\text{QCD}} \sim 200\,\text{MeV}$, typically the scale of hadronic masses; the transition between the perturbative and non-perturbative regimes is given by the Λ_{QCD} constant (below Λ_{QCD}, Eq. 2.15 has no validity). The variation, or *running*, of α_S is significant, as shown in Fig. 2.1.

The measured couplings have already been given in Tab. 2.2; they correspond to *renormalised* couplings. By convention, the *couplings*' values are given at the peak of mass of the Z^0 boson $Q^2 = M_Z^2 \approx (91.2\,\text{GeV})^2$, *i.e.* well in the perturbative regime. While α_S is the direct analog to α for the strong force, θ_W, the *Weinberg angle*, is of different nature, as it is involved in the mixing of electromagnetism and weak force within the electroweak theory. A recent review of the latest measurements [9] of the strong couplings may be found in [23].

2.2.4 Calculations with Feynman diagrams in the perturbative regime

Richard Feynman invented a graphical representation of the terms appearing in the Lagrangian to treat calculations in the perturbative regime, called *Feynman diagrams* [24].

These diagrams can be seen as representations of the dynamics, how the fields interact, and allow to understand the calculations. Although not used explicitly to perform calculations in this thesis, it is worthwhile giving some first principles in order to discuss theoretical predictions.

7. At low energy, it is given by $\alpha = \frac{e^2}{4\pi\epsilon_0\hbar c}$ and measured with a very high precision to $7.297,352,5698(24) \times 10^{-3}$ [15].

8. This corresponds to the renormalisation scheme called *minimal subtraction*.

9. In addition, on behalf of the CMS collaboration, I also had the opportunity to present the latest measurements at CMS at the LHCP 2017 conference; the proceedings may be found in App. C.

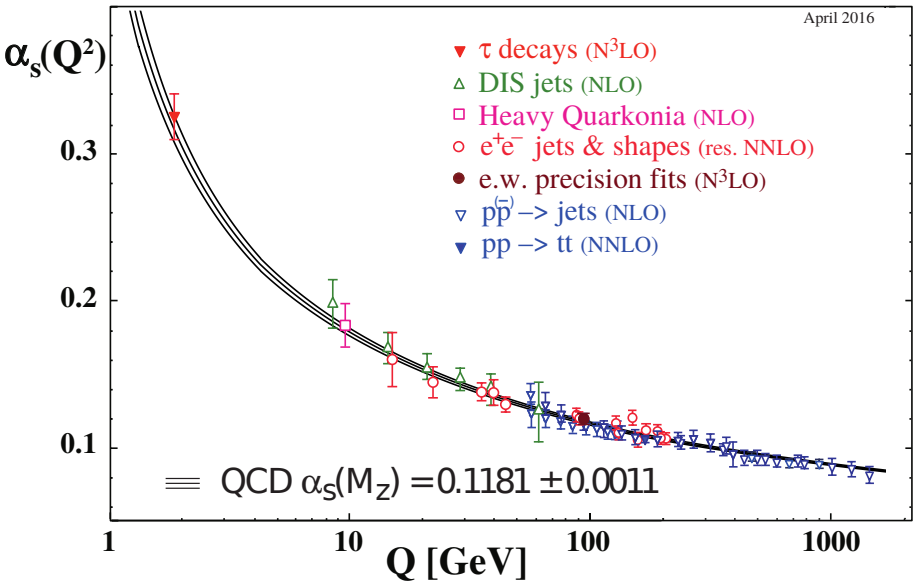
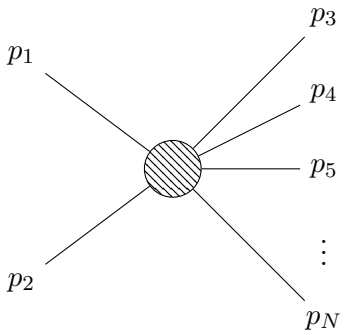


FIGURE 2.1 – "Summary of measurements of α_S as a function of the energy scale Q. The respective degree of QCD perturbation theory used in the extraction of α_S is indicated in brackets (…)." [15]

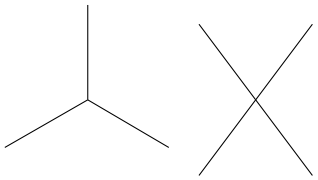
A Feynman diagram connects the particles in the initial (on the left hand side) and final states (on the right hand side):



The time (space) coordinate is represented horizontally (vertically). What happens in the blob is represented by *lines* and *vertices*:

lines The lines correspond to the type of field (see Tab. 2.1a). Each one corresponds to a factor in the calculation, called *propagator*.

vertex Interactions are represented by joint lines:



The four-momentum has to be conserved at every vertex.

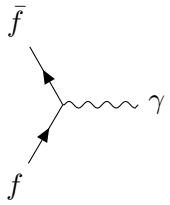
All particles inside (outside) of the blob are said to be *virtual* (*real*). A real process being characterised only by its initial and final states, as described in Eq. 2.1; virtual particles

cannot be observed. A virtual particle may even have a different mass than the mass of the real particle.

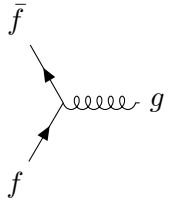
In principle, the interaction Hamiltonian will be made of *all* the possible combinations of vertices respecting the laws of conservations between the initial and final states. The exact *Feynman rules*, translating the diagrams into terms entering the computation of the ME, may be found for instance in Ref. [4]; here, we only give the principles.

From the Lagrangian, one can deduce all possible diagrams in QED and QCD:

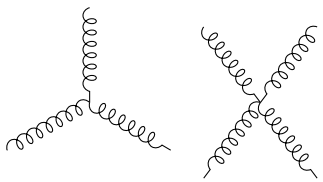
QED Only one type of vertex exists: the interaction of a fermion with a photon:



QCD The analog diagram to the QED one exists:



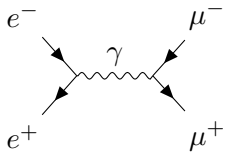
In addition, since the gluon can interact with itself, 3- and 4-leg vertices are possible:



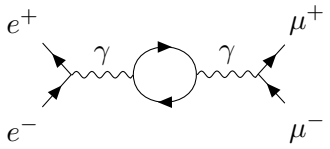
Real processes however always involve two scattering particles in the initial state and at least two product particles in the final state: 3-leg vertices do not correspond to any real process and must be combined at least in pairs; 4-leg vertices may however take place without being combined.

2.2.4.1 Example

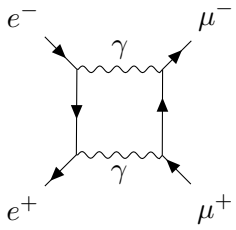
In principle, in order to compute the ME of a given process (see Eq. 2.1), one should consider *all* possible combinations of the interactions. To give a preview of the complexity, one can take the example of $e^+e^- \to \mu^+\mu^-$ at an energy scale $\mathcal{O}(1\,\text{GeV})$, where the situation is as simple as possible since only QED is involved with distinct initial and final states.

$i\mathcal{M}_{\mu^+\mu^-}$ = 

(tree-level) (2.16)

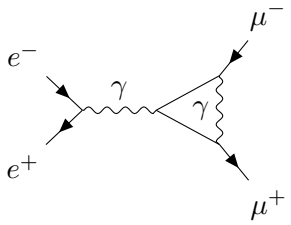
\+ 

(1-loop diagrams) (2.17)

\+ 

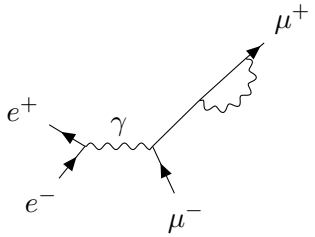
(2.18)

\+ one similar diagram (2.19)

\+ 

(2.20)

\+ one similar diagram (2.21)

\+ 

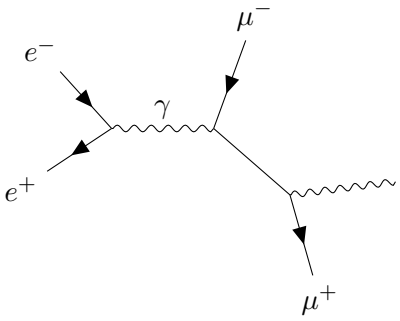
(2.22)

\+ three similar diagrams (2.23)

\+ ... (higher loop multiplicity) (2.24)

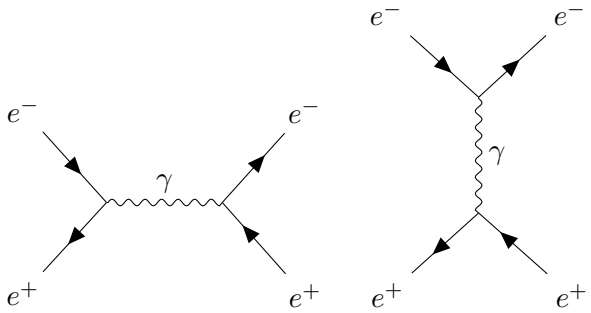
Including more complicated diagrams means computing higher-order terms in the calculation, but they are rarely considered. The given diagram is one of the simplest ones; in general, complications arise very soon:

— A more realistic view should be to consider additional radiations:



Such a radiation could take place before the annihilation (Initial-State Radiation, or ISR), or after the decay of the virtual photon (Final-State Radiation, or FSR). These additional radiations are important: *soft* (*i.e.* at low energy) and *collinear* (*i.e.* collimated with the emitter) radiations cause the ME calculation to diverge – this is called *infrared divergences*. It turns out that these may be absorbed by the computation of *loop diagrams* (such as Eq. 2.22).

— When describing $e^+e^- \to e^+e^-$, two diagrams (instead of only one) should be considered already at tree level:



From left to right, these two diagrams are respectively said to be in the s-channel and t-channel (see also App. 2.B.2). The multiplicity of diagrams can become extremely high as soon as one includes loop diagrams.

— At higher energies, one should also start considering diagrams involving weak bosons $W^\pm$ and Z^0, implying additional loop diagrams. Moreover, QCD loops should also be considered, adding yet other diagrams.

As a conclusion of this example, the representation in Feynman diagrams allows to see how even the most elementary processes may require advanced calculations.

2.2.4.2 Vocabulary

We introduce now some common terms employed to characterise calculations. The context is fixed to QCD.

Inclusive and exclusive cross sections. Given a certain process $ij \to F$, the *inclusive* (*exclusive*) cross section corresponds to the final state including (excluding) all extra radiations. The inclusive cross section can be seen as a sum of exclusive cross sections with X all extra radiations in the final state:

$$\hat{\sigma}^{\text{incl}}_{ij\to F} = \sum_k \hat{\sigma}^{\text{excl}}_{ij\to F+k} \tag{2.25}$$

where k stands for the number of additional legs to the original diagram.

Leading Order and Next-to-Leading Order. The inclusive cross section of $ij \to F$ can be further decomposed in MEs, the diagrams with identical final state correspond to loop diagrams:

$$\hat{\sigma}_F^{\text{incl}} = \sum_k \int \mathrm{d}\Phi_{F+k} \sum_l \left| \mathcal{M}_{F+r}^{(l)} \right|^2 \tag{2.26}$$

where l stands for the number of loops in the diagram ($\mathrm{d}\Phi$ is a common short notation of the whole element of phase space). Then, the terms may be rearranged in orders of the strong coupling:

$$\hat{\sigma}_F^{\text{incl}} = \sum_n \hat{\sigma}_F^{N^n LO} \tag{2.27}$$

where n stands for the order in the strong coupling. Each leg (loop) contributes with one (two) order(s) in α_S; then the squared modulus of the ME implies the presence of interference terms. One commonly defines:

$n = 0$ Leading Order

$n = 1$ Next to Leading Order

$n = 2$ Next to Next to Leading Order

(above, one simply writes N^{n}LO, but higher orders are rarely considered). The number of terms contributing to a cross section at the first orders is shown in Tab. 2.3. The *Born cross section* designates LO for $2 \to 2$ processes, *i.e.* in the most elementary final state. Finally, the *tree level* corresponds to diagrams without loops, regardless of the number of legs.

Virtuality. The conservation of the momentum only applies to the legs of a diagram; internal lines do not respect this constraint. Let p be the four-momentum of an internal line, p^2 is its *virtuality*. If $p^2 \approx m^2$ ($p^2 \neq m^2$) where m is the mass of the corresponding particle, then the particle is said to be *on-shell* (*off-shell*). In diagrams, off-shell particles are sometimes noted with a star (*e.g.* γ^* stands for a virtual photon).

additional terms	first factor		second factor	
	# loops	# legs	# loops	# legs
LO	0	0	0	0
NLO	1	0	0	0
	0	1	0	1
NNLO	2	0	0	0
	1	0	1	0
	1	1	0	1
	0	2	0	2
etc.	⋮	⋮	⋮	⋮

TABLE 2.3 – Number of loops and legs contributing to the different orders in the strong coupling.

2.2.5 Current limits of calculations within the SM

The SM is an very powerful theory and has been verified to an extreme precision; the nineteen input parameters have been measured (or at least constrained) with very good precision.

However, it still suffers from three major difficulties that we encounter in proton-proton collisions: the treatment of non-perturbative QCD, the complexity of the perturbative calculations, and the high multiplicity of the final states. This will motivate the phenomenological approach presented in Sect. 2.3 of this chapter.

2.2.5.1 Non-perturbative QCD

In QCD, low-energy systems are characterised by the property of *confinement*: quarks and gluons never appear alone and free, but always in bound states, *i.e.* in *hadrons* [10] [26, 27]. The study of the scattering of partons requires necessarily the scattering of hadrons; it is therefore essential to understand and to be able to describe these objects [11].

2.2.5.2 Complexity of perturbative calculations

The calculations of cross sections (Eq. 2.2) are most of the time too complex to be solved completely analytically, as soon as more than a few particles are present in the final state with several orders of accuracy in the perturbative expansion. For instance, the number of diagrams at tree level for $e^+e^- \to u\bar{u}X$ (where X stands for additional gluon radiations) is shown in Tab. 2.4. Most of them cannot be integrated analytically and require Monte Carlo (MC) techniques (introduced in Chap. 4); the different theoretical predictions will differ on the order of the calculation, and on the phenomenological treatment of the additional radiations (described in Sect. 2.3).

final state	# diagrams
$u\bar{u}g$	2
$u\bar{u}gg$	8
$u\bar{u}ggg$	50
$u\bar{u}gggg$	428
$u\bar{u}ggggg$	4670

TABLE 2.4 – Number of diagrams in e^+e^- collisions for final states with a pair $u\bar{u}$ and extra gluon radiations at the tree level [30].

2.2.5.3 High-multiplicity final states

Finally, in order to be able to apply perturbative QCD (pQCD), one needs to *deconfine* the partons inside of the hadron, and let them reach the regime of asymptotic freedom. This can only be achieved at energies $Q^2 \gg \Lambda^2_{\text{QCD}}$, which also means $Q^2 \gg m^2_{\text{proton}}$; in LHC conditions, the phase space is large enough to allow the production of hundreds of additional particles. Today, the production of such a high number of particles cannot be described directly within the SM and requires a phenomenological approach.

10. One should mention an extremely dense state of matter where partons are not confined in hadrons but deconfined in a very strong colour field: the *quark-gluon plasma* [25]. This kind of state is studied by the ALICE experiment at LHC.

11. Although we are presently not concerned with it, one should at least mention another, non-phenomenological approach of non-perturbative QCD: *lattice QCD* [28, 29]. In this approach, QCD systems are discretised and treated numerically with very large computation resources. Good progresses are being done, but lattice QCD is currently not used in the present context.

2.3 Phenomenology

According to the type of collisions under study different phenomenological approaches exist. For *elastic scattering* (both protons remain) and *diffractive dissociation* (at least one proton gets destroyed), the commonest approach to treat such collisions is the Regge theory, where the interaction is described in terms of exchanges of a *pomeron*[12] [33, 34].

In this thesis, we shall rather consider inelastic non-diffractive scattering at LHC, where large exchanges of momentum take place. Indeed, in this context, the property of *factorisation* can be applied in order to separate the scattering into a contribution from the perturbative regime and another contribution from the non-perturbative regime (Sect. 2.3.1). Starting from this property, we describe the phenomenology of proton-proton scattering, first with hadrons and jets (Sect. 2.3.2), then with an overview of a proton-proton scattering (Sect. 2.3.3). Finally, we close the chapter with a discussion of the different types of factorisations (Sect. 2.3.4).

2.3.1 Factorisation

An important property allows to separate the hadronic cross section into contributions from the interaction at parton level and from the hadron; this property is known as *factorisation*, and is expected to apply for the following processes [35]:

— in Deeply Inelastic Scattering (DIS):

$$l + h \to l' + X \tag{2.28}$$

where h (l, X) stands for any hadron (lepton, anything);

— in electron-positron scattering:

$$e^+ + e^- \to h + X \tag{2.29}$$

— the Drell-Yan (DY) process:

$$h + h' \to \mu^+ + \mu^- + X \tag{2.30}$$
$$h + h' \to e^+ + e^- + X \tag{2.31}$$
$$h + h' \to W + X \tag{2.32}$$
$$h + h' \to Z + X \tag{2.33}$$

— jet production:

$$h + h' \to j + X \tag{2.34}$$

where j stands for jet (see Sect. 2.3.2.2);

— heavy quark production:

$$h + h' \to Q + X \tag{2.35}$$

where Q stands for heavy quark.

12. As part of my contributions to the CMS collaboration, I also had the chance to participate to two measurements of the minimum-bias hadronic production in proton-proton collisions, where the different contributions of the total cross section are measured in detail [31, 32].

Theoretically, it has been proven only for the tree first processes for *leading twist*[13]. Experimentally however, the factorisation is believed to hold in presence of a *hard* process, *i.e.* at scale $\mu_F^2 \gg \Lambda^2_{\text{QCD}}$, corresponding the regime of asymptotic freedom; but the fact that the same Parton Distribution Functions (PDFs) work for different types of scatterings (typically *ep* and *pp* collisions) is a very strong argument in favour of the legitimacy of the factorisation.

Technically, a *factorisation scale* μ_F^2 must be chosen: below this scale, the contributions to the hadronic scattering is contained in the PDFs; above this scale, the process is described by the partonic scattering (described in Sect. 2.2). Since this scale μ_F^2 is not physical, a PDF does not represent any fundamental quantity: it is only a way of separating (or factoring out, hence the name) the non-perturbative regime from the whole process; however, this separation also is not exact, and part of the perturbative regime is also *de facto* included in the PDF. As of today, PDFs cannot be computed and must be extracted (or fitted) from measurements. Different PDF extractions, or *sets*, are available, according to the data used to perform the extraction or to the type of fit; for instance (alphabetically ordered, non-exhaustive list): ABM [36], CTEQ [37], HERAPDF [38], MMHT [39], MSTW [40] or NNPDF [41, 42].

Different schemes exist: we shall here consider the *collinear factorisation* with DGLAP[14] evolution. In this paradigm, PDFs are noted $f_{i/h}(x, \mu_F^2)$, where x_i corresponds to the *momentum fraction* of the hadron momentum carried by the parton i. (Other evolution and factorisation schemes will be discussed later in Sect. 2.3.4.)

The factorisation was initially introduced in the context of DIS:

$$\sigma^{\text{incl}}_{lh\to F} = \underbrace{\sum_{i\in g,q,\bar{q}} \int_0^1 \mathrm{d}x_i f_{i\in h}(x_i, \mu_F^2)}_{\text{extraction from PDF}} \times \underbrace{\hat{\sigma}^{\text{incl}}_{li\to F}(x_i, \mu_F^2)}_{\text{parton-level process}} \tag{2.36}$$

PDFs are universal in the sense that they should be common to all processes to which factorisation applies; however, different choices of PDFs apply according to the order of the ME. As an illustration, in Fig. 2.2 are shown the extractions at NLO and NNLO performed at the HERA collider (already introduced in Sect. 1.2.4) with the combined data from the H1 and ZEUS experiments [43].

Since they are universal, the same PDFs can in principle be used in the factorisation for hadron-hadron scattering as follows:

$$\sigma^{\text{incl}}_{h_1h_2\to F} = \sum_{i\in g,q,\bar{q}} \int_0^1 \mathrm{d}x_i f_{i\in h_1}(x_i, \mu_F^2) \times \tag{2.37}$$

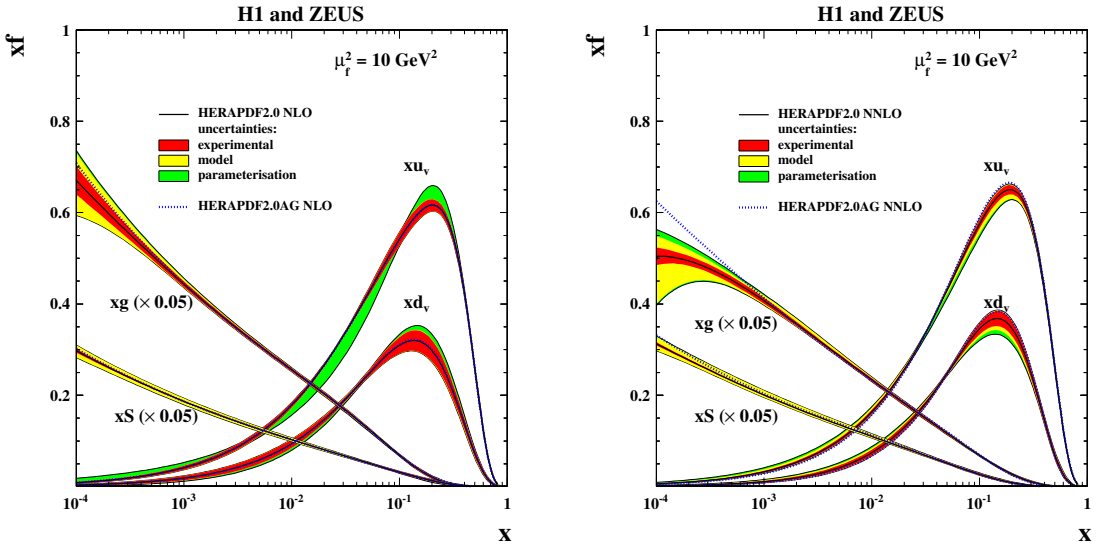
$$\times \sum_{j\in g,q,\bar{q}} \int_0^1 \mathrm{d}x_j f_{j\in h_1}(x_j, \mu_F^2) \times \tag{2.38}$$

$$\times \hat{\sigma}^{\text{incl}}_{ij\to F}(x_i, x_j, \mu_F^2) \tag{2.39}$$

The hadron-hadron cross section may also be seen, at a given factorisation scale μ_F, as the sum on the gluons and the flavours of quarks of the corresponding parton-parton cross sections weighted by the PDFs.

13. The twist corresponds to the mass dimension minus the spin; the cross section can be expanded in orders of the twist; in the factorisation, correction terms in $\ln\left(Q^2/\Lambda^2_{\text{QCD}}\right)^{m<2n}/Q^{2n}$ are neglected.

14. Named after the five physicists DOKSHITZER, GRIBOV, LIPATOV, ALTARELLI and PARISI.



(A) NLO (B) NNLO

FIGURE 2.2 – Extraction of the PDFs of the proton at the HERA collider in *ep* scattering [43]. The four curves correspond to up and down valence quarks, to gluons and to sea quarks. The PDFs are extracted in different sets according to the order. Both cases are qualitatively similar: the up and down valence quarks show a bump at high x, while the sea quarks and the gluons become dominant at low x.

Here, we shall focus on proton-proton interactions in the context of LHC, such as the ones that will be analysed in Part II. The application of factorisation will be described later on.

2.3.2 Objects in the initial and final states

Before detailing the interaction of the proton-proton cross section, it is necessary to define what can be found in the initial and final states of such a scattering. As already mentioned in Sect. 2.2.5.1, partons are never found free but always clustered in hadrons. Hadrons themselves are usually produced in jets, corresponding qualitatively to the production coming from a parton in the final state of a strong interaction.

2.3.2.1 Hadrons

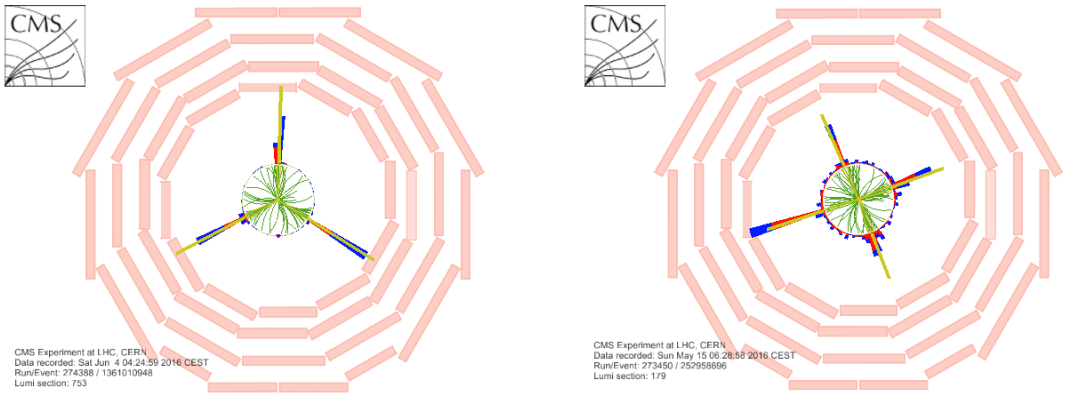
Hadrons are composite, colour-neutral, bound states holding as a result of confinement property of the strong interaction.

Two kinds of hadrons exist, according to their *valence quarks*: just as the valence electrons that determine the properties of chemical species, the valence quarks determine the properties of the hadrons[15]:

baryons which are triplets of quarks or antiquarks,

mesons which are pairs of quark-antiquark.

15. *Hadron* means *strong* in Old Greek, in relation to the strong force. *Baryon* (*mesons*) means *heavy* (*middle*). However, *leptons*, for *light*, are no hadrons. The etymology is purely based on empirical point of view.



(A) Tri-jet event (B) Four-jet event

FIGURE 2.3 – These two event displays show the final states of proton-proton collisions at CMS seen in the transverse plane [45]. In the innermost part, the green lines represent the reconstructed tracks corresponding to charged particles. The red and blue piles correspond to the energy deposits from all particles except muons and neutrinos; they are the signature of jets.

Hadrons are also characterised by their lifetimes. The proton is the only stable hadron [16], which is the reason for which, in practice, collisions are mostly performed with protons or atom nuclei [17]. However, many different types of hadrons can be directly seen in a detector; a list of hadrons with respective decay lengths is given in Tab. 2.5.

In addition to the valence quarks, other partons may exist in a hadron, namely the *sea* quarks and the gluons; their existence correspond to QCD fluctuations, and are permitted within the *Uncertainty Relation* $\Delta E \Delta t \geq \hbar/2$. The content of hadrons is described in terms of the aforementioned PDFs, and changes as a function of the energy scale at which it is considered. In Fig. 2.2, the contribution from valence and sea quarks is separated; since there is (are) only one down (two up) valence quark(s), the following sum rules apply:

$$\int_0^1 f_{u/p}(x)\,\mathrm{d}x = 2 \quad \text{and} \quad \int_0^1 f_{d/p}(x)\,\mathrm{d}x = 1 \tag{2.40}$$

In addition, the momentum sum rule must apply at any scale:

$$\int_0^1 \mathrm{d}x \sum_{i\in\{q,g\}} x f_{i/p}(x, Q^2) = 1 \tag{2.41}$$

2.3.2.2 Jets

Hadrons themselves are often found collimated in *jets*; this can be best illustrated by the two *event displays* in Fig. 2.3, where the final states of proton-proton collisions at CMS are shown; jets may be seen as the experimental pendants to the partons in the final state of a strong interaction. If jets can also be defined at parton level or at hadron level, there is no unique way of clustering hadrons in jets: clustering is partly arbitrary. To define a *jet clustering algorithm*, one take several important properties into account:

16. The neutron may be stable only if it is bound into a nucleus; this having been said, its lifetime being around fifteen minutes, it may also be considered as stable in the decay products of a collision.
17. One should still mention the pion-pion and proton-pion scatterings [44].

— The boundaries of a jet are in principle undefined.

— An algorithm has to be infrared and collinear safe (IRC), *i.e.* safe toward the emission of a soft gluon or of a collinear gluon.

— Small-size jets may miss important contributions from the showering of a hard parton, but large-size jets will be contaminated by extra activity in the event.

Different algorithms have been developed: the cone and the recombination algorithms [46, 47]. The difficulty in defining jets is that they need to be convenient both in predictions and in measurements [48].

Cone. These algorithms are defined in the (y, ϕ) space with rigid boundaries, where y (ϕ) stands for the rapidity [18] (azimuthal angle). Cone algorithms present the advantage of being easy to implement, but suffer from the difficulty of dealing with overlapping jets. Today, the most cited cone algorithm is the SIScone (Seedless and Infrared Safe Cone) algorithm [19] [49]; it is represented in the top left of Fig. 2.4. It presents certain advantages, as being IRC. However, the SIScone algorithm still relies on the choice of a non-physical parameter f to deal with overlapping cones. In addition, in comparison with the recombination algorithms (defined in the next paragraph), the SIScone is very time- and resource-consuming.

Recombination algorithms. All the recombination algorithms have the advantage of being IRC safe, and may be seen a particular case of the following algorithm:

— Define the distances d_{ij} between any two particles i and j of transverse momentum $k_{\perp i,j}$:

$$d_{ij} = \min\left(k_{\perp i}^{2p}, k_{\perp j}^{2p}\right) \frac{\Delta y_{ij}^2 + \Delta\phi_{ij}^2}{R^2} \tag{2.42}$$

where R is the cone size radius parameter, and for p the exponent parameter.

— Define the distances d_{iB} between any particle i and the beam B:

$$d_{iB} = k_{\perp i}^{2p} \tag{2.43}$$

— Then one proceeds iteratively:

1. Find first the minimum of the entire set of distances d_{ij}, d_{iB}.
2. If d_{ij} is smaller, than cluster i and j into a (proto) jet by summing their momenta; if d_{iB} is smaller, then label i as a jet.
3. If all particles have not been assigned to a jet, redefine the entire set of distances with the new objects and return to 2; else continue.

These algorithms are obviously invariant under boosts along the z-axis, along which the incident protons scatter. One should mention the three most frequently used ones:

k_T The k_T algorithm [50] (bottom left) is obtained for $p = 1$. This algorithm assume that the particles inside of a jet should have similar momenta. Soft particles are first clustered, explaining the irregular shape; this shape is very sensitive to the surrounding activity (the Underlying Event and the pile-up).

18. See App. 2.B for the description of the coordinates.

19. One can also mention the Iterative Cone algorithm with Progressive Removal procedure (collinear unsafe), or the Iterative Cone algorithm with Split Merge procedure (infrared unsafe). These two algorithms suffer from the fact that they rely on the somewhat arbitrary choice of a seed, *i.e.* a particle that would define the direction of the cone. The seeds lead to several ill-defined behaviours; the SIScone algorithm does not need any seed.

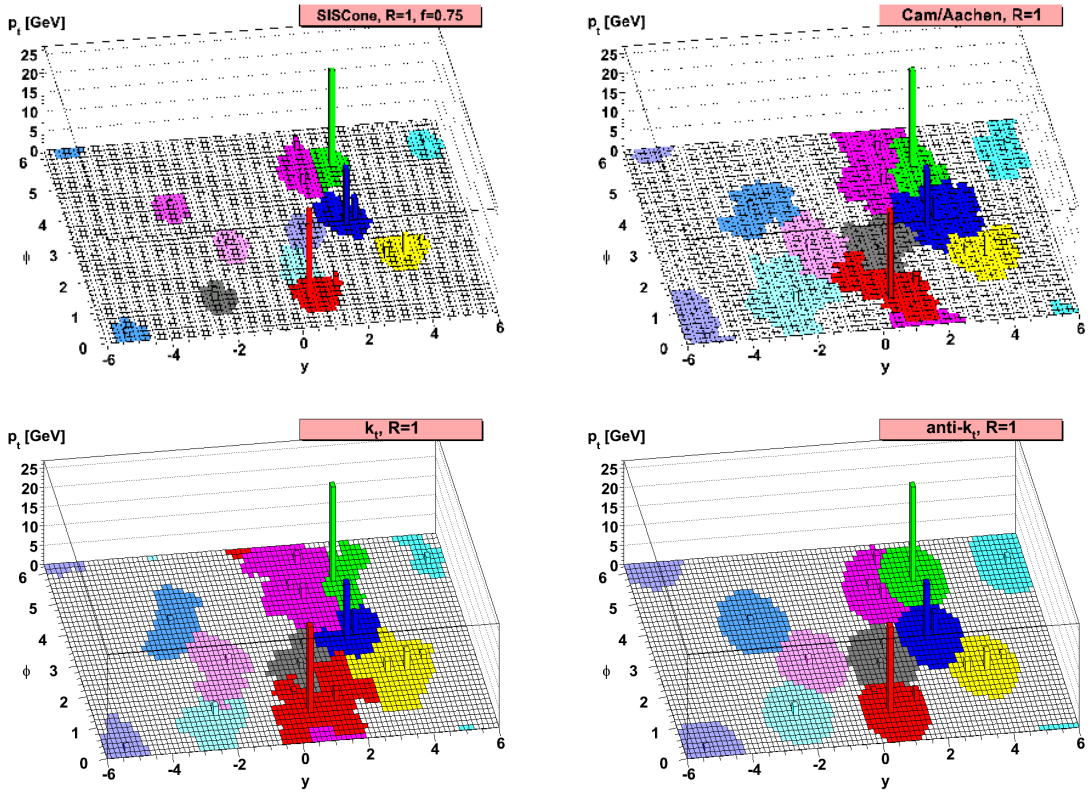


FIGURE 2.4 – Illustration of the four jet clustering algorithms with $R = 1$ for the same event from a simulation. Taken from [52].

Cambridge-Aachen The Cambridge-Aachen algorithm [51] (top right) is obtained for $p = 0$. The jet is also very sensitive to the surrounding activity but the substructure of the jet is totally conserved.

anti-k_T The anti-k_T algorithm [52] (bottom right) is obtained for $p = -1$. This algorithm is similar to the k_T algorithm but clusters first hard particles, which is related to the regular shape; this shape is robust against the surrounding activity.

As an illustration of the difference between the clustering algorithms, Fig. 2.4 shows the result of four different algorithms of jet clustering for the same event. At LHC Run-II, unless one is interested in the jet substructure, the standard algorithm used at CMS and ATLAS is the anti-k_T algorithm with cone size radius $R = 0.4$ or $R = 0.7$, in order to ease the comparison of results.

In this thesis, only one jet algorithm is considered: the standard anti-k_T algorithm shall be used with $R = 0.4$.

2.3.3 Overview of a proton-proton scattering at LHC

The overview will follow the sketch in Fig. 2.5; the goal is to go through the different components:

1. Matrix Element (ME),
2. Parton Shower (PS),
   — Initial-State Radiation (ISR),

   — Final-State Radiation (FSR);
3. Multi-Parton Interaction (MPI) and Beam-Beam Remnants (BBR);
4. hadronisation;
5. stable particles,
   — hadron decays,
   — soft photon radiation.

This description corresponds to the current understanding, as implemented in the General-Purpose MC Event Generators, whose implementation will be further discussed in Chap. 4.

2.3.3.1 Hard process

The starting point of the overview of the proton-proton scattering is the factorisation (Eq. 2.37).

First, the hard process (represented with a red blob in the figure) can be described in the perturbative regime, as introduced in Sect. 2.2. As of today, predictions can be performed typically at LO or NLO, and in some cases at higher orders.

The hard process defines the *hard scale* Q^2 of the process. At LO, a natural choice corresponds to take the virtuality of the internal propagator; at higher orders, there is no obvious way to proceed; in general, the choice of the hard scale is matter of debate.

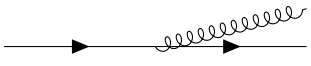
The hard scale is then used as factorisation scale, *i.e.* $\mu_F^2 = Q^2$, at which the PDFs have to be considered. In order to reach the right scales, the PDFs of the incident hadron (three green lines with a green blob) can be evolved as a function of the scale. This is achieved with the DGLAP equations:

$$\frac{\mathrm{d}f_a(x,\mu_F^2)}{\mathrm{d}\ln\mu_F^2} = \sum_{b\in\{q,g\}} \int_x^1 \frac{\mathrm{d}z}{z} P_{ba}(z) f_b\left(\frac{x}{z},\mu_F^2\right) \tag{2.44}$$

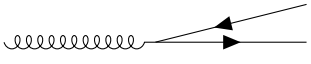
The (collinear) splitting functions P_{ba}s describe the transition (after emitting one or several partons) of a parton b into a parton a carrying a momentum fraction z of the initial parton. Their exact expressions depend on the order of precision in which the evolution is performed.

At LO, they can be deduced by comparing the $2 \to 2$ and $2 \to 3$ MEs [22, 54] and, for massless quarks, correspond to the following expressions:

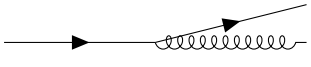
$$P_{qq}^{\mathrm{LO}}(z) = \frac{4}{3}\frac{1+z^2}{1-z} \tag{2.45}$$



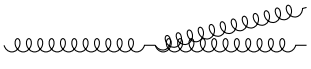
$$P_{gq}^{\mathrm{LO}}(z) = \frac{3}{2}\left(z^2+(1-z)^2\right) \tag{2.46}$$



$$P_{qg}^{\mathrm{LO}}(z) = \frac{4}{3}\frac{1+(1-z)^2}{z} \tag{2.47}$$



$$P_{gg}^{\mathrm{LO}}(z) = 3\left(\frac{z}{1-z}+\frac{1-z}{z}+z(1-z)\right) \tag{2.48}$$



In order to deal with the divergences in the evolution, a more rigorous expression of the splitting functions should include some regularisation [53]; since it is not useful for the present discussion, the regularisation is neglected for clarity. Moreover, additional splitting

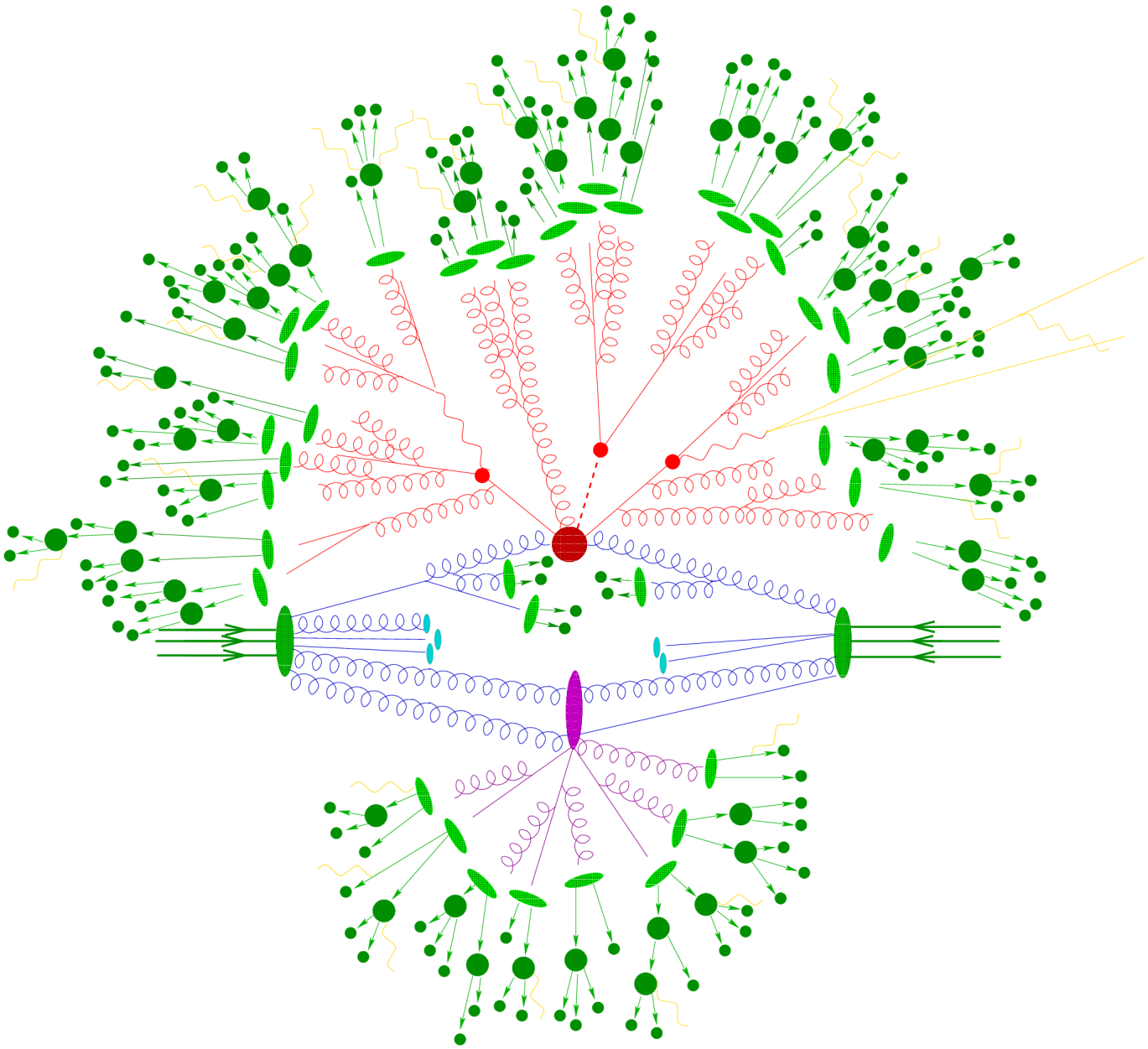


FIGURE 2.5 – "Sketch of a hadron-hadron collision as simulated by a Monte Carlo event generator. The red blob in the center represents the hard collision, surrounded by a tree-like structure representing Bremsstrahlung as simulated by parton showers. The purple blob indicates a secondary hard scattering event. Parton-to-hadron transitions are represented by light green blobs, dark green blobs indicate hadron decays, while yellow lines signal soft photon radiation." The incident partons, after extraction from the PDFs, are in blue. Finally, multi-parton interaction are shown in cyan. [53]

functions also exist to include the photons and leptons in the evolution; again for clarity, we restrict the discussion to QCD.

The interpretation of DGLAP equations (Eq. 2.44) is the following: when evolving to a higher scale, the PDF must account for with additional splittings, *i.e.* finer fluctuations can be resolved. This having been said, it is important to stress that the evolution can only be applied in the perturbative regime; a PDF cannot be evolved lower than the hadronisation scale. In other words, the DGLAP equations only describe the evolution of the perturbative component of the PDF.

The interpretation in terms of fluctuations explains how non-u and non-d quarks may appear. However, it may not always be correct to neglect the masses of HF quarks, since these belong to scales of the perturbative regime. Today, different *flavour schemes* exist, according to whether they are treated as massive or as massless:

1. In the *5-flavour scheme*, the b quark is treated as light and c quarks, *i.e.* it comes from PDFs; b quarks may be produced already at LO, but will be massless as any other quarks (except top).
2. In the *4-flavour scheme*, $b\bar{b}$ pairs can only be produced explicitely in the ME; in that case, calculations including higher-order terms are required, but on the other hand the mass of b's can be correctly described.
3. Similarly, one can also define a *3-flavour scheme*.

Unlike all other quarks, the top quark is always considered as massive.

2.3.3.2 Parton Shower

When applying scrupulously the factorisation and using a fixed-order ME, the extra ISRs are not described explicitly, but only accounted for in the PDFs; however, radiations in the perturbative regime (namely the *shower*) should be *resolvable* in the detector (red an blue lines). Moreover, additional FSRs are also expected to take place, which apply to legs outgoing either from the hard process of from ISR legs.

Fig. 2.6a illustrates the PS in a $gg \to gg$ scattering, showing a realistic scenario of the expected multiplicity, while the NLO or even NNLO calculations can only treat up a few legs. Therefore, in order to obtain a description of the high multiplicity in the final state, one resorts to the evolution equations again [53, 22, 55, 56]. The formulation of the DGLAP equation in Eq. 2.44 is *inclusive* in the sense it only allows to change the scale of the PDF without describing explicitly *when* a branching occurs; for this, it is necessary to rewrite it in an *exclusive* form.

The evolution can be performed iteratively from the hard scale of the hard process down to the hadronisation scale $Q_0 \sim \Lambda_{\text{QCD}}$. The treatment of ISR and FSR is similar, despite some differences:

— First, the ISR (FSR) is space-like (time-like).

— Then, the ISR is evolved *backward* from the ME back to initial hadron (though only for efficiency reasons), while the FSR evolves without such constraint since the hadronic final state is *a priori* not known; in fact, in the context of FSR, the DGLAP evolution involves Fragmentation Functions (FFs) instead of PDFs.

Final-State Radiation. The FSR is the simpler and therefore discussed first. In order to determine when a splitting takes place, the DGLAP equation is reformulated in terms

of probability for a parton a to undergo a splitting and become a parton b:

$$\mathrm{d}^2\mathbb{P}_a(z,Q^2) = \frac{\mathrm{d}f_a(z,Q^2)}{f_a(z,Q^2)} \tag{2.49}$$

$$= \mathrm{d}\ln Q^2 \frac{\alpha_S}{2\pi} P_{ab}(z)\,\mathrm{d}z \tag{2.50}$$

The probability of branching at a given scale regardless of the momentum fraction is described by the integral of Eq. 2.50 over all resolvable values of $z_{\min} < z < z_{\max}$ (which excludes *de facto* the divergences):

$$I_{ab}(Q^2) = \int_{z_{\min}}^{z_{\max}} \mathrm{d}z \frac{\alpha_S}{2\pi} P_{ab}(z) \tag{2.51}$$

For an infinitesimal evolution $\delta\ln Q^2$, the probability for branching is given as follows:

$$\mathbb{P}_a(Q^2, Q^2+\delta Q^2) = \sum_{b\in\{q,g\}} I_{ab}(Q^2)\delta\ln Q^2 \tag{2.52}$$

The probability for no branching is given by the complementary probability. Then, for an evolution from Q_1^2 to Q_2^2, the probability for no branching is given by a product of probabilities of not branching on infinitesimal intervals; eventually, it is given by an exponential and is called *Sudakov (form) factor*:

$$\Delta_a^{\mathrm{FSR}}(Q_1^2,Q_2^2) \equiv \bar{\mathbb{P}}_a(Q_1^2,Q_2^2) = \exp\left(-\int_{Q_1^2}^{Q_2^2} \mathrm{d}\ln Q^2 \sum_{b\in\{q,g\}} I_{ab}(Q^2)\right) \tag{2.53}$$

where $\bar{\mathbb{P}}_a$ stands for the probability for a *not* to branch. It can be interpreted in analogy with the decay of a particle, described with a Poisson process, $P(t) = \exp(-Nt)$ by identifying the (logarithm of the) scale $\ln Q^2$ with the time and the probability of branching with the decay rate N, with the only difference that in the case of the PS, the decay rate depends on the time. Moreover, a branching can take place on any outgoing leg of a diagram (unless this leg has already reached the hadronisation scale); the solution for the dependence in "time", or more exactly in the scale, is to proceed iteratively with a Sudakov factor associated to all outgoing legs:

1. Start from the scale of the hard process.
2. The Sudakov factors determine the scale of next branching (at a lower scale).
3. The branching of highest scale defines the new scale; one leg is added to the diagram.
4. If the scale has not reached the hadronisation scale yet, the procedure is iterated from Item 2; otherwise, stop the evolution.

The implementation is further described in Chap. 4. At each branching, the conservation laws are applied (four-momentum and quantum numbers in general).

Initial-State Radiation. The treatment of ISR is similar, with the difference that since the evolution is performed backward, the PDF has to be taken into account explicitly in the probability of branching:

$$\frac{\mathrm{d}f_b(x,Q^2)}{\mathrm{d}f_b(x,Q^2)} = \mathrm{d}\ln Q^2 \frac{\alpha_S}{2\pi} \int_x^{z_{\max}} \frac{\mathrm{d}z}{z} \frac{f_a(x/z,Q^2)}{f_b(x,Q^2)} P_{ab}(z) \tag{2.54}$$

This results in a Sudakov factor weighted by the PDF, therefore also keeping a dependence in the momentum fraction carried by the parton:

$$\Delta_b^{\mathrm{ISR}}(x, Q_1^2, Q_2^2) = \exp\left(- \int_{Q_1^2}^{Q_2^2} \mathrm{d}\ln Q^2 \int_x^{z_{\max}} \frac{\mathrm{d}z}{z} \frac{f_a(x/z, Q^2)}{f_b(x, Q^2)} \right) \tag{2.55}$$

Given this difference, the procedure is the same.

Strong ordering. The addition of a leg to the cross section formally corresponds to the following factorisation of the cross section:

$$\mathrm{d}\sigma_{F+k+1} = \mathrm{d}\sigma_{F+k} \Delta_a(Q_k^2, Q_{k+1}^2) \frac{\alpha_S}{2\pi} P_{ab}(z) \, \mathrm{d}z \, \mathrm{d}\ln Q^2 \tag{2.56}$$

In principle, adding a leg to a diagram gives rise to an interference term; it can be shown that this interference can be neglected in the context of *strong ordering* of the scale:

$$Q^2_{\text{hadronisation}} \ll \cdots \ll Q^2_{F+k+1} \ll Q^2_{F+k} \ll \ldots \ll Q^2_{\text{hard process}} \tag{2.57}$$

Moreover, several possibilities exist for the scale, and may differ in the ISR and FSR [56].

— The most natural choice consists in identifying the scale with the virtuality p^2 of the particles [57].

— An equivalent choice is to consider the *angular ordering*, since $\mathrm{d}Q^2/Q^2 = \mathrm{d}\theta/\theta$ [55].

— Finally, another possibility is to use transverse-momentum ordering [56].

In general, any scale such that $Q^2 = f(z)p^2$ with f a "reasonable" function can hold [56].

Intrinsic k_T. In order to reflect the Fermi motion inside of the proton, an *intrinsic* k_T, or *primordial* k_T, is given to the initial partons, such that $k_T \sim \mathcal{O}(100\,\mathrm{MeV})$. Conceptually, this is related to the fact that the DGLAP evolution only applies to the perturbative regime and treats the partons in a purely collinear way. In practice, a Gaussian is used to describe the primordial k_T. Then, the successive branchings make the partons acquire larger and larger transverse momenta.

Parameters. The PS comes with a few parameters, typically

— the strong coupling α_S,

— the hadronisation scale Q_0,

— the maximum momentum fraction $z_{\max}$,

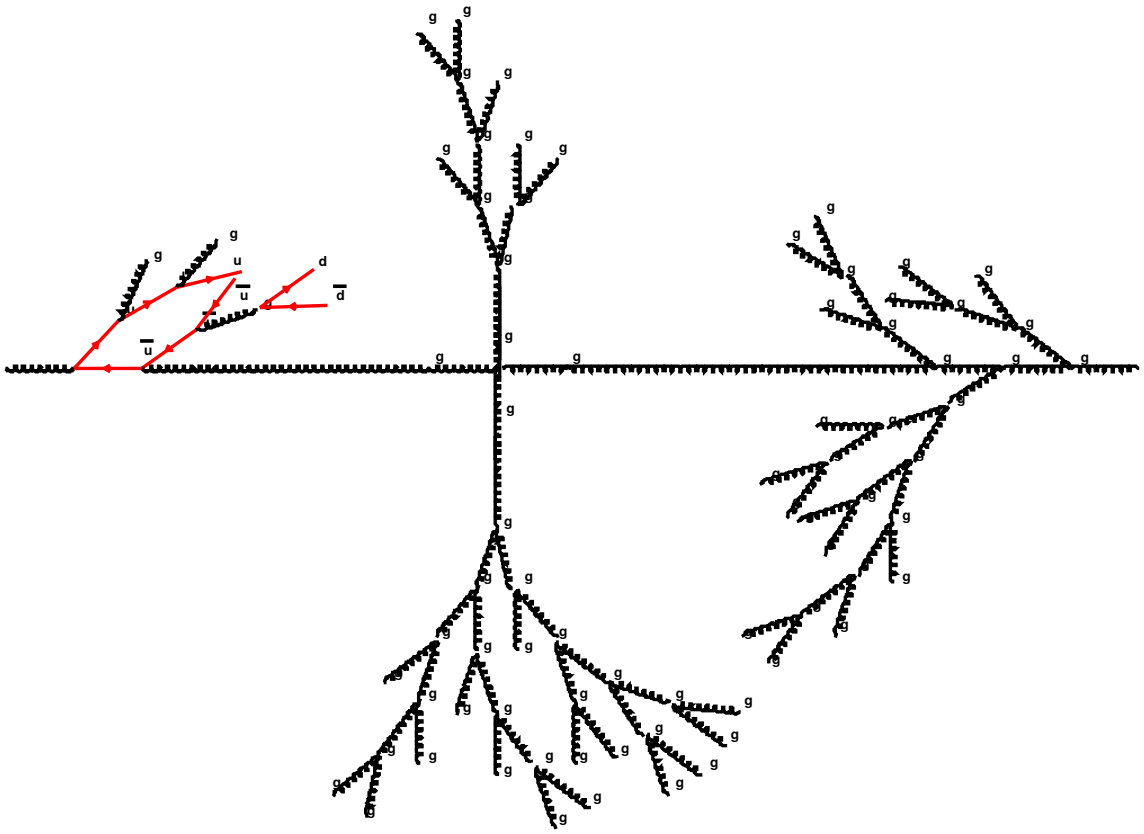
— and the width σ of the Gaussian for the intrinsic k_T.

2.3.3.3 Multi-Parton Interactions and Beam-Beam Remnants

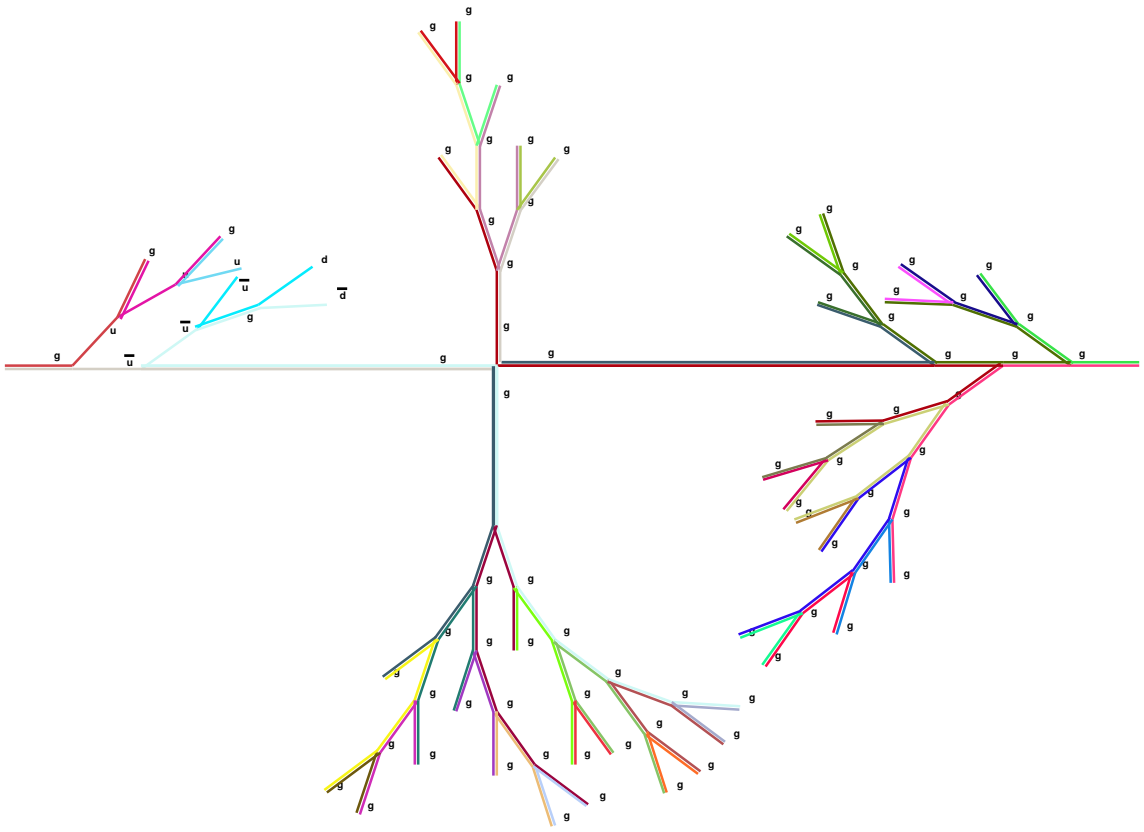
In the strict application of the factorisation, the underlying colour interactions between the incident protons (purple blob and lines) is neglected. In this picture, the brute calculation of the cross section for $2 \to 2$ QCD processes is divergent:

$$\frac{\mathrm{d}\hat{\sigma}}{\mathrm{d}p_T} \propto \frac{\alpha_S(p_T^2)}{p_T^2} \tag{2.58}$$

Historically, the possibility of MPI was suggested by the divergent behaviour of the calculation for $p_T \sim 3 - 5\,\mathrm{GeV}$. In fact, additional hard interactions at similar scales may even take place; in that case, the formalism of Double-Parton Scattering (DPS) can be applied [59].



(A) Parton Shower.



(B) Colour flow.

FIGURE 2.6 – Sketch of PS with colour flow [58]. The two incident (product) hard gluons are represented horizontally (vertically); the other lines corresponds to the result of showering. Above, the lines respect the representation in Feynman diagrams; below, the colour flow in the large-N limit is depicted for the same event.

Smoothing factor. In order to cope with the divergent behaviour at low transverse momentum, the PS of a branching is then either interrupted below a certain value p_{T0} [55] or tamed with a smoothing factor in the cross section in Eq. 2.58 [56]:

$$F(p_T) = \left(\frac{\alpha_S(p_{T0}^2 + p_T^2)}{\alpha_S(p_T^2)} \frac{p_T^2}{p_{T0}^2 + p_T^2} \right)^2 \tag{2.59}$$

In both cases, the parameter p_{T0} has to be fitted from data [20]

Number of interactions and interleaved Parton Shower. However, if Eq. 2.59 solves the divergence at low transverse momentum, it does not reproduce the multiplicity in the final state of the event. Several MEs are used to describe the different hard scatterings, but share the same Underlying Event (UE):

— The average number of interactions can be estimated from the ratio of the hard and non-diffractive cross sections [57]:

$$\langle n \rangle = \frac{\sigma_{\text{hard}}}{\sigma_{\text{non-diff}}} \tag{2.61}$$

The number of interactions is follows a Poisson distribution.

— Since the different hard processes come from the same protons, the PS of the different hard processes must be *interleaved* [56]. The same procedure is applied, but all outgoing legs are taken into account simultaneously to perform the shower. This is illustrated with p_T-ordering with four interactions in Fig. 2.7. Actually, two hard scatterings may come from a common ancestor, as it is the case for the interaction 2 and 3 on the figure; this case is called *joined interactions* [56].

Colour reconnection. Another aspect of the MPI is the fact that the different interactions must share the same colour flow, which has an impact on the hadronisation process (described in the next subsection). A procedure to reconnect the branches of the different interactions can be applied [57]. A basic model of the colour reconnection simply consists in reconnecting partons randomly; but this is insufficient for two reasons: first gluons seem to be ordered so as to minimise colour interaction, secondly non-trivial correlations can still take place as well. A more advanced model, as will be considered in the analysis in Part II, accounts for all this in an iterative procedure where low-p_T interactions, ordered in scales, are given a probability to reconnect with the interaction of highest-p_T, defined as follows:

$$p_{\text{reconnect}}(p_T) = \frac{(R + p_{T0})^2}{(R + p_{T0})^2 + p_T^2} \tag{2.62}$$

where R is the free parameter of the model, and p_{T0} is taken as in Eq. 2.60.

20. In fact, it is itself decomposed in further parameters with a power-low function of the centre-of-mass energy of the collisions:

$$p_{T0}\left(\sqrt{s}\right) = p_{T0}^{\text{ref}} \left(\frac{\sqrt{s}}{\sqrt{s_0}} \right)^{\epsilon} \tag{2.60}$$

The three parameters p_{T0}^{ref}, $\sqrt{s_0}$ and ϵ can be determined with data samples from different experiments. This is not crucial for the current discussion but will be discussed in Part. II.

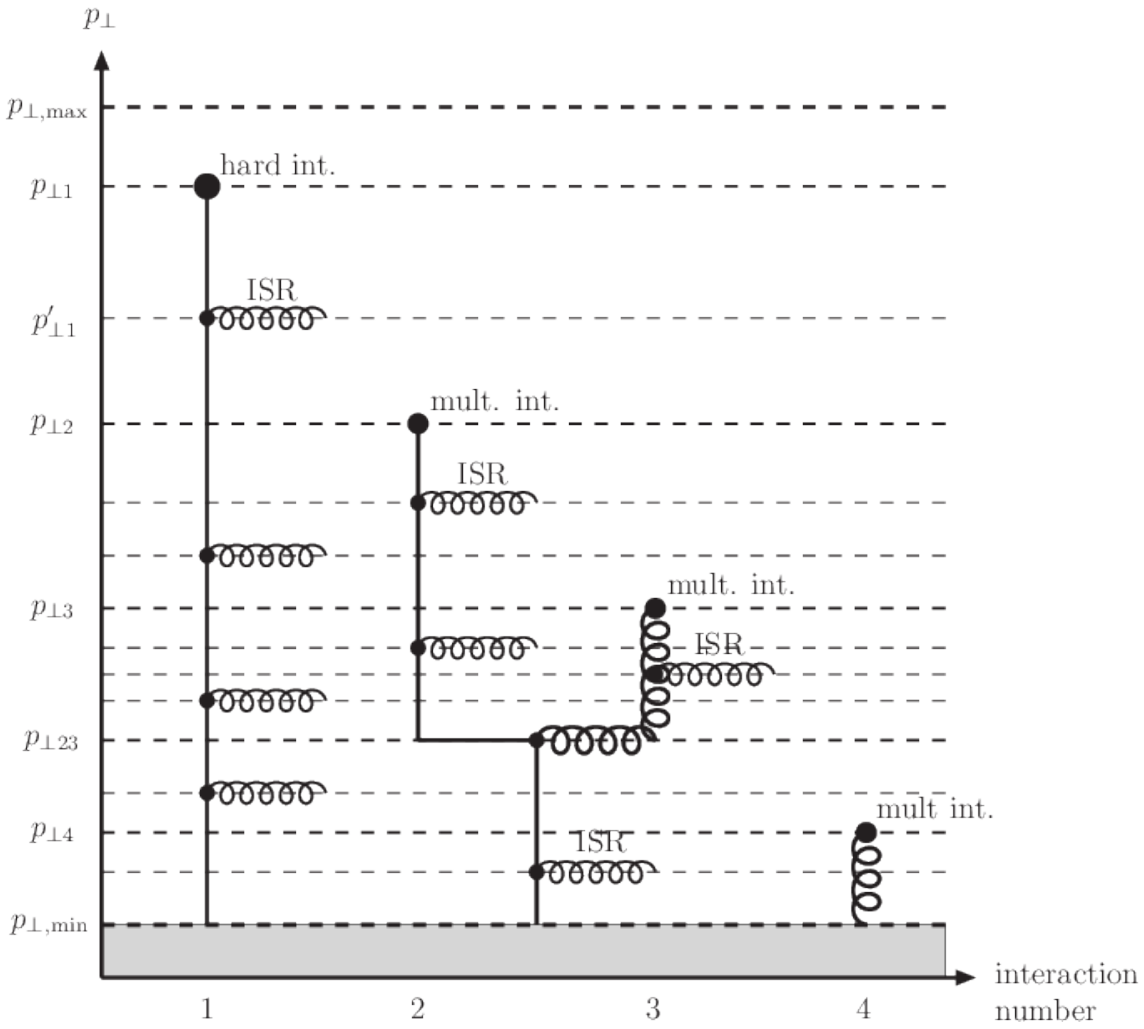


FIGURE 2.7 – "Schematic figure illustrating one incoming hadron in an event with a hard interaction occurring at p_{T1} and three further interactions at successively lower p_T scales, each associated with (the potentiality of) initial-state radiation, and further with the possibility of two interacting partons (2 and 3 here) having a common ancestor in the parton showers. Full lines represent quarks and spirals gluons. The vertical p_T scale is chosen for clarity rather than realism; most of the activity is concentrated to small p_T values." [56]

Beam-Beam Remnants. Eventually, the dynamics of the spectator partons in the beam remnants need to be treated for several reasons:

— consistent treatment of the colour flow in the hadronisation,

— compensation for the intrinsic k_T given to the hard partons,

— conservation of the quantum numbers.

The PDFs are adapted according to the partons that participated to the MPI: for n interactions, there can be up to $n+3$ partons in the beam remnant; if a valence quark interacts, it is taken into account in the remaining valence band; if a sea quark interacts, it must leave a partner in the beam. Being coloured, the spectator partons are consequently responsible of a part of the hadronic activity in the forward region.

2.3.3.4 Hadronisation

Once the PS has stopped at a scale, the hadronisation takes place. In Fig. 2.6b, the *colour flow* after showering is illustrated. Partons are connected to one another and the hadronisation will transform the (coloured) partons to (white) hadrons. Given the chain of colour-connected partons, even causally separated partons may be involved in the same hadronisation process.

Formally, the hadronisation is described by FFs, which are analog to the PDFs for the transition from the perturbative regime to the non-perturbative regime. For processes defined by hadrons in the final state, the factorisation (Eq. 2.37) can be rewritten with additional FFs.

Two models exist: the *string model* and the *cluster model*, illustrated in Fig. 2.9 with e^+e^- scattering. Both treat colour in the large-N colour structure, where gluons may be considered as carrying one colour charge and one anti-colour charge.

Lund string fragmentation model. The interaction of any two connected partons is described by a *string* [57, 60]. Observations from Lattice QCD confirm that the colour field is concentrated in the form of a string [61]. In their centre-of-mass frame, the partons move apart from one another nearly at the speed of light; due to the property of confinement, the kinetic energy is transformed in potential energy; the potential energy is in turn converted into a $q\bar{q}$ pair of quarks, breaking the string by screening effect; this is illustrated in Fig. 2.8, where, on the left, the three stages from the string to the pair creations are shown, as initial quarks are getting further and further apart from one another, and on the right, the strings are represented and illustrate the screening effect at the production of a new pair. In the Lund fragmentation model, the FF are defined as follows:

$$f(z, m_T) = \frac{1}{z}(1-z)^a \exp\left(-b\frac{m_T^2}{z}\right) \tag{2.63}$$

where z is the longitudinal momentum fraction of the hadron from the quark, and $m_T^2 = m^2 + p_T^2$. The parameters a and b are not given by the theory and can depend on the flavour. One usually distinguishes different pairs of parameters (a, b) for light quarks (u, d, s), for charm quarks and for beauty quarks, but not for the top quark, since it is too heavy to participate to the hadronisation. Moreover, for HF quarks, an additional factor has been proposed [62]:

$$f_Q(z, m_T) = \frac{1}{z^{r_Q b m_Q^2}} \times f(z) \tag{2.64}$$

where Q is a generic notation for HF quark.

Cluster fragmentation model. Alternatively, the hadronisation is treated in smaller, white ensembles of partons with invariant mass in practice below 3 GeV, where the gluons are forced to split into quarks are forced to split into $q\bar{q}$ pairs (Fig. 2.9b) [63, 64, 55]. Similar dynamics as in the string fragmentation model is then adopted.

In the measurement presented in Part II, the final state is defined in terms of B hadrons rather than in terms of b quarks. Therefore we should in principle distinguish the notation of b jets and B jets; however, in order to stay consistent with the Compact Muon Solenoid (CMS) convention [66], we shall keep the notation "b jets".

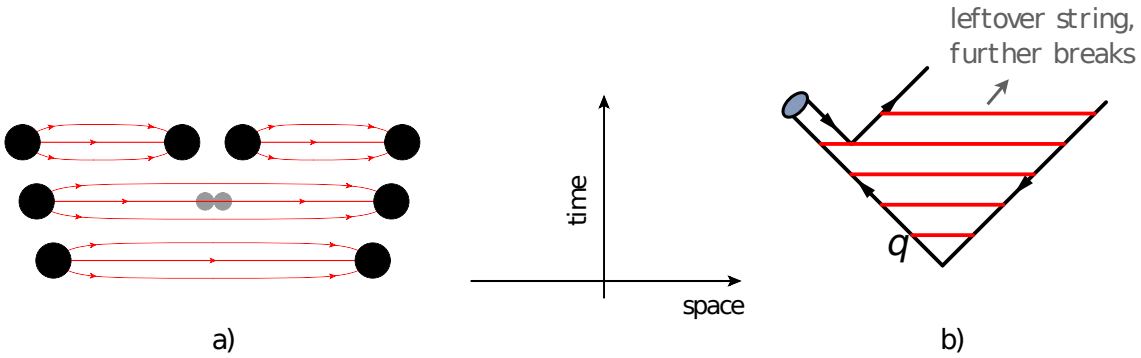
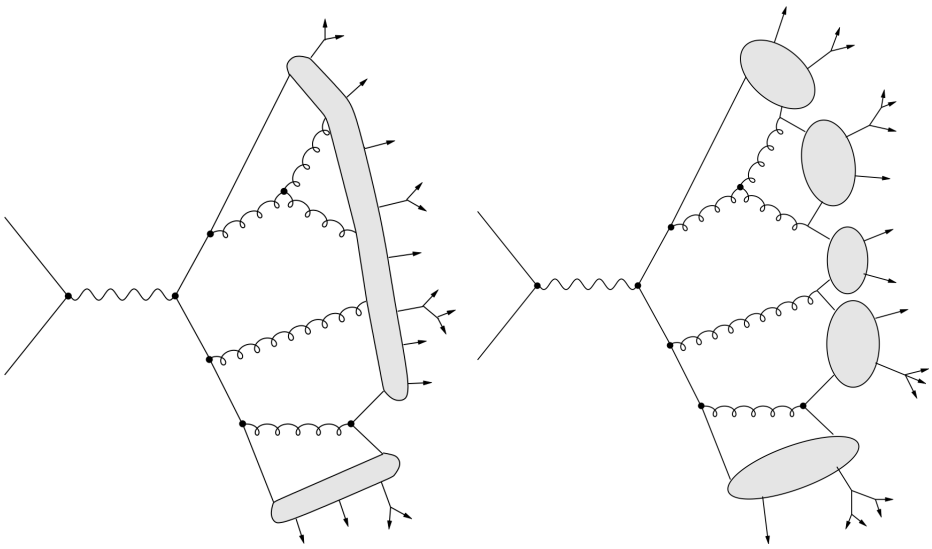


FIGURE 2.8 – "a. Illustration of string breaking by quark pair creation in the string field. b. Illustration of the algorithmic choice to process the fragmentation from the outside-in, splitting off a single on-shell hadron in each step." [22].



(A) String model. (B) Cluster model.

FIGURE 2.9 – Sketch of the two main hadronisation models in *ep* scattering [65]. The blobs correspond to the region where hadronisation processes take place. In the string model, far away separated partons may be taken into account in the same process; in the cluster model, smaller ensemble are first distinguished.

2.3.3.5 Stable particles

Hadrons differ significantly in terms of lifetime, which is important to treat in order to reflect a real event in a detector. In Tab. 2.5, the decay length of the most frequent hadrons are given, corresponding to the lifetime in the rest frame; in practice, these objects may be boosted, changing significantly the actual distance.

In parallel, the radiation of soft photons take place.

2.3.4 More on evolution

The discussion of the previous section was entirely conducted in the *collinear factorisation*, described with the DGLAP evolution. Albeit very successful for the description of many measurements, it is conceptually not completely satisfying, since it does not include a proper description of the transverse momentum (*cf.* intrinsic k_T in the PS).

particle	symbol	content (if composite)	$c\tau$
electron	e	—	∞
muon	μ	—	$\sim 659\,\mathrm{m}$
tauon	τ	—	$\sim 87\,\mu\mathrm{m}$
proton	p	uud	∞
neutron	n	udd	$2.64\,\mathrm{km}$
charged pion	$\pi^{\pm}$	$\bar{u}d$	$\sim 7.80\,\mathrm{m}$
neutral pion	π^0	$\bar{u}u$	$\sim 25.5\,\mathrm{nm}$
charged kaon	$K^{\pm}$	$\bar{u}s/\bar{s}u$	$\sim 3.71\,\mathrm{m}$
neutral short-lived kaon	K^0_S	$\bar{d}s/\bar{s}d$	$\sim 2.68\,\mathrm{cm}$
neutral long-lived kaon	K^0_L	$\bar{d}s/\bar{s}d$	$\sim 15.4\,\mathrm{m}$
positive sigma	Σ^+	uus	$\sim 2.40\,\mathrm{cm}$
negative sigma	Σ^-	dds	$\sim 4.43\,\mathrm{cm}$
neutral xi	Ξ^0	uss	$\sim 8.71\,\mathrm{cm}$
charged xi	$\Xi^{\pm}$	dss	$\sim 4.91\,\mathrm{cm}$
lambda	Λ^0	uds	$\sim 7.89\,\mathrm{cm}$
charmed lambda	$\Lambda^{\pm}_c$	udc	$\sim 59.9\,\mu\mathrm{m}$
bottomed lambda	Λ^0_b	udb	$\sim 439\,\mu\mathrm{m}$
charged D meson	$D^{\pm}$	$\bar{c}d$	$\sim 312\,\mu\mathrm{m}$
neutral D meson	D^0	$\bar{c}u$	$\sim 123\,\mu\mathrm{m}$
strange D meson	$D^{\pm}_s$	$\bar{c}s$	$\sim 150\,\mu\mathrm{m}$
charged B meson	$B^{\pm}$	$b\bar{d}$	$\sim 491\,\mu\mathrm{m}$
neutral B meson	B^0	$b\bar{u}$	$\sim 455\,\mu\mathrm{m}$
strange B meson	$B^{\pm}_s$	$b\bar{s}$	$\sim 453\,\mu\mathrm{m}$
charmed B meson	B^0_c	$b\bar{c}$	$\sim 152\,\mu\mathrm{m}$

TABLE 2.5 – Selection of particles typically found in the final state in the detector with decay lengths [15]. The value are given without uncertainty (hence the $\sim$ symbol) to give a general survey of the behaviour of the hadrons in a typical modern experiment. Unless relevant, the antiparticle is not specified.

In general, evolution equations can be seen as renormalisation equation for certain quantities like PDFs [54]. The DGLAP evolution equation is derived in the context of the collinear factorisation; other factorisation scheme exist, as will be described in this section.

2.3.4.1 BFKL evolution

The BFKL [21] [67, 68, 69, 70] consists in an integro-differential equations in k_T and x for unintegrated gluon distribution G:

$$\frac{\mathrm{d}G(x,k_T^2)}{\mathrm{d}\ln 1/x} \sim \alpha_S \int \frac{\mathrm{d}q^2}{q^2} K\left(\frac{k_T^2}{q^2}\right) G(x,q^2) \tag{2.65}$$

where the unintegrated distribution can be related to the integrated gluon distribution by integrating over the transverse momentum k:

$$g(x,Q^2) = \int^{Q^2} \mathrm{d}^2 k G(x,k^2) \tag{2.66}$$

21. Named after the four physicists BALITSKY, FADIN, KURAEV and LIPATOV.

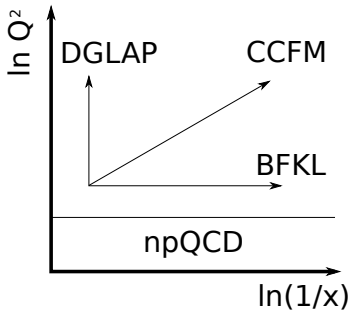


FIGURE 2.10 – Diagram of QCD evolution. The different schemes are represented: DGLAP describes evolution according to the scale Q^2; BFKL describes evolution according to the momentum fraction x; CCFM describes evolution as a function of both.

The condition for the application of this equation is a strong ordering in the successive momentum fractions, rather than in the scale. The situation can be seen in analogy to the DGLAP evolution where the rôles of x and Q^2 have swapped. At very low x, recombination effects from gluons are expected to become significant; this effect is called *saturation*. In case of saturation, BFKL does not apply anymore.

Experimental conditions to test BFKL equations require to go in the very forward region of the detector, because small x values means large rapidity separation:

$$\Delta y \sim \ln \frac{1}{x} \tag{2.67}$$

Experimental evidence was found with the so-called *Mueller-Navelet jets* [71]. In general, large rapidity separations will not be discussed in this thesis.

2.3.4.2 CCFM evolution

The CCFM[22] [72, 73, 74, 75] evolution equation is also interesting since it includes a dependence in transverse momentum of PDFs:

$$f(x, Q^2) = \int \mathrm{d}k_T \mathcal{A}(x, k_t, Q^2) \tag{2.68}$$

where f ($\mathcal{A}$) is the *collinear PDF* (TMD, or Transverse-Momentum-Dependent PDF).

In this approach, Eq. 2.37 can be explicitly written as follows:

$$\begin{aligned}\sigma_{h_1 h_2 \to X} = \sum_{a \in \{q,g\}} \int \mathrm{d}x_a \, \mathrm{d}k_{Ta} \mathcal{A}_{a \in h_1}(x_a, k_{Ta}, \mu_F^2) \times \\ \times \sum_{b \in \{q,g\}} \int \mathrm{d}x_b \, \mathrm{d}k_{Tb} \mathcal{A}_{b \in h_2}(x_b, k_{Tb}, \mu_F^2) \, \mathrm{d}\sigma_{ab}\end{aligned} \tag{2.69}$$

The angular ordering is applied. Nonetheless, only gluons are included in this evolution; therefore it is not suited for high transverse momentum regions, where quarks have a significant contribution, such as in this thesis.

22. Named after the four physicists CATANI, CIAFALONI, FIORANI and MARCHESINI.

Conclusion

In this chapter, the basic elements of theory to understand proton-proton collisions at LHC have been introduced. The complexity of performing predictions has been explained, and the difficulties inherent to hadrons, related to the property of confinement, has been presented. This portrait of HEP has been intended to be general.

Apart of the next chapter, dedicated to the description of the experimental set-up, we shall focus more on topics related to b physics, from generalities to the measurement of the production of B jets at hadron colliders.

In appendix of this chapter are a historical introduction to HEP (with a more theoretical point of view than in Chap. 1) and a description of some typical coordinates and variables used in the context of general-purpose hadron colliders.

2.A Historical perspective

Modern physics started with Galileo Galilei, at the beginning of the 17th century, who first formulated the *Principle of Relativity*, stating that the laws of mechanics should be the same in all (inertial) frames [76].

The formulation of the three *laws of motion*[23] came up at the end of the 17th century with Isaac Newton [1]. These laws assume that the evolution of a particle follows a trajectory.

In parallel of his work on mechanics, Newton also wrote a treatise on *optics* [77], where he assumed the light of *corpuscular* nature.

At the beginning of the 18th century, Thomas Young performed the *double-slit experiment* with lights, highlighting the wave nature of light, in contradiction with the idea of light made of classical corpuscles.

During the 19th century, James Clerk Maxwell synthesised the *laws of electromagnetism* in four equations[24] [78, 79, 80]. These four equations had two consequences: first, solutions to the Maxwell equations led to a description of light in terms of electromagnetic waves, confirming Young's interpretation; secondly, these equations did not respect the Principle of Relativity as expressed by Galilei.

At the end of the 19th century, Hendrik Lorentz and Albert Einstein found a new formulation of the Principle of Relativity, where the speed of light c appears as a funda-

23. Personal translation from Latin:

Principle of Inertia: « Every body will to stay in its state of rest or of uniformly straight motion, unless it is compelled to change its state by a force that applies on it. »

Fundamental principle of dynamics: « The change of the motion is proportional to the force that is applied on it, and is applied in straight line with respect to it. (...) »

Principe of reciprocal actions: « The reaction is always equal and opposite to the action (...) »

24. Given the electric charge density ρ and is the electric current density $\mathbf{j}$, the electric and magnetic fields $\mathbf{E}$ and $\mathbf{B}$ are described by the four following equations:

Gauss $\nabla \cdot \mathbf{E} = \rho/\epsilon_0$

Maxwell $\nabla \cdot \mathbf{B} = 0$

Faraday $\nabla \times \mathbf{E} = -\partial\mathbf{B}/\partial t$

Ampère $\nabla \times \mathbf{B} = \mu_0\mathbf{j} + \epsilon_0\mu_0\partial\mathbf{E}/\partial t$

mental constant of physics, valid in all frames [81, 82, 83]; today, the "relativistic" vocable refers to this version of the Principle of Relativity. In addition, Maxwell noticed that the electric and magnetic field were invariant under a *Gauge transformation*[25].

At the beginning of the 20th century, in parallel of the development of relativity, Max Planck, in his attempts to describe the *radiation of the black body*, first used discrete levels of energy [84, 85, 86]. Although he himself did not believe at first in it, it turned out that *quantisation* could explain several phenomena: for instance, the *photoelectric effect* [87] or the *spectral lines* with the atomic *Bohr model* [88, 89]. (The repetition of Young's experiment with electrons or larger molecules, shown in Fig. 2.11, would come only in the second half of the 20th century [90, 91].) The general description of (non-relativistic) quantic states was obtained by Erwin Schrödinger[26] [92, 93, 94, 95, 96]; while Newton's laws of motion describe the evolution of a particle with a trajectory, the Schrödinger equation describes the evolution of a particle with a *wave function*, which is then understood as a probability amplitude. Experimentally, Davisson and Germer highlighted the wave-like character of the motion of electrons [97], similarly to the light.

In the context of quantum mechanics, a new quantity, showing the properties of an intrinsic angular momentum, was discovered: the *spin*. The *Stern-Gerlach experiment* [99, 100, 101], illustrated in Fig. 2.12, showed that its value is non-integer; it would turn out that it can take any half-integer for values. In a phenomenological approach, Wolfgang Pauli formulated an equation[27] with a 2-component wave function, called a *spinor*.

The direct generalisation of the Schrödinger equation to the relativistic regime is called the Klein-Gordon equation[28]. Unlike the Schrödinger equation, the interpretation of the wave function as a probability amplitude was not so clear anymore, because the conserved quantity — which corresponds to the probability in the case of the Schrödinger equation — was no more positive definite. Moreover the Klein-Gordon equation is only able to describe the dynamics of spinless particles, which does not include particles like the electron; experimental possibilities were limited at the time of its derivation.

25. The electric and magnetic fields can be deduced from a scalar and a vectorial potentials ϕ and $\mathbf{A}$:

$$\mathbf{E} = -\nabla\phi - \frac{\partial \mathbf{A}}{\partial t}, \quad \mathbf{B} = \nabla \times \mathbf{A} \tag{2.70}$$

These potentials are not unique; certain transformations, called *Gauge transformations*, lead to the same evolution:

$$\phi \longrightarrow \phi + \frac{\partial \Lambda}{\partial t}, \quad \mathbf{A} \longrightarrow \mathbf{A} + \nabla\Lambda \tag{2.71}$$

for any function Λ.

26. Schrödinger equation:

$$\left(\frac{\widehat{\mathbf{p}}}{2m} + \widehat{\mathbf{U}}(t)\right)|\psi\rangle = E\,|\psi\rangle \tag{2.72}$$

27. Pauli equation:

$$\left[\frac{1}{2m}\left(\sigma \cdot (\mathbf{p} - q\mathbf{A})\right)^2 + q\phi\right]|\psi\rangle = i\frac{\partial}{\partial t}|\psi\rangle \tag{2.73}$$

for a half-spin particle in an electric (magnetic) field ϕ ($\mathbf{A}$), with Pauli matrices:

$$\sigma_1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \sigma_2 = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} \quad \sigma_3 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \tag{2.74}$$

28. While the Schrödinger equation is built up on the dispersion relation $E = p^2/2m$, the Klein-Gordon equation is built up on its relativistic analog $E^2 = m^2 + p^2$: $\Box\phi + m^2\phi = 0$.

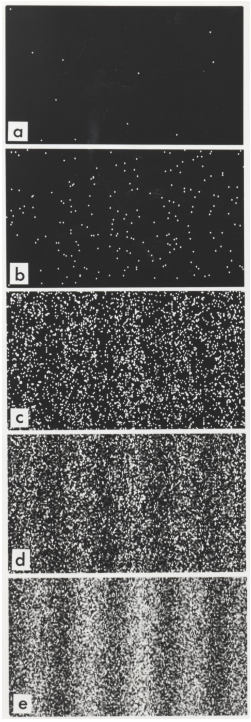


FIGURE 2.11 – "Results of a double-slit-experiment performed by Dr. TONOMURA showing the build-up of an interference pattern of single electrons. Numbers of electrons are 11 (a), 200 (b), 6000 (c), 40000 (d), 140000 (e)." [98]

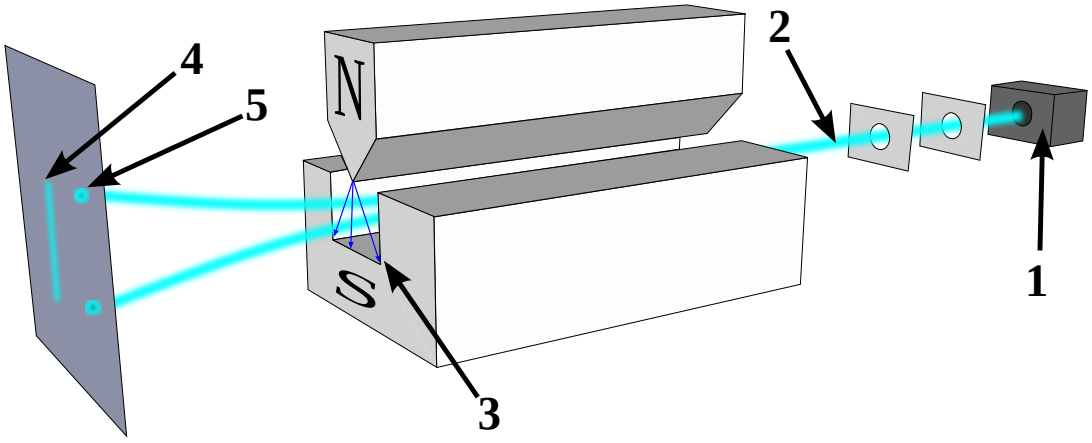


FIGURE 2.12 – "Stern-Gerlach experiment: silver atoms travel through an inhomogeneous magnetic field and are deflected up or down depending on their spin. 1: furnace. 2: beam of silver atoms. 3: inhomogeneous magnetic field. 4: expected result. 5: what was actually observed." [102]

Paul Dirac was the first in 1928 to derive an equation [29] [103, 104] in the relativistic regime to describe (free) particles with half spin, *e.g.* electrons. Expecting 2-components wave functions like in the Pauli equation, Dirac built up his equation in such a way that each component would respect the Klein-Gordon equation. However, it turned out that in order to fulfill all requirements, the wave function needed not two but four components; in addition, some solutions seemed to correspond to negative energies. By introducing an *ad hoc* term to describe interactions with an electromagnetic field [30], one could find in the non-relativistic limit the Pauli equation for two of the four components; the two additional components suggested the existence of *antiparticles*, soon confirmed by Carl David Anderson with positrons [105].

Eventually, the interpretation of the wave function was abandoned — some solutions had negative energies, and the causality was not respected — for that of *quantum field*; a quantic and relativistic theory must fundamentally treat of multi-particle systems. Instead of describing the quantic particle itself and its probability amplitude of being in a given state, the quantum field describes creation and annihilation of particles and antiparticles. Particles are only excitation states of the quantum field, which fills up the whole space.

Soon after, the Proca equation [31], the EOM for spin-1 particles, came up, as well as the *spin-statistics theorem* [106] establishing the relations between the spin and the statistics of the particles; one could therefore distinguish two types:

fermions half-integer-spin particles (obeying the *Pauli Exclusion Principle*);

bosons integer-spin particles (which can be superposed abritrarily).

In parallel, after the discovery of the neutron by James Chadwick [107], Hideki Yukawa developped a theory of nuclear interaction.

The next decades would show the discovery of many new particles and phenomena. Before closing this section, two of them at least should still be mentioned: first, the experimental discovery of partons, predicted by Gell'mann and Feynman at Stanford National Accelerator Laboratory (SLAC) in 1968 (see Sect. 1.2.4) [108], and then the discovery of the colour charge with the Δ^{++}. This marked the day of birth of the QCD, which we are treating in this thesis.

The SM was developed in the following decades with growing successes. Today, all particles predicted by the SM have been discovered, all parameters have been measured. Still, many theoretical questions stay open, like the group structure of SM and some apparent symmetries.

29. Given a four-component wave function ψ:

$$(i\gamma^\mu \partial_\mu - m)\,\psi = 0 \tag{2.75}$$

where

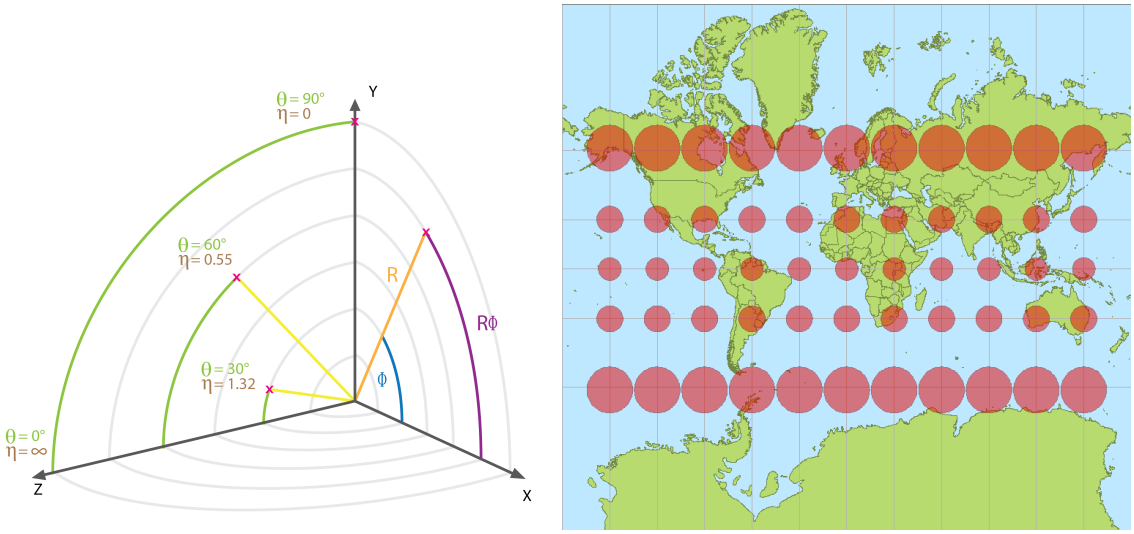
$$\gamma^0 = \begin{bmatrix} 1_2 & 0 \\ 0 & 1_2 \end{bmatrix}, \quad \gamma^i = \begin{bmatrix} 0 & -\sigma_i \\ \sigma_i & 0 \end{bmatrix} \tag{2.76}$$

30. Given a four-component wave function ψ:

$$(\gamma^\mu\,(i\partial_\mu - A_\mu) - m)\,\psi = 0 \tag{2.77}$$

31.

$$\partial_\mu F^{\mu\nu} + m^2 A^\nu = 0 \tag{2.78}$$



(A) Correspondance with spherical coordinates [109].

(B) Mercator projection of the planet Earth [110]: the red disks correspond to disks of same area on the globe, illustrating the effect of the conformal transformation in Eq. 2.79.

FIGURE 2.13 – Conventional coordinates at CMS. z is the axis of the beam, $\eta = -\ln\tan\frac{\theta}{2}$, ϕ the azimuthal angle.

2.B Variables and coordinates

We recapitulate here some of the standard variables and coordinates used in HEP.

2.B.1 Coordinate system

We adopt here the conventional coordinate system used at CMS, as shown in Fig. 2.13a:

— The x axis points toward the centre of ring of the collider.

— The y axis points to the sky.

— The z axis is defined along the colliding beams with the right hand rule.

Given the symmetry of the detectors and of the topology of the collisions, cylindrical coordinates are often used.

However, instead of using a polar angle θ, one uses *pseudorapidity*:

$$\eta = -\ln\tan\frac{\theta}{2} \tag{2.79}$$

The pseudorapidity η is convenient since the difference of pseudorapidities $\Delta\eta$ are *almost* Lorentz-invariant quantity. The pseudorapidity also corresponds to the Mercator projection used for maps of the Earth (see Fig. 2.13b).

The *rapidity* y is defined as follows:

$$y = \frac{1}{2}\ln\frac{E + p_z}{E - p_z} \tag{2.80}$$

In cylindrical coordinates, the difference of rapidities Δy is Lorentz invariant along the z axis, and can be related to the polar angle for small masses, *i.e.* $m \ll |\mathbf{p}|$:

$$\begin{aligned} y &\approx \frac{1}{2} \ln \frac{|\mathbf{p}| + p_z}{|\mathbf{p}| - p_z} \\ &= \frac{1}{2} \ln \frac{1 + \cos\theta}{1 - \cos\theta} \\ &= -\ln \tan \frac{\theta}{2} \equiv \eta \end{aligned}$$

The transformation from the polar angle to the pseudorapidity is a conformal transformation, which means that angles are conserved and that a circle is transformed in a circle; it is therefore well suited for the definition of jet cone.

2.B.2 Mandelstam variables

In the case of $2 \to 2$ processes, it is convenient to define the *Mandelstam variables*. The Mandelstam variables correspond to three Lorentz-invariant quantities with interesting properties.

2.B.2.1 Definition

The Mandelstam variables are defined as follows:

$$s = (p_1 + p_2)^2 = (p_3 + p_4)^2 \tag{2.81}$$

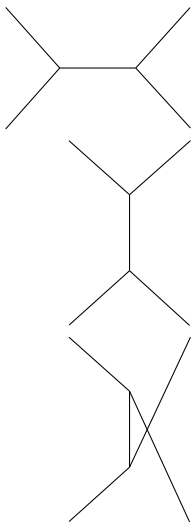
$$t = (p_1 - p_3)^2 = (p_4 - p_2)^2 \tag{2.82}$$

$$u = (p_1 - p_4)^2 = (p_2 - p_3)^2 \tag{2.83}$$

For instance, s coincides with the squared centre-of-mass energy of a collision, and is an important parameter for colliders; t coincides with the *momentum transfer* of a collision.

2.B.3 Channels.

The Mandelstam variables are also used to designate the three channels at tree level:



(s-channel) (2.84)

(t-channel) (2.85)

(u-channel) (2.86)

The permutation of two Mandelstam variables corresponds to the permutation of legs in a diagram; the MEs of different processes may be easily related thanks to this property.

Parton convention. If the scattering involves hadrons, the Mandelstam variables at the parton-level (hadron-level) are used with (without) hat. In the collinear factorisation, the s variables at hadron and parton levels can be related with the momentum fractions; for DIS and pp scatterings respectively:

$$\hat{s} = x_1 s \qquad \text{(DIS)} \tag{2.87}$$

$$\hat{s} = x_1 x_2 s \qquad (pp) \tag{2.88}$$

2.B.4 Properties

General. Let m_i be the masses of the interacting particles:

$$s + t + u = \sum_{i=1}^{4} m_i^2 \tag{2.89}$$

Centre-of-mass frame. The t and u variables can be related to the polar angle in the centre-of-mass (c.m.s.) frame.

$$t = \frac{-s}{2}(1 - \cos\theta) \tag{2.90}$$

$$u = \frac{-s}{2}(1 + \cos\theta) \tag{2.91}$$

$$\tag{2.92}$$

Massless approximation. If the mass can be neglected, the form of the Mandelstam variables is simplified to the following:

$$s \approx 2p_1 \cdot p_2 \tag{2.93}$$

$$t \approx -2p_1 \cdot p_3 \tag{2.94}$$

$$u \approx -2p_1 \cdot p_4 \tag{2.95}$$

References

[1] Isaac Newton. *Philosophiae Naturalis Principia Mathematica*. Ed. by Wikimedia. https://la.wikisource.org/wiki/Philosophiae_Naturalis_Principia_Mathematica. 1687.

[2] Günther Dissertori, Ian G Knowles, and Michael Schmelling. *Quantum chromodynamics: high energy experiments and theory*. Vol. 115. Oxford University Press, 2003.

[3] F. Halzen and Alan D. Martin. *Quarks and Leptons: An Introductory Course In Modern Particle Physics*. 1984.

[4] Michael E. Peskin and Daniel V. Schroeder. *An Introduction to quantum field theory*. Reading, USA: Addison-Wesley, 1995. URL: http://www.slac.stanford.edu/~mpeskin/QFT.html.

[5] Thomas Hambye. "Introduction au Modèle Standard". Course of 1st master year, ULB. No written version. 2013.

[6] Thomas Hambye and Michel H.G. Tytgat. "An Introduction to Elementary Particles". Theoretical aspects of the course of 3rd bachelor year, ULB. 2012.

[7] Marc Henneaux. "Groupes et représentations". http://www.solvayinstitutes.be/pdf/Marc/SO3SO3_1_18112014.pdf. b Lectures at ULB.

[8] Marc Henneaux. "Représentation des groupes et application à la physique". http://www.solvayinstitutes.be/pdf/Marc/Groupes22062015.pdf. c Lectures at ULB.

[9] B. Felsager. *Geometry, Particles and Fields*. Graduate Texts in Contemporary Physics. Odense: Univ.Pr., 1981. DOI: 10.1007/978-1-4612-0631-6.

[10] C. N. Yang and R. L. Mills. "Conservation of Isotopic Spin and Isotopic Gauge Invariance". In: *Phys. Rev.* 96 (1 Oct. 1954), pp. 191–195. DOI: 10.1103/PhysRev.96.191. URL: https://link.aps.org/doi/10.1103/PhysRev.96.191.

[11] Howard Georgi. "Lie algebras in particle physics". In: *Front. Phys.* 54 (1999), pp. 1–320.

[12] Gerardus' t Hooft. "Renormalization of massless Yang-Mills fields". In: *Nuclear physics: B* 33.1 (1971), pp. 173–199.

[13] Glenn Barnich. "Localité et symétries en théorie classique des champs". http://homepages.ulb.ac.be/%7Egbarnich/symlectures.pdf. Lectures at ULB.

[14] Makoto Kobayashi and Toshihide Maskawa. "CP-Violation in the Renormalizable Theory of Weak Interaction". In: *Progress of Theoretical Physics* 49.2 (1973), pp. 652–657. DOI: 10.1143/PTP.49.652. eprint: /oup/backfile/content_public/journal/ptp/49/2/10.1143/ptp.49.652/2/49-2-652.pdf. URL: +%20http://dx.doi.org/10.1143/PTP.49.652.

[15] C. Patrignani et al. "Review of Particle Physics". In: *Chin. Phys.* C40.10 (2016), p. 100001. DOI: 10.1088/1674-1137/40/10/100001.

[16] R Keith Ellis, W James Stirling, and Bryan R Webber. *QCD and collider physics*. Cambridge university press, 2003.

[17] Serguei Chatrchyan et al. "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC". In: *Physics Letters B* 716.1 (2012), pp. 30–61.

[18] Georges Aad et al. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". In: *Physics Letters B* 716.1 (2012), pp. 1–29.

[19] F. Englert and R. Brout. "Broken Symmetry and the Mass of Gauge Vector Mesons". In: *Phys. Rev. Lett.* 13 (1964), pp. 321–323. DOI: 10.1103/PhysRevLett.13.321.

[20] Peter W. Higgs. "Broken Symmetries and the Masses of Gauge Bosons". In: *Phys. Rev. Lett.* 13 (1964), pp. 508–509. DOI: 10.1103/PhysRevLett.13.508.

[21] Laurent Favart. "Physique auprès des collisionneurs". Course of 1st master year, ULB. 2013.

[22] Peter Skands. "Introduction to QCD". In: *Proceedings, Theoretical Advanced Study Institute in Elementary Particle Physics: Searching for New Physics at Small and Large Scales (TASI 2012): Boulder, Colorado, June 4-29, 2012*. 2013, pp. 341–420. DOI: 10.1142/9789814525220_0008. arXiv: 1207.2389 [hep-ph]. URL: https://inspirehep.net/record/1121892/files/arXiv:1207.2389.pdf.

[23] Daniel Britzger et al. "Determination of the strong coupling constant from inclusive jet cross section data from multiple experiments". In: (2017). arXiv: 1712.00480 [hep-ph].

[24] Richard Phillips Feynman. "Space-time approach to quantum electrodynamics". In: *Physical Review* 76.6 (1949), p. 769.

[25] Henrik Bohr and H.B. Nielsen. "Hadron production from a boiling quark soup: A thermodynamical quark model predicting particle ratios in hadronic collisions". In: *Nuclear Physics B* 128.2 (1977), pp. 275–293. ISSN: 0550-3213. DOI: https://doi.org/10.1016/0550-3213(77)90032-3. URL: http://www.sciencedirect.com/science/article/pii/0550321377900323.

[26] Yoichiro Nambu. "The confinement of quarks". In: *Scientific American* 235.5 (1976), pp. 48–63.

[27] Kenneth G. Wilson. "Confinement of quarks". In: *Phys. Rev. D* 10 (8 Oct. 1974), pp. 2445–2459. DOI: 10.1103/PhysRevD.10.2445. URL: https://link.aps.org/doi/10.1103/PhysRevD.10.2445.

[28] Rajan Gupta. "Introduction to lattice QCD: Course". In: *Probing the standard model of particle interactions. Proceedings, Summer School in Theoretical Physics, NATO Advanced Study Institute, 68th session, Les Houches, France, July 28-September 5, 1997. Pt. 1, 2*. 1997, pp. 83–219. arXiv: hep-lat/9807028 [hep-lat].

[29] Heinz J Rothe. *Lattice gauge theories: an introduction*. World Scientific, 2005.

[30] J. Alwall et al. "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations". In: *JHEP* 07 (2014), p. 079. DOI: 10.1007/JHEP07(2014)079. arXiv: 1405.0301 [hep-ph].

[31] "Pseudorapidity distribution of charged hadrons in proton – proton collisions at s=13 TeV". In: *Physics Letters B* 751.Supplement C (2015), pp. 143–163. ISSN: 0370-2693. DOI: https://doi.org/10.1016/j.physletb.2015.10.004. URL: http://www.sciencedirect.com/science/article/pii/S0370269315007558.

[32] CMS Collaboration. "Measurement of pseudorapidity distributions of charged particles in proton-proton collisions at $\sqrt{s} = 13$ TeV by the CMS experiment." In: (2016).

[33] A.C. Irving and R.P. Worden. "Regge phenomenology". In: *Physics Reports* 34.3 (1977), pp. 117–231. ISSN: 0370-1573. DOI: https://doi.org/10.1016/0370-1573(77)90010-2. URL: http://www.sciencedirect.com/science/article/pii/0370157377900102.

[34] P.D.B. Collins. "An Introduction to Regge Theory and High Energy Physics Cambridge Univ". In: *Press, Cambrige, NY, Melbourne* (1977).

[35] John C. Collins, Davison E. Soper, and George F. Sterman. "Factorization of Hard Processes in QCD". In: *Adv. Ser. Direct. High Energy Phys.* 5 (1989), pp. 1–91. DOI: 10.1142/9789814503266_0001. arXiv: hep-ph/0409313 [hep-ph].

[36] S. Alekhin, J. Blumlein, and S. Moch. "The ABM parton distributions tuned to LHC data". In: *Phys. Rev.* D89.5 (2014), p. 054028. DOI: 10.1103/PhysRevD.89.054028. arXiv: 1310.3059 [hep-ph].

[37] J. Pumplin et al. "New generation of parton distributions with uncertainties from global QCD analysis". In: *JHEP* 07 (2002), p. 012. DOI: 10.1088/1126-6708/2002/07/012. arXiv: hep-ph/0201195 [hep-ph].

[38] Amanda M. Cooper-Sarkar. "HERAPDF1.5LO PDF Set with Experimental Uncertainties". In: *PoS* DIS2014 (2014), p. 032.

[39] L. A. Harland-Lang et al. "Parton distributions in the LHC era: MMHT 2014 PDFs". In: *Eur. Phys. J.* C75.5 (2015), p. 204. DOI: 10.1140/epjc/s10052-015-3397-6. arXiv: 1412.3989 [hep-ph].

[40] A. D. Martin et al. "Parton distributions for the LHC". In: *Eur. Phys. J.* C63 (2009), pp. 189–285. DOI: 10.1140/epjc/s10052-009-1072-5. arXiv: 0901.0002 [hep-ph].

[41] Richard D. Ball et al. "Parton distributions with LHC data". In: *Nucl. Phys.* B867 (2013), pp. 244–289. DOI: 10.1016/j.nuclphysb.2012.10.003. arXiv: 1207.1303 [hep-ph].

[42] Richard D. Ball et al. "Parton distributions for the LHC Run II". In: *JHEP* 04 (2015), p. 040. DOI: 10.1007/JHEP04(2015)040. arXiv: 1410.8849 [hep-ph].

[43] H. Abramowicz et al. "Combination of measurements of inclusive deep inelastic $e^{\pm}p$ scattering cross sections and QCD analysis of HERA data". In: *Eur. Phys. J.* C75.12 (2015), p. 580. DOI: 10.1140/epjc/s10052-015-3710-4. arXiv: 1506.06042 [hep-ex].

[44] D. Madigozhin. "Pion scattering lengths from NA48". In: *Nuclear Physics B - Proceedings Supplements* 164 (2007). QCD 05, pp. 85–88. ISSN: 0920-5632. DOI: https://doi.org/10.1016/j.nuclphysbps.2006.11.069. URL: http://www.sciencedirect.com/science/article/pii/S0920563206009212.

[45] Albert M Sirunyan et al. "Azimuthal correlations for inclusive 2-jet, 3-jet, and 4-jet events in pp collisions at $\sqrt{s} = 13$ TeV". In: (2017). arXiv: 1712.05471 [hep-ex].

[46] Ryan Atkin. "Review of jet reconstruction algorithms". In: *Journal of Physics: Conference Series* 645.1 (2015), p. 012008. URL: http://stacks.iop.org/1742-6596/645/i=1/a=012008.

[47] Gavin P. Salam. "Towards Jetography". In: *Eur. Phys. J.* C67 (2010), pp. 637–686. DOI: 10.1140/epjc/s10052-010-1314-6. arXiv: 0906.1833 [hep-ph].

[48] John 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. URL: http://lss.fnal.gov/cgi-bin/find_paper.pl?conf-90-249.

[49] Gavin P. Salam and Gregory Soyez. "A Practical Seedless Infrared-Safe Cone jet algorithm". In: *JHEP* 05 (2007), p. 086. DOI: 10.1088/1126-6708/2007/05/086. arXiv: 0704.0292 [hep-ph].

[50] Stefano Catani et al. "Longitudinally-invariant k $\perp$-clustering algorithms for hadron-hadron collisions". In: *Nuclear Physics B* 406.1-2 (1993), pp. 187–224.

[51] Yu L Dokshitzer et al. "Better jet clustering algorithms". In: *Journal of High Energy Physics* 1997.08 (1997), p. 001.

[52] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. "The Anti-k(t) jet clustering algorithm". In: *JHEP* 04 (2008), p. 063. DOI: 10.1088/1126-6708/2008/04/063. arXiv: 0802.1189 [hep-ph].

[53] Stefan Höche. "Introduction to parton-shower event generators". In: *Proceedings, Theoretical Advanced Study Institute in Elementary Particle Physics: Journeys Through the Precision Frontier: Amplitudes for Colliders (TASI 2014): Boulder, Colorado, June 2-27, 2014*. 2015, pp. 235–295. DOI: 10.1142/9789814678766_0005. arXiv: 1411.4085 [hep-ph]. URL: http://inspirehep.net/record/1328513/files/arXiv:1411.4085.pdf.

[54] Hannes Jung. *Lectures on QCD and Monte Carlo*. http://desy.de/~jung/qcd_and_mc_2015/lecture-writeup.pdf. 2015.

[55] M. Bahr et al. "Herwig++ physics and manual". In: *Eur. Phys. J. C* 58 (2008), p. 639. DOI: 10.1140/epjc/s10052-008-0798-9. arXiv: 0803.0883 [hep-ph].

[56] T. Sjostrand and Peter Z. Skands. "Transverse-momentum-ordered showers and interleaved multiple interactions". In: *Eur. Phys. J.* C39 (2005), pp. 129–154. DOI: 10.1140/epjc/s2004-02084-y. arXiv: hep-ph/0408302 [hep-ph].

[57] Torbjörn Sjöstrand, Stephen Mrenna, and Peter Skands. "PYTHIA 6.4 physics and manual". In: *JHEP* 05 (2006), p. 026. DOI: 10.1088/1126-6708/2006/05/026. arXiv: hep-ph/0603175 [hep-ph].

[58] Radek Zlebcik. *Illustration of initial- and final-state parton shower*. Private communication.

[59] Paolo Gunnellini. "Study of double parton scattering using four-jet scenarios in proton-proton collisions at $\sqrt{s} = 7$ TeV with the CMS experiment at the Large Hadron Collider". PhD thesis. U. Hamburg, Dept. Phys., 2014. DOI: 10.1007/978-3-319-22213-4, 10.3204/DESY-THESIS-2015-010. URL: http://pubdb.desy.de/search?of=hd&p=id:DESY-THESIS-2015-010.

[60] B. Andersson et al. "Parton fragmentation and string dynamics". In: *Physics Reports* 97.2 (1983), pp. 31–145. ISSN: 0370-1573. DOI: https://doi.org/10.1016/0370-1573(83)90080-7. URL: http://www.sciencedirect.com/science/article/pii/0370157383900807.

[61] GS Bali and K Schilling. "Static quark-antiquark potential: Scaling behavior and finite-size effects in SU (3) lattice gauge theory". In: *Physical Review D* 46.6 (1992), p. 2636.

[62] M. G. Bowler. "e+ e- Production of Heavy Quarks in the String Model". In: *Z. Phys.* C11 (1981), p. 169. DOI: 10.1007/BF01574001.

[63] Daniele Amati and Gabriele Veneziano. "Preconfinement as a property of perturbative QCD". In: *Physics Letters B* 83.1 (1979), pp. 87–92.

[64] B. R. Webber. "A QCD Model for Jet Fragmentation Including Soft Gluon Interference". In: *Nucl. Phys.* B238 (1984), pp. 492–528. DOI: 10.1016/0550-3213(84)90333-X.

[65] B. Webber. "Parton shower Monte Carlo event generators". In: *Scholarpedia* 6.12 (2011). revision #128236, p. 10662. DOI: 10.4249/scholarpedia.10662.

[66] Albert M Sirunyan et al. "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV". In: (2017). arXiv: 1712.07158 [physics.ins-det].

[67] Victor S. Fadin, E. A. Kuraev, and L. N. Lipatov. "On the Pomeranchuk Singularity in Asymptotically Free Theories". In: *Phys. Lett.* 60B (1975), pp. 50–52. DOI: 10.1016/0370-2693(75)90524-9.

[68] E. A. Kuraev, L. N. Lipatov, and Victor S. Fadin. "Multi - Reggeon Processes in the Yang-Mills Theory". In: *Sov. Phys. JETP* 44 (1976). [Zh. Eksp. Teor. Fiz.71,840(1976)], pp. 443–450.

[69] E. A. Kuraev, L. N. Lipatov, and Victor S. Fadin. "The Pomeranchuk Singularity in Nonabelian Gauge Theories". In: *Sov. Phys. JETP* 45 (1977). [Zh. Eksp. Teor. Fiz.72,377(1977)], pp. 199–204.

[70] I. I. Balitsky and L. N. Lipatov. "The Pomeranchuk Singularity in Quantum Chromodynamics". In: *Sov. J. Nucl. Phys.* 28 (1978). [Yad. Fiz.28,1597(1978)], pp. 822–829.

[71] Alfred H Mueller and H Navelet. "An inclusive minijet cross section and the bare pomeron in QCD". In: *Nuclear Physics B* 282 (1987), pp. 727–744.

[72] Marcello Ciafaloni. "Coherence effects in initial jets at small Q2/s". In: *Nuclear Physics B* 296.1 (1988), pp. 49–74.

[73] S Catani, F Fiorani, and G Marchesini. "QCD coherence in initial state radiation". In: *Physics Letters B* 234.3 (1990), pp. 339–345.

[74] S Catani, F Fiorani, and G Marchesini. "Small-x behaviour of initial state radiation in perturbative QCD". In: *Nuclear Physics B* 336.1 (1990), pp. 18–85.

[75] Giuseppe Marchesini. "QCD coherence in the structure function and associated distributions at small x ★". In: *Nuclear Physics B* 445.1 (1995), pp. 49–78.

[76] Galileo Galilei. *Dialogo sopra i due massimi sistemi del mondo tolemaico e copernicano*. Ed. by Wikimedia. https://it.wikisource.org/wiki/Dialogo_sopra_i_due_massimi_sistemi_del_mondo_tolemaico_e_copernicano. 1632.

[77] Isaac Newton. *Opticks*. Ed. by Wikimedia. https://en.wikisource.org/wiki/Opticks_(4th_Ed). 1704.

[78] J. Clerk Maxwell. "On Faraday's Lines of Force". In: *Transactions of the Cambridge Philosophical Society* 10 (1864), p. 27.

[79] James Clerk Maxwell. *A dynamical theory of the electromagnetic field*. The Society, 1864.

[80] James Clerk Maxwell. *A treatise on electricity and magnetism*. Vol. 1. Clarendon press, 1881.

[81] Hendrik Antoon Lorentz. "Simplified theory of electrical and optical phenomena in moving systems". In: *Koninklijke Nederlandse Akademie van Wetenschappen Proceedings Series B Physical Sciences* 1 (1898), pp. 427–442.

[82] HA Lorentz. "Electromagnetic phenomena in a system moving with any velocity smaller than that of light". In: *KNAW Proceedings*. Vol. 6. 1904, pp. 1903–1904.

[83] A. Einstein. "Zur Elektrodynamik bewegter Körper". In: *Annalen der Physik* 322.10 (1905), pp. 891–921. ISSN: 1521-3889. DOI: 10.1002/andp.19053221004. URL: http://dx.doi.org/10.1002/andp.19053221004.

[84] Max Karl Ernst Ludwig Planck. "Zur theorie des gesetzes der energieverteilung im normalspectrum". In: *Verhandl. Dtsc. Phys. Ges.* 2 (1900), p. 237.

[85] Max Planck. "Zur Theorie der Wärmestrahlung". In: *Annalen der Physik* 336.4 (1910), pp. 758–768.

[86] Max Planck. *Vorlesungen über die Theorie der Wärmestrahlung*. Leipzig, 1921.

[87] A. Einstein. "Über einen die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt". In: *Annalen der Physik* 322.6 (1905), pp. 132–148. ISSN: 1521-3889. DOI: 10.1002/andp.19053220607. URL: http://dx.doi.org/10.1002/andp.19053220607.

[88] N. Bohr. "On the Constitution of Atoms and Molecules". In: *Phil. Mag. Ser.6* 26 (1913), pp. 1–24. DOI: 10.1080/14786441308634955.

[89] N. Bohr. "On the Constitution of Atoms and Molecules. 2. Systems containing only a Single Nucleus". In: *Phil. Mag. Ser.6* 26 (1913), p. 476. DOI: 10.1080/14786441308634993.

[90] Olaf Nairz, Markus Arndt, and Anton Zeilinger. "Quantum interference experiments with large molecules". In: *American Journal of Physics* 71.4 (2003), pp. 319–325. DOI: 10.1119/1.1531580. eprint: https://doi.org/10.1119/1.1531580. URL: https://doi.org/10.1119/1.1531580.

[91] O. Donati, G. P. Missiroli, and G. Pozzi. "An Experiment on Electron Interference". In: *American Journal of Physics* 41.5 (1973), pp. 639–644. DOI: 10.1119/1.1987321. eprint: https://doi.org/10.1119/1.1987321. URL: https://doi.org/10.1119/1.1987321.

[92] E. Schrödinger. "Quantisierung als Eigenwertproblem". In: *Annalen Phys.* 384.4 (1926). [Annalen Phys.79,no.Ser.IV,361(1926)], pp. 361–376. DOI: 10.1002/andp.19263840404.

[93] E. Schrödinger. "Quantisierung als Eigenwertproblem". In: *Annalen Phys.* 385.13 (1926). [Annalen Phys.80,no.Ser.IV,437(1926)], pp. 437–490. DOI: 10.1002/andp.19263851302.

[94] E. Schrodinger. "Der stetige Ubergang von der Mikro- zur Makromechanik". In: *Naturwiss.* 14 (1926), pp. 664–666. DOI: 10.1007/BF01507634.

[95] E. Schrödinger. "Quantisierung als Eigenwertproblem". In: *Annalen Phys.* 384.6 (1926). [Annalen Phys.79,no.Ser.IV,489(1926)], pp. 489–527. DOI: 10.1002/andp.19263840602.

[96] E. Schrödinger. "Quantisierung als Eigenwertproblem". In: *Annalen Phys.* 386.18 (1926). [Annalen Phys.81,no.Ser.IV,109(1926)], pp. 109–139. DOI: 10.1002/andp.19263861802.

[97] C.J. Davisson and L.H. Germer. "Reflection of electrons by a crystal of nickel". In: *Proceedings of the National Academy of Sciences* 14.4 (1928), pp. 317–322.

[98] Dr. Tanamura. *Double-slit experiment*. https://commons.wikimedia.org/wiki/File:Double-slit_experiment_results_Tanamura_2.jpg. Accessed: 2018-01-13.

[99] Walther Gerlach and Otto Stern. "Der experimentelle Nachweis der Richtungsquantelung im Magnetfeld". In: *Zeitschrift für Physik* 9.1 (1922), pp. 349–352.

[100] Walther Gerlach and Otto Stern. "Das magnetische Moment des Silberatoms". In: *Zeitschrift für Physik A Hadrons and Nuclei* 9.1 (1922), pp. 353–355.

[101] Walther Gerlach and Otto Stern. "Der experimentelle Nachweis des magnetischen Moments des Silberatoms". In: *Zeitschrift für Physik A Hadrons and Nuclei* 8.1 (1922), pp. 110–111.

[102] Theresa Knott. *Stern-Gerlach experiment*. https://commons.wikimedia.org/wiki/File:Stern-Gerlach_experiment_svg.svg. Accessed: 2017-12-12.

[103] Paul A. M. Dirac. "The quantum theory of the electron". In: *Proc. Roy. Soc. Lond.* A117 (1928), pp. 610–624. DOI: 10.1098/rspa.1928.0023.

[104] P. A. M. Dirac. "A Theory of Electrons and Protons". In: *Proc. Roy. Soc. Lond.* A126 (1930), p. 360. DOI: 10.1098/rspa.1930.0013.

[105] Carl D. Anderson. "The Positive Electron". In: *Phys. Rev.* 43 (6 Mar. 1933), pp. 491–494. DOI: 10.1103/PhysRev.43.491. URL: https://link.aps.org/doi/10.1103/PhysRev.43.491.

[106] M. Fierz and W. Pauli. "On relativistic wave equations for particles of arbitrary spin in an electromagnetic field". In: *Proc. Roy. Soc. Lond.* A173 (1939), pp. 211–232. DOI: 10.1098/rspa.1939.0140.

[107] James Chadwick. "Possible existence of a neutron". In: *Nature* 129.3252 (1932), p. 312.

[108] M. Breidenbach et al. "Observed Behavior of Highly Inelastic Electron-Proton Scattering". In: *Phys. Rev. Lett.* 23 (16 Oct. 1969), pp. 935–939. DOI: 10.1103/PhysRevLett.23.935. URL: https://link.aps.org/doi/10.1103/PhysRevLett.23.935.

[109] Thomas Lenzi. "Development and Study of Different Muon Track Reconstruction Algorithms for the Level-1 Trigger for the CMS Muon Upgrade with GEM Detectors". MA thesis. U. Brussels (main), 2013. arXiv: 1306.0858 [physics.ins-det]. URL: http://inspirehep.net/record/1236817/files/arXiv:1306.0858.pdf.

[110] Stefan Kühne. *A Mercator projection map with Tissot's indicatrices*. https://en.wikipedia.org/wiki/File:Tissot_mercator.png. Accessed: 2018-01-08.

3

The Large Hadron Collider and the Compact Muon Solenoid

The European Organisation for Nuclear Research (CERN) [1] is an international centre for experimental physics [1]. Originally founded by twelve European countries in 1952 to associate their research programmes, many other countries from all around the world have now joined CERN at various levels, and other topics of research have come up. CERN is nowadays one of the leading centres for research in nuclear and particle physics, for an internal budget of around one billion euros a year.

CERN hosts the largest complex of particle accelerators in the world (illustrated with the diagram in Fig. 3.2), and is therefore a unique place for performing all kinds of experiments involving high-energy beams, from particle physics to meteorology. In particular, it hosts the Large Hadron Collider (LHC), where beams are scattered head on at extremely high energy densities. Such a configuration implies the building of very large and complex detectors, such as the Compact Muon Solenoid (CMS). The present chapter is devoted to their descriptions.

FIGURE 3.1 – CERN logo [2].

In this section, we give a description of LHC and CMS, emphasising on aspects that will matter in the physics analysis.

3.1 Large Hadron Collider

The LHC project was approved in 1992 by the CERN Council, with the intention of succeeding the Large Electron Positron Collider (LEP), in the same tunnel and in the physics programme, especially in the search of the BEH boson [4, 5].

In this section, after a global overview of the project, we give the basic principles of particle acceleration and apply them to the LHC.

1. The acronym stands for the original name in French of the organisation *Conseil Européen pour la Recherche Nucléaire*, and has been kept for the proximity with the Germanic root *kern*, meaning "nucleus".

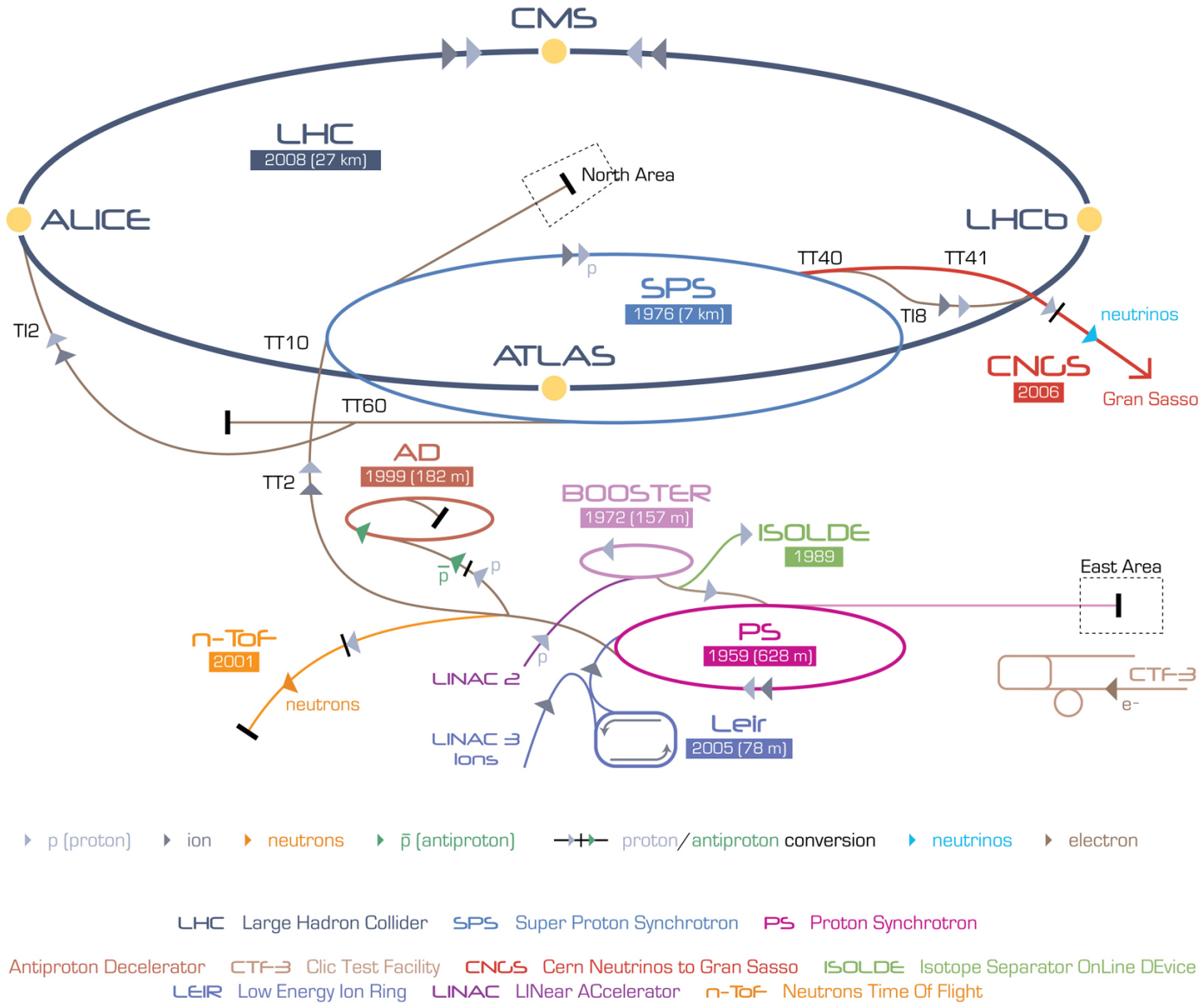
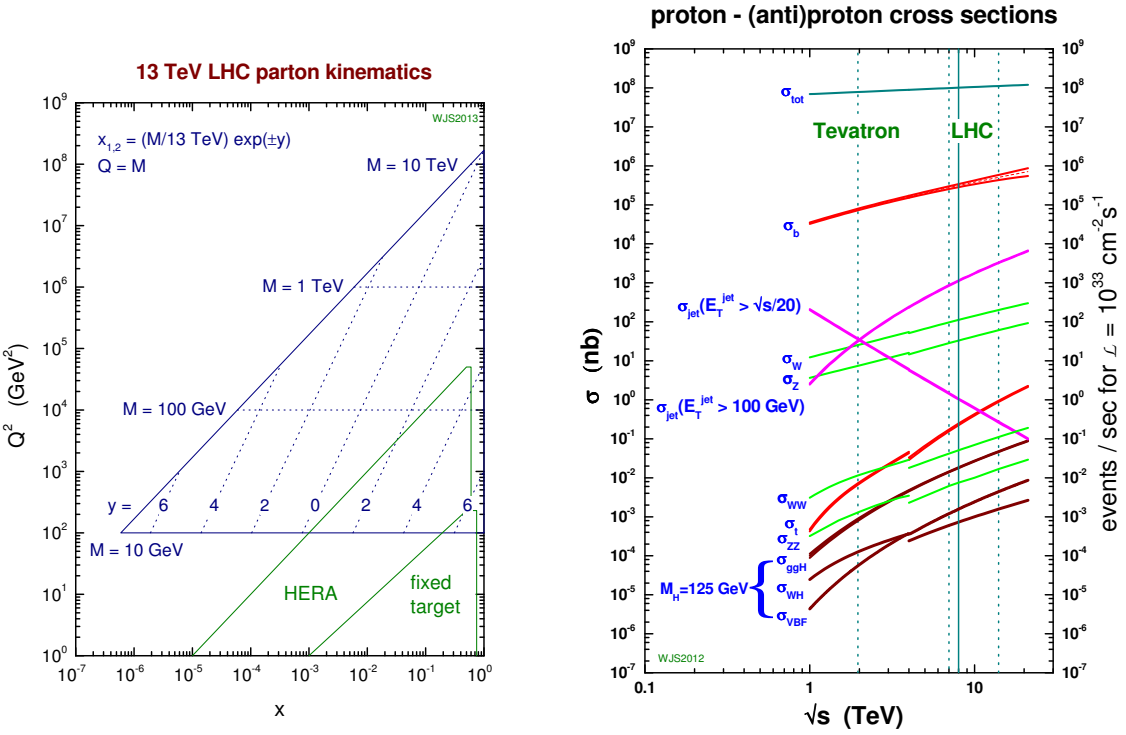


FIGURE 3.2 – CERN accelerator complex [3]. Protons are extracted from a simple bottle of hydrogen based at the LINAC 2. They are collected in the Booster that accelerate them from 50 MeV to 1.4 GeV, then to the PS that speeds them up to 26 GeV and separates the beams in smaller bunches, then to the SPS that makes them go to 450 GeV. And eventually, they can go to the LHC, whose nominal energy per beam in 2016 is of 6.5 TeV. Lead nuclei can be as well accelerated, but start from the LINAC 3 and go through the LEIR.



(A) Phase space in (x, Q^2) coordinates, compared to HERA.

(B) Simulation of the cross sections, compared to TEVATRON.

FIGURE 3.3 – Comparisons of the accessible phase space at LHC compared with previous experiments [6].

3.1.1 Overview

As can be seen from Fig. 3.2, the LHC is at the end of a long chain of accelerators, providing beams of protons. It is built underground to reduce the impact of cosmic rays in measurements. As of 2016, each beam can carry up to 6.5 TeV of energy in the centre-of-mass system of the colliding particles, highest energy in the world. This opens up new regions of the phase space to perform precision measurements — as in the case of the present thesis — or searches — like the search of the *Higgs boson*. Especially, such a high energy allows the production of b jets up to transverse momenta of order from $\mathcal{O}(10\,\text{GeV})$ up to $\mathcal{O}(1\,\text{TeV})$. Fig. 3.3 shows how LHC enlarges the phase space with respect to the HERA and TEVATRON experiments.

LHC stands for Large Hadron Collider:

Collider As its name suggests, the LHC is a *collider*, *i.e.* it accelerates hadrons in opposite directions and makes them collide in flight. Collider experiments are to be opposed to *fixed-target* experiments, where particles are all accelerated in the same direction on a fixed target. Although easier to set up and although providing higher luminosity, the energy in the centre-of-mass system grows as the square root $\sqrt{E}$ of the beam energy E of the incident proton, while it grows as $\sqrt{E_1 E_2}$, the square root of the product of the energies $E_{1,2}$ of the colliding beams. Therefore, colliders are more suited to investigate region of

the phase space of higher energy.

Hadron Furthermore, the LHC collides hadrons. In practice, these are either protons or lead nuclei:

— Protons, since the *proton* is the only stable and charged hadron that can be found in nature, allowing the study of fundamental parameters of the SM or BSM. LHC currently provides proton collisions at 13 TeV in the centre-of-mass frame.

— Lead nuclei, since the *lead* is the heaviest stable nucleus (around 208 times the mass of the proton), allowing the study of *quark-gluon plasma*. LHC currently provides collisions at around 1150 TeV in the centre-of-mass frame, corresponding to around 2.75 TeV per nucleon inside of the lead nucleus.

In this thesis, we shall only discuss proton-proton collisions.

Large Finally, the LHC is a circular collider, i.e. it is made of two superimposed rings in which protons circulate in respectively opposite directions. *Circular colliders* are to be opposed to *linear colliders*.

— In a linear (circular) collider, the accelerator and the bunches can be only once (many times). Yet reusing the same beams may save energy and time.

— However, circular colliders are affected by the *synchrotron radiation*[2] [7], *i.e.* part of the energy that is dedicated to the acceleration of particles is lost in the form of light rays. At each revolution, the loss of energy ΔE the goes as follows:

$$\Delta E = \frac{1}{3}\frac{e^2}{\epsilon_0}\left(\frac{E}{R(mc^2)}\right)^4 \tag{3.1}$$

where

— e is the electric elementary charge,

— ϵ_0 is the permittivity,

— E stands for the energy

— and R stands for the radius of the ring.

For protons at $E = 6.5\,\text{TeV}$, $\Delta E \approx 0.01\,\text{MeV}$. In addition, from this formula, one can also see the advantage of using protons instead of electrons: since protons are much heavier, the radiation is much more suppressed with a factor of $(m_e/m_p)^4 \approx 10^{-13}$.

The circumference of LHC is of 26.659 km, the *large* radius[3] of around 4.3 km allowing to limit the synchrotron radiation; the size of the ring can be compared to the size of the surroundings in Fig. 3.4.

The LHC machine is in constant development. Since the start of data taking in 2008, it has already undergone two *runs*. The 2016 data that we shall analyse later belongs to the second run, a.k.a. LHC Run-II. Further upgrades are intended in order to reach higher beam energies and luminosities for later data-taking periods.

2. Note that in general, any charge that has a non-uniform movement should radiate according to the Maxwell equations; however the radiation in case of a linear collider are usually negligible.

3. Note that the LHC is not perfectly circular, some section being linear.

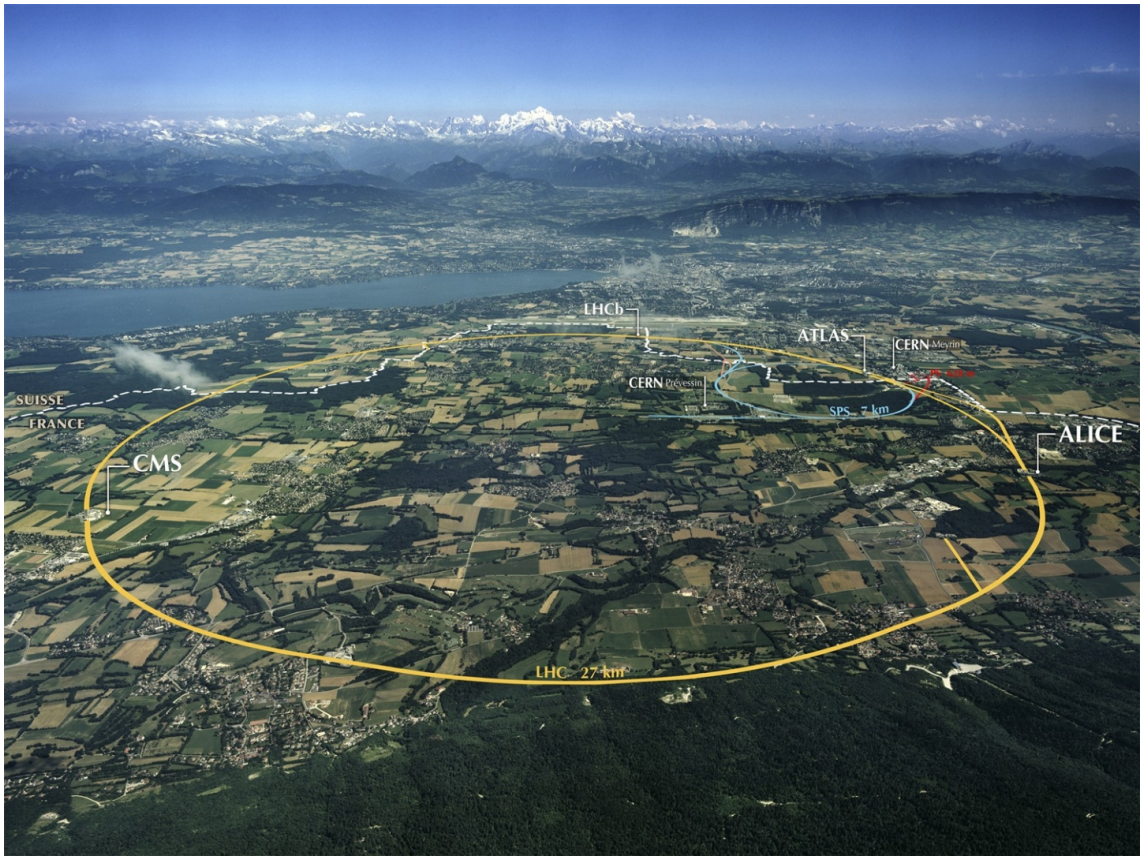


FIGURE 3.4 – Surroundings of CERN and Geneva with the Lake Léman, highlighting the position of LHC underground (yellow line) and SPS (light blue line) [8]. The two campuses of CERN can be seen (Meyrin and Prévessin), as well as the four big experiments (CMS, ATLAS, LHCb and ALICE). It can also be compared to the airport to better appreciate its size.

3.1.2 Principles of acceleration

Some general principles of acceleration are here given [9, 1].

The techniques used to accelerate particles are relativistic but non-quantum-mechanical. Particles are only manipulated through classical electromagnetic interactions.

From the extraction from a bottle of gaseous hydrogen to the collisions, four main aspects related to proton acceleration may be distinguished:

acceleration Given $\mathbf{F} = q\mathbf{E}$, the main technique in use relies on *radiofrequency cavities* (or *resonance cavities*): cavities provide an alternating electric field $\mathbf{E}(t)$, where charges will pass successively. In order to be only sensitive to acceleration phases of the electric fields, the charges stay in cavities only during a half period; during the other half periods, charges travel through tubes with no ambient electric field. The working principle is illustrated in Fig. 3.5. The length of the tubes and the frequency of the alternating electric field depend on the configuration:

— At the Linear Accelerator 2 (LINAC 2), since the protons start from rest, the first tubes are of variable size. Then, since the velocity gets closer and closer to the speed of light, the length of last tubes can be constant.

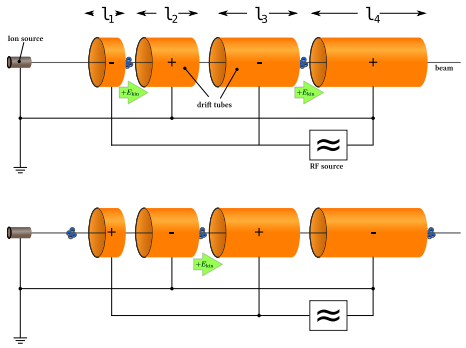


FIGURE 3.5 – Illustration of the working of radiofrequency cavities [10]. Cavities and tubes succeed to one another. Each cavity provides an alternating electric field. Here, the frequency is taken as constant and the length of the tube is adapted such that charges get only accelerated by the successive cavities according to $L = v \times \frac{T}{2}$.

— However, in the case of a circular accelerator made of a ring, such at the LHC, since the same cavities are reused several times, the frequency needs to be varied. An accelerator relying on this principle is called *synchrotron*, such as most of the circular accelerators nowadays [4].

clustering and bunch splitting Particles in the beams are not all synchronous, *i.e.* they are not all perfectly synchronised with the electric fields in the resonance cavities. Early (late) particles will be more (less) accelerated, resulting in longitudinal oscillations of particles.

steering Dipole magnets are in principle sufficient to steer particles. At LHC, the technical difficulty lies in getting strong enough magnets to steer the beams, requiring 8.33 T of magnitude. This has been made possible by the superconducting magnet technology.

focusing Focusing the beams is required for several reasons:

— to keep same-charge particles confined in the beam;

— to control the transverse oscillation of the protons in the beams (*i.e.* protons should stay inside of the pipe);

— to focus strongly the two beams shortly before the collision so as to increase the cross section (see Fig. 3.6).

This may be achieved thanks to pairs of *quadrupole magnets*: one quadrupole magnet focusing in one transverse direction while defocusing in the other, two quadrupole magnets with alternating poles will act as consecutive convergent and divergent lenses in optics.

In addition, all operations on the beams require an very deep vacuum inside of the pipe, in order to avoid collisions with particles in the medium.

Given these elements, and given the diagram of the complex of accelerators at CERN in Fig. 3.2, we can now detail the different phases of acceleration of particles and beams of particles up to the collision at LHC:

LINAC 2 The Linear Accelerator 2 is an injector, *i.e.* it is only designed to provide protons to other, more powerful accelerators. Its rôles are:

4. By opposition, the first circular accelerators were no synchrotrons but *cyclotrons*, made of a disk rather than a ring, with constant frequency but variable curvature of particles.

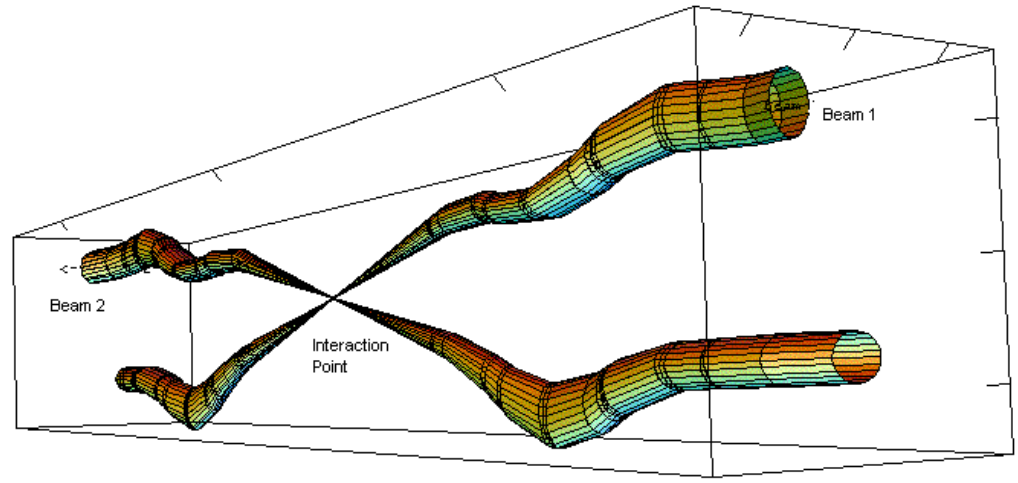


FIGURE 3.6 – Increase of the density of the beams while approaching the IP1 (ATLAS) [11]. Note that such a focus could be achieved continuously, since it comes together with an increase of the amplitude of oscillation of the particles inside of the beam.

— extraction of protons from bottle of hydrogen,

— acceleration from rest to 50 MeV (for a speed around $0.3c$),

— continuous beam production.

Booster The Proton Synchrotron Booster [12] is made of four superimposed rings for different functions; each ring may accelerate one or two bunches from the injection of protons from the LINAC 2. In its rôle of injector for LHC [5], its functions are:

— acceleration from 50 MeV to 1.4 GeV,

— accumulation of protons from LINAC 2 (or LINAC 3 for heavy nuclei),

— production of six or eight bunches (different schemes are possible).

PS The Proton Synchrotron is the oldest accelerator still in use at CERN. It is also the very first synchrotron at CERN, with a length of 628 m. After a period dedicated to physics in the fifties, it underwent several upgrades and is now only used as an acceleration facility of various kinds of particles to different machines. In the chain of proton acceleration to LHC, getting bunches from the Booster, it is used for:

— acceleration from 1.4 GeV to 24 GeV,

— bunch splitting (typically from 6 to 72 bunches [13]).

SPS The Super Proton Synchrotron was the successor of the PS, with nearly 7 km of circumference, and has also been recycled as an accelerating facility:

— acceleration from 24 GeV to 450 GeV.

LHC Finally, protons reach the LHC, which will take care of the following:

— acceleration from 450 GeV to (as of 2016) 6.5 TeV,

5. It is interesting to keep in mind that most protons accelerated at CERN are not intended to be delivered at LHC. The Booster has various working modes and deliver beams of different flavours. In fact, the experience requiring most of the protons at CERN is ISOLDE.

— collision of beams at CMS and other experiments.

Once the beams have reached their cruising speed, their bunches measure approximately as long as a knitting needle and follow one another with a distance around 7.5 m. Collisions may take place at the different interaction points (IPs), corresponding to the different experiments.

3.1.3 Collisions

In normal running conditions, crossing bunches typically provide a pile-up of a few tens of collisions. The term *pile-up* generally refers to the fact that more than one *pp* collision may occur at the same time in a single bunch crossing [6]; it will be discussed in more detail in Sect. 7.2 while performing the analysis. The pile-up is the price to pay to increase of the rate of interesting processing without degrading too much the quality of the recorded data.

The *total cross section* [7] σ_{tot} at LHC at $\sqrt{s} = 13\,\text{TeV}$ is around 100 mb [15, 16]. It can be distinguished between the *elastic cross section* (around 30 mb) and *inelastic cross section* (around 70 mb).

Since the total cross section defines the collision rate, one defines the *luminosity* of an accelerator to relate it to the number of delivered collisions:

$$N_{\text{tot}} = \sigma_{\text{tot}} \times L \tag{3.2}$$

It characterises the performance of the accelerator. This formula can be applied to any process, *i.e.* to any cross section:

$$N = \sigma \times L_{\text{int}} \tag{3.3}$$

The luminosity itself is independent of the process, and is only related to the amount of delivered collisions.

In practice, one defines the *instantaneous luminosity* and the *integrated luminosity*.

Circular colliders with identical beams deliver a luminosity according to the following formula:

$$L_{\text{inst}} = fn\frac{N^2}{A} \tag{3.4}$$

where

— f is the revolution frequency (around 11 kHz),

— n the number of *bunches* in the ring (around 2800),

— N the average number of particles in one bunch (around 10^{11})

— and A the *cross-sectional area* [8] of the beams (around $10^{-5}\,\text{cm}^2$).

6. "Pile-up occurs when the readout of a particle detector includes information from more than one event. This specifically refers to cases where there are other (background) collisions somewhere within a timing window around the signal collision." [14]

7. Note that in principle, given the Coulomb interaction, the total cross section is divergent. The total cross section accounts therefore for nuclear interactions.

8. The cross-sectional area can be further described as follows:

$$A = \pi\frac{\epsilon_n\beta^*}{\gamma}F \tag{3.5}$$

where

This quantity is characteristic of the production of collisions by the machine per unit of time and is called the *instantaneous luminosity*. The smallest unit of time on which the luminosity is measured, is called Lumi Section (LS), corresponding to 2^{20} orbits (around 93 s) [14]. For LHC Run-II, the instantaneous and integrated luminosities are shown in Fig. 3.7.

In the regard of a physics analysis, only the *integrated luminosity* is given, since only the total amount of data matters to describe the statistics [18]. The total luminosity of the data with which we shall work is:

$$L_{\text{int}} = \int L_{\text{inst}}\, \mathrm{d}t = 35.2 \pm 0.8\,\text{fb}^{-1} \tag{3.6}$$

Note that this value does not match with the recorded luminosity of $37.76\,\text{fb}^{-1}$ from Fig. 3.7: indeed, while and after recording, each run is carefully monitored and only *certified* runs can be considered in physics analyses. Moreover, some special runs are dedicated to specific analyses, like the measurement of the Minimum Bias (MB) cross section, which requires low-pile-up conditions.

Since the cross-sectional area A varies with time (see Fig. 3.8), the beam luminosity decreases with time. At the LHC, a *run*[9] corresponds to the time that a beam stays in the pipe; runs are of variable luminosities and durations, but can reach a total luminosity of $\mathcal{O}(100\,\text{pb}^{-1})$ and last twelve hours. The smallest unit of count for the instantaneous luminosity is called LS, corresponding to $\sim 2 \times 10^{20}$ orbits (around 93 s) [14].

3.2 Compact Muon Solenoid

The CMS detector [20] is one of the four main[10] detectors at LHC: CMS, ATLAS [25] (A Toroidal LHC ApparatuS), LHCb [26] (LHC beauty) and ALICE [27] (A Large Ion Collider Experiment). Like ATLAS, it is a general-purpose detector, *i.e.* it has been built in order to investigate various aspects of HEP (thus corroborating findings); it is to be opposed to LHCb — which was built mainly to investigate heavy-flavour physics and CP violation — or to ALICE — which was built mainly to investigate heavy-ion collisions and in particular the *quark-gluon plasma* state (see Sect. 2.2.5.1).

The acronym of CMS stands for Compact Muon Solenoid:

— ϵ_n is the *normalised transverse beam emittance*, which is representative of the intrinsic dispersion of the beam depending only on the initial conditions,

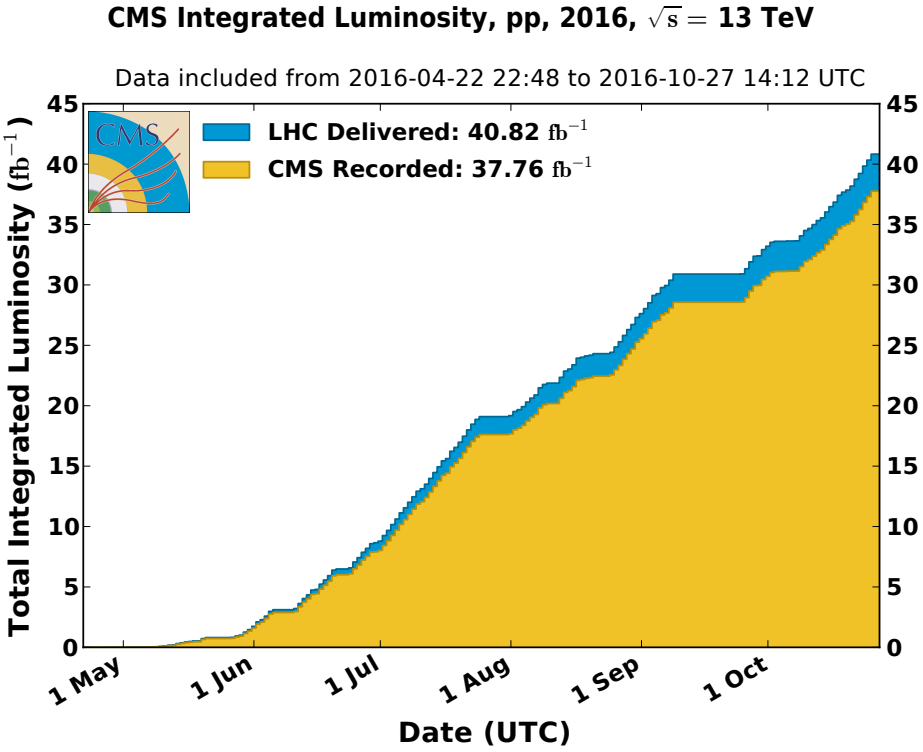
— β^* is the *amplitude function* at the IP, which is representative of the power of focusing of the magnets,

— γ is the relativistic factor,

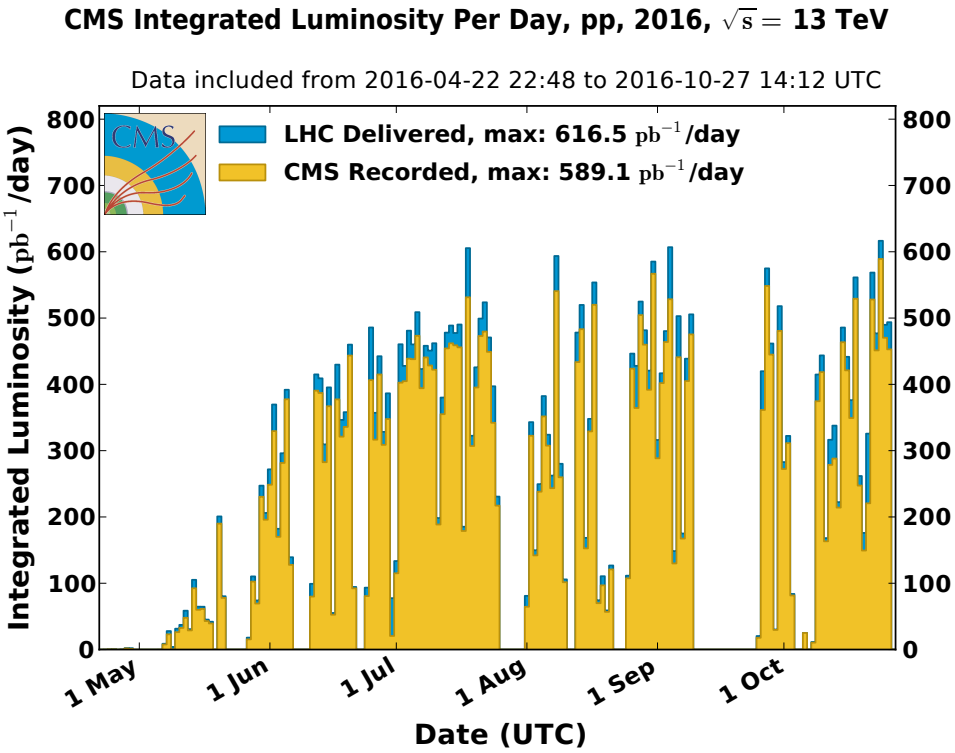
— and F is correction coming from the fact the beams do not collide exactly head on, but with a small angle of 300μrad.

9. The utilisation of the word "run" may be quite confusing, since it is used at three different scales: 1. it stands for long periods of data taking such at LHC Run-I and LHC Run-II (usually written with a capital letter and followed by a Roman number); 2. it stands for variable periods of data taking corresponding to the equivalent of a few weeks of data taking (in this context, also called *era*, and is usually followed by the year of data taking plus some capital letter(s)); 3. finally, as in this section, it corresponds to the data taken from a single pair of beams (corresponding to a 6-digit number).

10. There are a few additional experiments of smaller size: TOTEM [21], LHCf [22], MoEDAL [23], CASTOR [24], ...



(A) "Cumulative offline luminosity versus day delivered to (blue), and recorded by CMS (orange) during stable beams and for *pp* collisions at 13- TeV centre-of-mass energy in 2016. The delivered luminosity accounts for the luminosity delivered from the start of stable beams until the LHC requests CMS to turn off the sensitive detectors to allow a beam dump or beam studies. Given is the luminosity as determined from counting rates measured by the luminosity detectors after offline validation."



(B) "Same as the above, but not cumulative."

FIGURE 3.7 – Integrated luminosity at CMS and LHC in 2016 for proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$ (plots and captions taken from [17]).

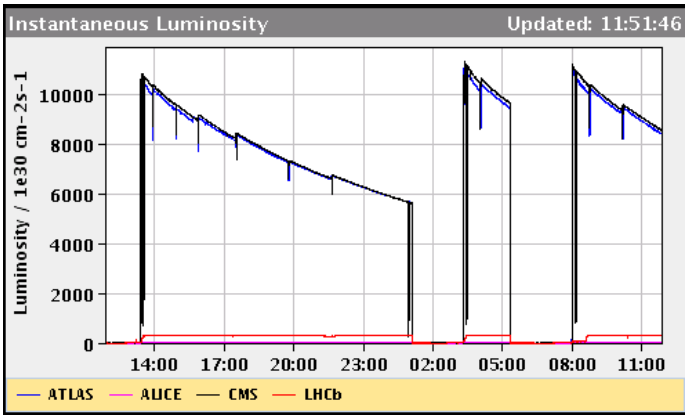


FIGURE 3.8 – Screenshot of instantaneous luminosity (taken on 8 September 2017) [19] delivered to the different experiments at LHC. A beam can last up to twelve hours; afterward, the luminosity gets divided by two with respect to its original value, and the beams are dumped and replaced.

Compact means that most of the devices are placed inside the magnet, the trackers and the calorimeters except for the muon chambers (building such a large magnet is a technical achievement);

Muon means that the detector has been designed to be very sensitive to the muons, as they are involved in some important processes;

Solenoid stands by opposition to *toroidal*, which is one of the two possible geometries for the magnetic field to be parallel to the beams in order to act only on the produced particles [11].

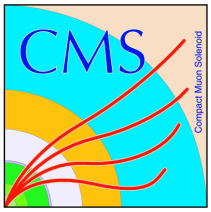


FIGURE 3.9 – Logo of the CMS collaboration. The different layers represent the subdetectors and the four tracks represent muons crossing the detector.

As the name suggests, CMS is designed for particles scatterings in the *c*entre-of-*m*ass *s*ystem (even though it is not the case when studying proton-lead collisions).

The technical proposal [28] was written in 1994 and the construction began in 1997. A picture of the CMS detector is shown in Fig. 3.10 and a few key figures are given in Tab. 3.1.

The design of the detector was developed according to the following points [20]:

1. good muon identification up to 1 TeV and good resolution on the mass of dimuons (see muon chambers in Sect. 3.2.2.3);
2. good tracking and vertexing [12], especially regarding the tagging of b jets and tauons (see tracker system in Sect. 3.2.2.1);
3. good resolution on electromagnetic energy and on the mass of diphotons and dielectrons (see ECAL in Sect. 3.2.2.2);
4. good resolution on Missing Transverse Energy (MET) and on the mass of dijets (see HCAL in Sect. 3.2.2.2).

11. Alternatively, ATLAS was designed with a toroidal configuration.
12. In the context of reconstruction, a vertex corresponds to the crossing of several tracks, usually corresponding to the decay of a particle into several other particles.

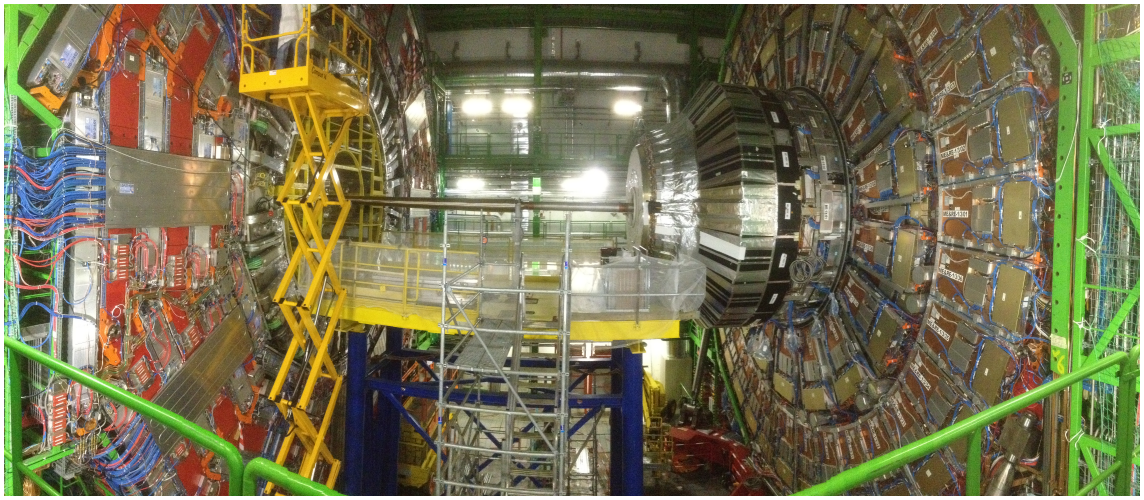


FIGURE 3.10 – Panoramic picture of the CMS detector, taken in January 2017. During this shutdown period, the detector was opened in order to check and repair some components; one can see the beam pipe in the middle, and on the RHS (LHS) the *forward* (*barrel*) parts of the detector.

CMS	A380
100 m underground	1.25× of the length
21 m long	0.3× fuselage size
15 m diameter	same height
14000 t	28 airplanes
5000 persons	10 airplanes

(A) Comparison to an A380 airplane.

category	number
active people	5250
— staff physicists	1963
— physics doctoral students	922
— undergraduates	994
— engineers	995
— technicians	279
— other	97
institutes	198
countries & regions	45

(B) CMS in numbers, at the time of writing the thesis.

TABLE 3.1 – A few key figures comparing CMS to an A380 airplane (see also [1]).

All these points are crucial to the analysis presented in this thesis:

— the identification of secondary vertices, which may indicate the decay of a B hadron, mostly relies on the tracker;

— the identification of electrons and muons plays a rôle in b tagging, since a non-isolated soft lepton may also be the sign of the decay of a B hadron;

— the electromagnetic and hadronic calorimeters are essential in the reconstruction of jets.

In this thesis, we shall focus on issues related to heavy flavour rather than jet reconstruction.

In the next subsections, we give some general principles of reconstruction; then we explain how these are applied at CMS by reviewing its different subdetectors; eventually, we describe how (and when) an event is reconstructed.

3.2.1 General principles of detection and reconstruction

We give here general principles of particle detection [29] and their specific application at CMS for the event reconstruction.

The interaction of particles with media are considered as classical. An event is described by the list of the particles, their four-momenta (see Fig. 2.13a for the conventional coordinate system) of the different processes that have happened at the IP.

Most of the particles created at the IP do not live long enough to be directly detected. As already addressed in Sect. 2.3.2.1, only the proton and the neutron can be considered as stable hadrons, whereas all other hadrons are expected to decay; the decay lengths of hadrons have been summarised in Tab. 2.5). Eventually, in the detector, one can find the following particles (and their antiparticles if relevant):

- photons (γ),
- electrons (e),
- muons (μ),
- charged pions ($\pi^{\pm}$),
- charged kaons ($K^{\pm}$),
- neutral kaons (K^{0}),
- protons (p),
- neutrons (n).

Neutrinos do not decay but interact too weakly to be detected in collider experiments; their presence will be estimated using the missing energy.

Those particles' interactions with media are very well-known:

Photons
- *photoelectric effect*,
- *Compton effect*
- and *pair production* by interaction with nuclei.

In principle, all types of interaction can take place, but at the energy scale of incoming photons in an event at CMS, the pair production is more significant, down to $\sim 10\,\text{MeV}$.

Electrons and muons
- *Bremsstrahlung*,
- *ionisation*
- and *multiple scattering*.

For the detection of electrons (muons), the Bremsstrahlung (ionisation) is the most significant interaction with the media down to $\sim 10\,\text{MeV}$.

Protons, charged pions and charged kaons Regarding the electromagnetic interactions, they are similar to the muons. However, they mainly interact by nuclear interactions.

Neutrons, neutral kaons Only have nuclear interactions.

In addition, mesons may also decay in flight: charged pions may decay weakly into muons for instance.

In practice, except for muons that continue through the magnet and through the *muon chambers* and neutrinos that interact too weakly to be seen, all particles produced at the IP go through the *tracker*, and should be stopped in one of the *calorimeters*:

tracker The aim is to reconstruct the trajectory of all charged particles coming from the IP. Their trajectories are curved thanks to the magnetic

field, and the curvature of a trajectory can be related to its transverse momentum, according to the following relation:

$$\frac{p_T}{\text{GeV}} = 0.3 \frac{B}{\text{T}} \frac{\rho}{\text{m}} \tag{3.7}$$

where ρ is the curvature radius of the charged particle; in 3-dimensional space, this corresponds to a helical motion [13].

calorimeter The principle of a calorimeter is to stop all incident particles (except muons and neutrinos) and measure their energy deposits. If possible, a deposit will be associated to one of the charged particles seen by the tracker at reconstruction (in the tracker acceptance).

muon chambers Muon chambers are a kind of external tracker designed especially for muons, the largest component of the CMS detector.

Given these elements, we can now give a description of the components of the CMS detector.

3.2.2 Application at CMS

A transversal view of the CMS detector can be seen in Fig. 3.11, and the pseudorapidity coverage is given in Tab. 3.2 according to the different subdetectors. From now on, the discussion will be specific to CMS, especially focusing on the parts of the detector that need to be covered to apply b-tagging techniques. Since b-tagging requires all components of the detector, it is limited to the acceptance of muon chambers, *i.e.* $|\eta| < 2.4$.

subdetector	coverage
tracker	$\|\eta\| < 2.5$
ECAL	$\|\eta\| < 3.0$
HCAL	$\|\eta\| < 5.2$
muon chambers	$\|\eta\| < 2.4$

TABLE 3.2 – Pseudorapidity coverage of the different subdetectors at CMS.

13. In general, the motion of a particle in an electromagnetic field is given by *Lorentz force*:

$$\mathbf{F} = q\,(\mathbf{E} + \mathbf{v} \times \mathbf{B}) \tag{3.8}$$

Here,

— the electric field $\mathbf{E}$ is null,

— the magnetic field $\mathbf{B}$ is constant and parallel to the axis,

— and the motion is supposed to be circular under the assumption that the crossing of the tracker is negligible.

Therefore, the transverse momentum can be obtained as follows:

$$m\frac{v_T^2}{\rho} = q v_T B \quad \rightarrow \quad p_T = qB\rho \tag{3.9}$$

Finally, the modulus of the momentum is simply obtained with trigonometry.

$$p = \frac{p_T}{\sin\theta} \tag{3.10}$$

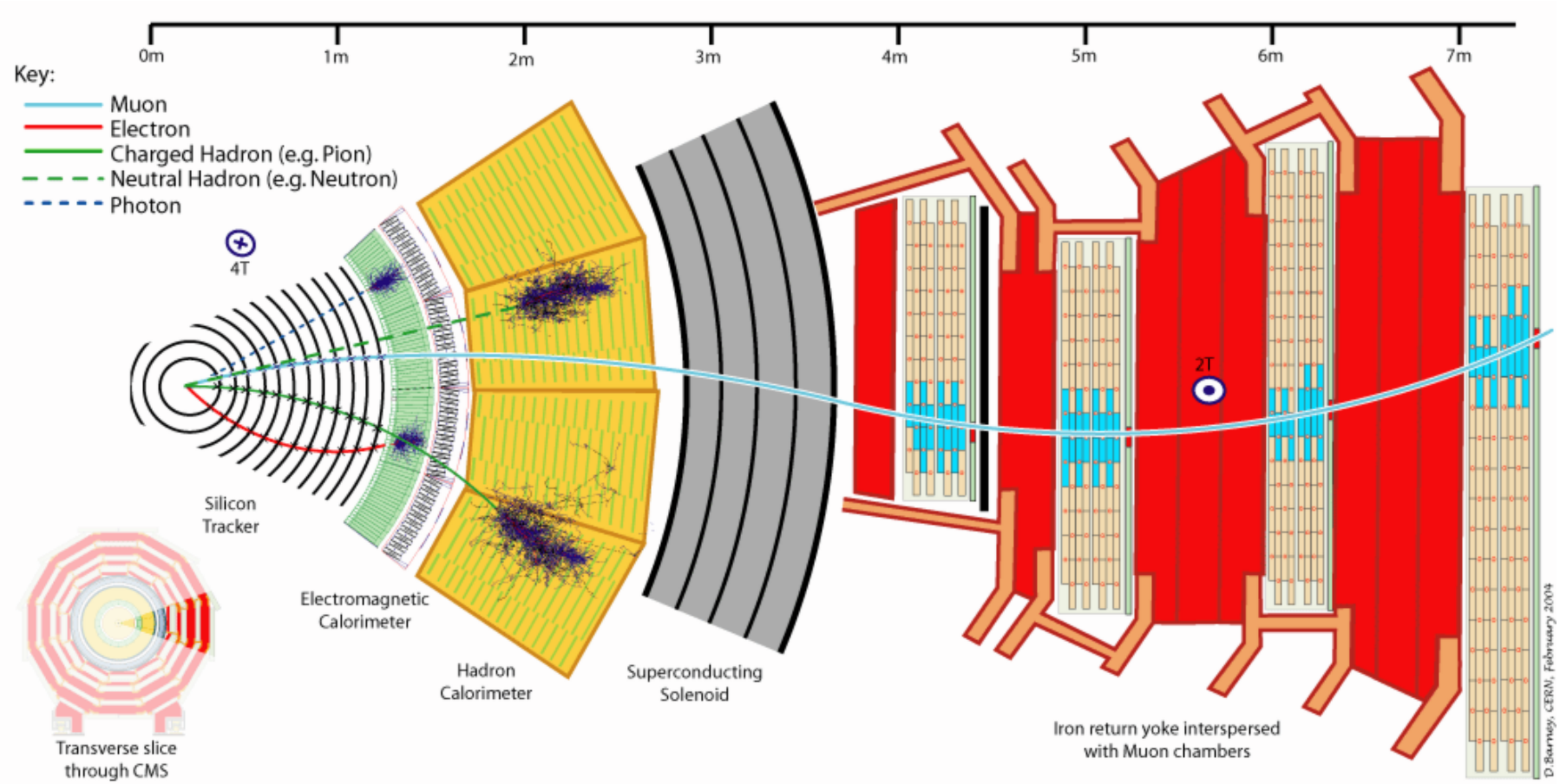


FIGURE 3.11 – Diagram of the barrel of the detector (end-caps have a similar structure) [30]. The ideal trajectories of the different types of particles have been drawn (see colour code in top right). A particle coming from the IP first goes through the tracker; it is seen only if it is charged. Then it goes through the electromagnetic calorimeters, which should only stop photons and electrons. If it has not been absorbed yet, the particle continues through the hadronic calorimeter, which should stop all hadrons. Only muons or neutrinos should still survive these steps, going then through the magnet and the steel structure, and through the muon chambers. There, muons will be detected while neutrinos will continue.

3.2.2.1 Tracking and vertexing

Overview. As already mentioned, the tracker's rôle consists in reconstructing the tracks of outgoing charged particles; especially, thanks to the magnetic field, it allows to determine their transverse momenta. At CMS, the tracker itself is made of *semi-conductor modules*, assembled in layers; the magnet is made of superconducting niobium-titanium, providing an homogeneous magnetic field of 3.8 T around the beam pipe. The interest of the semi-conductor technology is that the valence band of the electrons is just below the conduction band; the excitation of valence electrons makes them jump in the conduction band. Therefore, when a high-energetic charged particle goes through a module, a signal is induced, called a *hit*; combining hits, one can in principle reconstruct the tracks of all charged particles crossing the tracker, and consequently identify vertices.

Description. The semi-conductor modules are made of *npn*-doped junctions in silicon. At CMS, one distinguishes to types of modules:

silicon strip tracker From $R = 55\,\text{cm}$ to $R = 110\,\text{cm}$, the tracker consists of layers of doped semi-conductor detectors. The strips measure $25\,\text{cm} \times 180\,\mu\text{m}$ and are arranged in stereo to get the two components of the coordinate.

pixel tracker Closer to the beam pipe, from $R = 20\,\text{cm}$ to $R = 55\,\text{cm}$, pixel cells are used in order to provide high accuracy. Pixel cells measure $100 \times 150\,\mu\text{m}^2$ to be as precise as the silicon strip tracker; the principles of working and detection are the same. The pixel cells must be made of materials resisting to the important radiations; however, significant degradations may be observed during the data-taking periods and continuous calibration is needed.

Pictures of pixel and strip modules are shown in Fig. 3.12. The pixel modules provide a finer resolution than the strip modules, as described in Tab. 3.3.

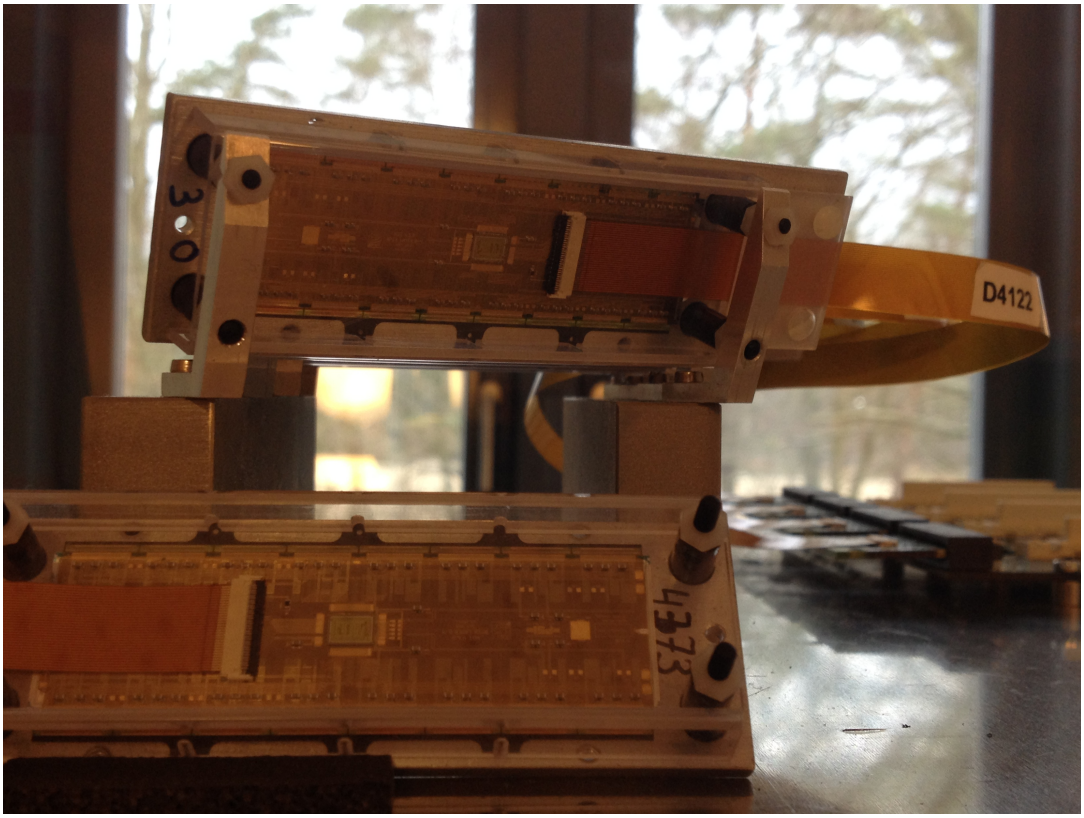
	units	hit resolution
pixel	1440	$9\,\mu\text{m}$
strip	15 148	$20 - 60\,\mu\text{m}$

TABLE 3.3 – Description of the pixel and strip modules.

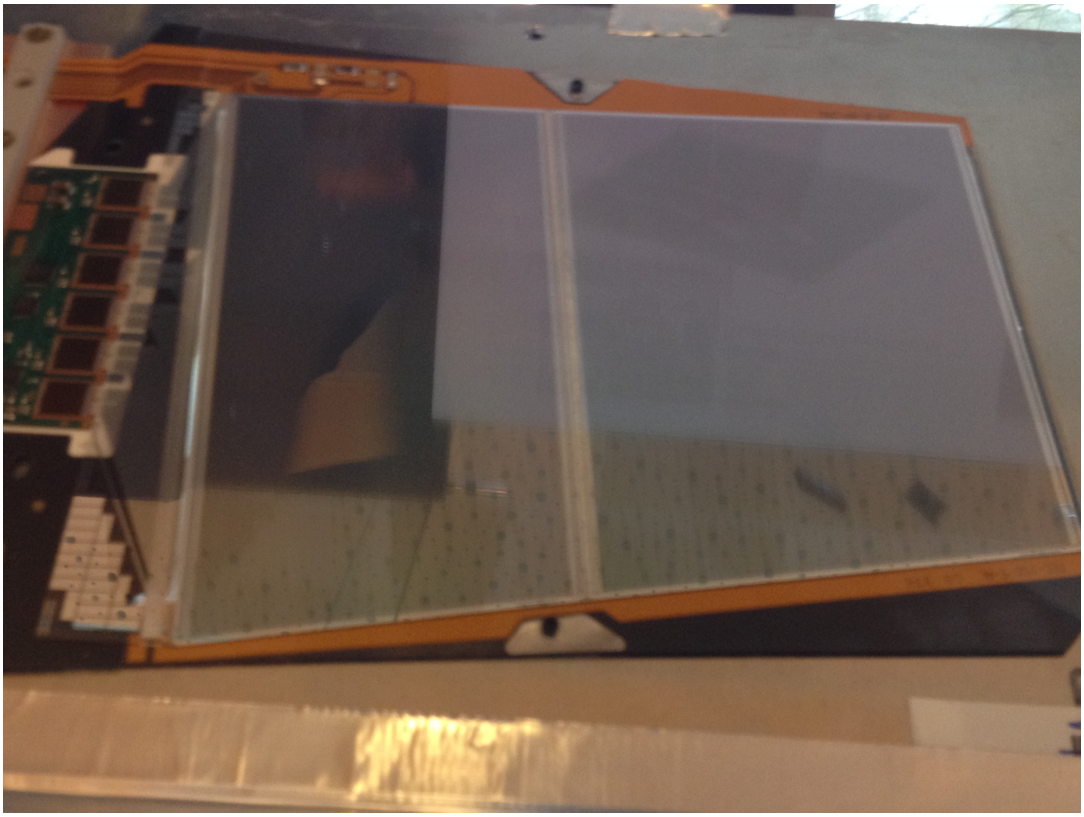
Track reconstruction. By combining tracks in successive layers, the tracking system allows to reconstruct tracks of charged particles; these can be later associated to Primary Vertices (PVs) or Secondary Vertices (SVs). At CMS, the tracking algorithm [31] is based on the *Kalman filter* [32]; it is an iterative procedure, applied layer by layer, as shown in Fig. 3.13:

1. On the first layer, one can only measure the position ; the momentum is arbitrarily set to 5 GeV, with a very large uncertainty.
2. On the next layer, the position can be both a) determined experimentally and b) predicted by extrapolation from the previous layer(s) with the Equations of Motions. The *filtered position* is then determined as a compromise between the prediction and the measurement. The balance accounts for the respective uncertainties.
3. The track parameters are updated with the last filtered position, and the position on the next layer is predicted. Item 2 is repeated till the last layer.

Many refinements exist:



(A) Two pixel modules. One sees here the read-out chips. The sensors are on the other side of the modules.



(B) A strip module. One can see the read-out chip on the left, and the two sensors (the reflecting surfaces) of the strip module.

FIGURE 3.12 – Silicon modules as of 2016 (picture taken at DESY in February 2018).

— The filter can be tuned for noisy environment, where many tracks have crossed the tracker. In particular, at CMS, a preselection of candidate tracks is considered before applying the reconstruction, in order to reduce the combinatorics.

— It can also account for inefficiencies (*i.e.* for missing hits) and for the thickness of the material (*i.e.* for multiple scattering).

— The filter can be applied twice to reduce sensitivity to the 5-GeV seed: once starting from the innermost layer, once starting from the outermost layer. It can also be reiterated taking into account the vertex to which it belongs, once the vertices have been reconstructed.

At CMS, each track is typically reconstructed with 20 hits in the tracker, and is efficient down to 1 GeV.

Vertex reconstruction. Given the reconstructed tracks, one can reconstruct PVs and SVs. In practice, in normal run conditions, a few tens of vertices are expected per bunch crossing. The reconstruction of vertices is another problem of pattern recognition. The adopted strategy at CMS is called Adaptive Vertex Reconstruction (AVR) [31]; given the complexity of the problem of vertexing, it can here only be outlined:

1. Based on the position with respect to the beam pipe (z), candidate vertices are proposed by the *deterministic annealing* technique [33]. The deterministic annealing technique treats the problem in analogy with statistical mechanics, defining a free energy and a temperature; starting from only one cluster of tracks, the temperature is gradually lowered, a phase transition corresponding then to the splitting of a cluster. The difficulty lies in finding the right balance between clustering a variable number of tracks in a variable number of vertices, avoiding as much as possible to split a genuine interaction into two clusters of tracks. In this work, only tracks with $z < 24\,\text{cm}$ have been considered.
2. Based on the 3D coordinates (x, y and z), an Adaptive Vertex Fitter (AVF) [34] is applied to determine the best estimates of the vertex configuration from all track parameters. It consists in a modified, more robust Kalman filter, where all tracks are weighted according to their compatibility with the vertex; in particular, outliers are downweighted.

This method is used to determine all vertices. SVs are selected among all vertices with additional requirements, depending on the analysis; another vertex fitter, the Inclusive Vertex Fitter (IVF), is sometimes used [35].

Resolution. The resolution on the momentum (Eq. 3.10) is essentially limited by two factors [36]:

curvature The higher the transverse momentum, the straighter the curve. The tendency increases with the transverse momentum:

$$\frac{\sigma_p}{p} \propto p_T \tag{3.11}$$

multiple scattering The lower the momentum, the more sensitive to multiple scattering. The effect is almost constant with respect to transverse momentum but becomes significant when the curvature becomes smaller:

$$\frac{\sigma_p}{p} \propto \frac{1}{\sin\theta} \tag{3.12}$$

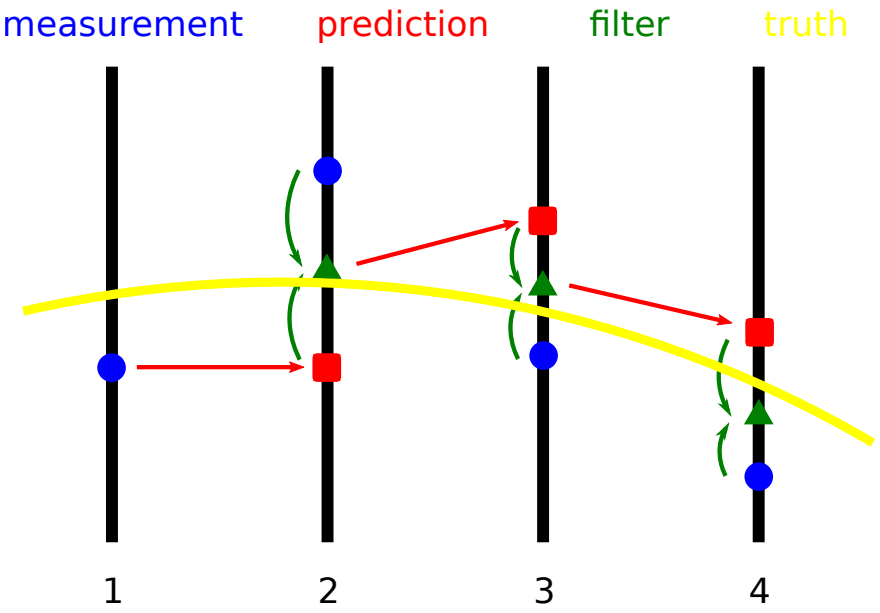


FIGURE 3.13 – Illustration of Kalman filter in a simplistic case with 4-layer tracker: the iterative procedure is performed layer by layer. The layers are represented transversally in black; the true track is represented with the bended yellow curve. The position on the second layer is predicted from the position on the first layer (red arrow); then a compromise — the *filter* (green triangles) — is found (green arrows) between the measurement (blue points) and the prediction (red squares). The prediction for the position on the third layer is performed taking into account the kinematics from the two first layers, etc.

Alignment. The tracker needs to be aligned for two reasons:

1. The precision of mounting of the modules at assembly is lower than their hit resolution.
2. A systematic misalignment may introduce a systematic bias in the reconstruction, which may then lead in a systematic bias in any measurement.

Therefore the resolution can be improved by aligning the modules. Details on alignment procedure may be found in App. A.

Calibration. A high-voltage tension of a few hundreds Volts is used to control the sensitivity of the modules, and needs to be adjusted in order to get a uniform sensitivity throughout the whole volume of the tracker. Moreover, due to the high radiation environment during data taking, the sensitivity region of the modules reduces significantly with time; to counter-balance the ageing of the modules, the high-voltage tension needs to be readjusted regularly. Additional details on the calibration of the modules may also be found in App. A.

3.2.2.2 Calorimetry

Overview. While the principle of a tracker is to measure the momenta of the charged particles without significant interaction, the principle of a calorimeter is to measure the total energy by stopping completely charged and neutral particles (only muons and neutrinos are not affected). Incoming high-energetic particles provoke cascades of particles of smaller energy. At CMS, one distinguishes two calorimeters: the Electromagnetic CALorimeter

	Coverage			Crystals	
	Rapidity ($\lvert\eta\rvert$)	Radius (r/m)	Position ($\lvert z\rvert$/m)	Dimensions (V/cm^3))	Number
EB	$0 - 1.48$	$1.24 - 1.75$	$0 - 3.0$	$2.2 \times 2.2 \times 23$	61 200
EE	$1.48 - 3.0$	$0.32 - 1.71$	$3.2 - 3.9$	$2.86 \times 2.86 \times 22$	2×7324

(A) ECAL

	Coverage			Towers	
	Rapidity ($\lvert\eta\rvert$)	Radius (r/m)	Position ($\lvert z\rvert$/m)	Coverage ($\Delta\eta \times \Delta\phi$)	Number
HB	$0 - 1.4$	$1.8 - 2.9$	$0 - 4.0$	$0.087 \times 10°$	2304
HO	$0 - 1.26$	$3.85 - 4.1$	$0 - 2.5$	$0.087 \times 30°$	
HE	$1.4 - 3.0$	$0.45 - 2.9$	$3.9 - 5.7$	$0.09 - 0.35 \times 5 - 10°$	2304
HF	$3.0 - 5.2$	$0.15 - 1.3$	$11.2 - 12.85$	$0.1 - 0.3 \times 10 - 20°$	1800

(B) HCAL

TABLE 3.4 – Geometry of the calorimeters at CMS [37]. The numbers may sometimes be rounded up for more readability.

(ECAL) and the Hadronic CALorimeter (HCAL); the former (latter) is more suited to the detection of photons and electrons (hadrons).

Description. The ECAL and HCAL [37] are (almost) entirely placed between the tracker system and the magnet, as shown in Fig. 3.14. However, they have different structures and are made of different materials:

ECAL The ECAL is an *homogeneous calorimeter*, which means that the entire volume is used to collect the signal, situated between the tracking system and the HCAL. According to the region of the detector, one distinguishes the ECAL Barrel (EB) and the ECAL End-caps (EE); the geometry of the ECAL is given in Tab. 3.4a. Both regions are made of a lead tungstate crystals with photodetector glued onto the back. The high density (8.3 g/cm^3, each crystal weigh around 1.5 kg) make the electrons and positrons radiate, and make the photons produce electron-positron pairs. The scintillator medium is transparent for photons of a certain wavelength that is subsequently collected by photodetectors. In addition to the EB and EE, a *pre-shower* detector is installed in $|\eta| < 0.9$ and $1.65 < |\eta| < 2.61$ in order to help distinguishing pions and photons.

HCAL The HCAL exploits the nuclear interactions of the hadrons to produce lighter hadrons and photons. It consists in a *sampling calorimeter*, structured in *towers* made of alternate layers of absorbers and scintillators, the former (latter) being used to slow down the hadrons (collect the produced photons). The HCAL consists in four regions: the HCAL Barrel (HB), the HCAL End-caps (HE), the HCAL Forward (HF) and the HCAL Outer (HO); the geometry of the HCAL is given in Tab. 3.4b. It is situated outside of the ECAL and inside of the magnet, except the HO, placed outside of the magnet together with the muon chambers.

The high granularity of both calorimeters is crucial to perform a fine description of the jet substructures, and to determine the contamination from the pile-up activity. It also implies that the alignment of the crystals with one another and with the tracker is crucial; however, this topic will not be treated here.

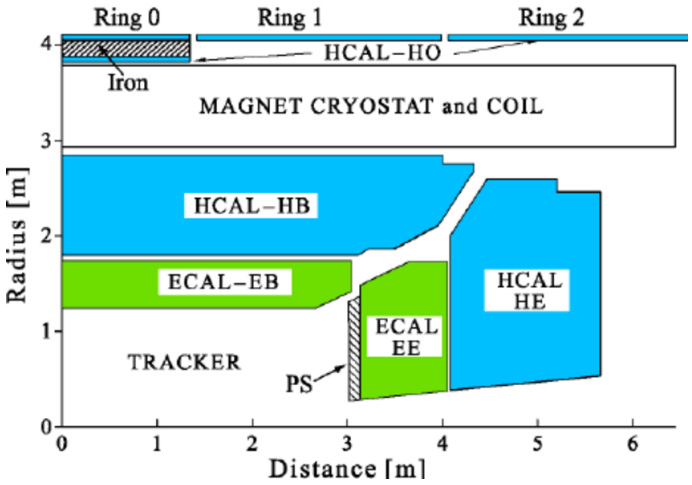


FIGURE 3.14 – Diagram showing the longitudinal view of the calorimeters and of the magnet [38]. The blue (green) region stands for the ECAL (HCAL). Only the HF is not represented. The rings correspond to the mechanical structures holding the whole detector.

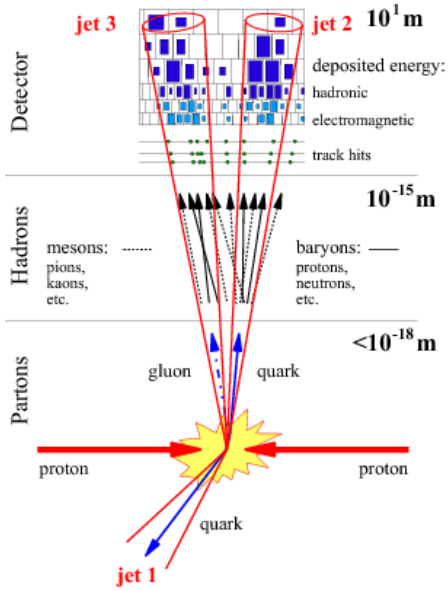


FIGURE 3.15 – Sketch of the jet reconstruction [39]: first the collision takes place and partons are produced in collimated regions of the detector; after hadronisation, the hadrons leave go through the detector, going through the tracker and through the calorimeters (energy deposit in calorimeters).

Reconstruction. The energy deposits are collected with low-energy photons in scintillators, using Compton effect. Deposits in the calorimeter will be compared to tracks in the tracker and in the muon chambers, distinguishing thus deposits from charged particles and deposits from neutral particles. Jets are then defined by clustering the particles[14], following one of the algorithms described in Sect. 2.3.2.2; this is illustrated in Fig. 3.15.

14. Note that it is also possible to define "calo jets" uniquely from the energy deposits in the calorimeters. However, since this does not allow to apply *b*-tagging techniques based on secondary vertices, this has not been considered in the present work.

Resolution. In general, the energy resolution is described by the following formula [40]:

$$\left(\frac{\sigma_E}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + \left(\frac{N}{E}\right)^2 + C^2 \tag{3.13}$$

where

— S is the stochastic term, accounting for statistical fluctuations in the cascade detection;

— N is the electric noise,

— C is a constant contribution, standing for miscalibrations.

The stochastic term is more (less) important for the HCAL (ECAL) since it is a sampling (homogeneous) calorimeter, but its effect becomes less relevant at high energy. In addition, given the structure of the CMS detector and the different rapidity coverages of the different subdetectors, the energy resolution in the calorimeters also depends on the pseudorapidity; details may be found in [41]. Details related to the jet energy calibration will be addressed while performing the analysis in Chap. 7.

3.2.2.3 Muon detection

Overview. There are three types of *muon chambers*. They all are gaseous detectors: the principle is to fill a volume with a gas and an electric field, so that an incoming particle may ionise the gas and produce an electric signal. Different types of gaseous detectors exist, according to the handling of the electric signal.

Description. In the case of the muon chambers, one may distinguish three types of detectors:

CSC Cathode Strip Chambers are based on the principle of multi-wire proportional chambers. They are made of arrays of anode wires and cathode strips, arranged perpendicularly within a gas volume. The gas is ionised by the passage of a muon, its position being determined by the intensity of currents induced on the strips. The 540 CSCs are arranged in six layers in the end-caps.

DT Drift Tubes are situated in the barrel ($|\eta| < 1.6$). Their dimensions are $4\,\text{cm} \times 2\,\text{m} \times 2.5\,\text{m}$, in which an anode wire is stretched; they are arranged in layers perpendicularly to the muon trajectories. When a muon goes through a DT, the gas is ionised, and the motion of the released electrons and ions induces a signal. The delay of the signal is then used in order to determine the position of the muon. As a DT gives only one coordinate, 3×60 DTs are arranged in three layers: the first and the last layers are used to measure the perpendicular coordinates, whereas the middle one is used to measure the coordinate parallel to the beam pipe.

RPC Resistive Plate Chambers are made of an anode and a cathode plates separated by a gas volume. The material that are used are highly resistive to intense electric fields: $E \simeq 50\,\text{kV/cm}$. 480 (288) RPCs are arranged in four concentric cylinders (four disks) in the barrel, $|\eta| < 1.6$ (in each end-cap, $|\eta| > 1.6$). RPCs are also used to obtain an estimation of the momenta of the muons in real time; however, the spatial resolution is moderate.

The muon chambers are combined with the central tracker in the reconstruction of the trajectories of the muons; they improve both the identification and the determination of the transverse momenta. A procedure of alignment of the muon system with itself and with the tracker is also required, but is not discussed here.

Reconstruction. Roughly, the reconstruction of a muon is performed in different steps:

local inside a muon chamber,

standalone combination of all muon chambers

global combination of the muon system and the tracker.

Refinements exists; for instance, muons are sometimes primarily reconstructed in the tracker and related to hits in the muons chambers if the hits in the muon chambers were not enough to determine the passage of a muon. The identification of muons is rather performant as they must have been seen in the muon chambers. Several levels of quality have been defined to describe the muon reconstruction.

tight optimised for weak boson studies, with reconstruction from both the tracker and in two different muons chambers;

soft optimised for b quark decays, with reconstruction in one CSC or in one DT;

loose optimised for multiple-muons events, with reconstruction after the full combination of the tracker, the muon chambers and the calorimeter;

Muons are mainly mentioned for CMS has been especially designed to have a performant muon identification; however, muons do not play a crucial rôle in this thesis.

Resolution. The resolution follows the same principles as for standard tracking.

3.2.2.4 Global event reconstruction

In a nutshell, the reconstruction goes as follows:

— Muons are identified easily thanks to the muon chambers.

— Electrons and positrons are reconstructed thanks to the association of a curved track and a deposit of energy.

— Photons are identified only thanks to the energy deposits in the ECAL.

— Hadrons are associated to all the left deposits in the HCAL.

— Jets are clustered from collimated particles (see Fig. 2.3).

— Neutrinos are not reconstructed but their presence is estimated from the MET.

This technique of reconstruction is called Particle-Flow (PF) algorithm [42]. In practice, the PF algorithm is tuned for each experiment, according to the specificities.

Moreover, the algorithms of reconstruction of the whole event are usually more complex than the ones introduced in this chapter, accounting for magnetic effects, pile-up, superposition of the tracks coming from the same bunch crossing, decays in flight, detector inefficiencies, etc.

3.2.2.5 Trigger and storage systems

The LHC is designed to acquire data at a very high collision rate, so high as 40 MHz; in 2016, bunches follow one another every 25 ns. At each bunch crossing, tens of collisions take place—the *pile-up*—, and each collision may provide hundreds of particles. Given that one particle must be detected in several parts of the detectors that have to be combined to reconstruct its trajectory and determine its nature, the total amount of data produced at each second is equivalent to several thousands of Wikipedia encyclopedias. This represents a too high volume to be treated or even stored. Therefore, a *trigger system* is set up,

selecting potentially interesting events [43]. In practice, a very few events are accepted: around one event for a few hundred thousands.

At CMS, the trigger system is composed of two levels:

L1 The Level 1 applies at the *hardware* system, only involving the calorimeters and the muon chambers for a stream of order of 100 kHz.

HLT The High-Level Trigger corresponds to the *software* system and is itself organised in different sub-levels:

- the `Level 2` also relies on the information from the calorimeters and from the muon chambers.
- the `Level 2.5` combines the information at `Level 2` with tracks in the pixel tracker in order to determine the region of extrapolation in the strip tracker.
- finally the `Level 3` combines `Level 2.5` with the strip tracker.

The stream delivered by HLT is of order of 100 Hz.

The whole event is then registered, including all electronic signals in order to investigate possible biases or dysfunctions. The data is sent to different computing centres: first the `Tier 0` at CERN, then to other computing and storing centres called `Tier 1` and `Tier 2` based in participating institutes and universities. The data files will then be further investigated, or *certified*, in order to determine whether they are valid for physics analyses.

Summary

The particles have followed a long procedure of extraction and acceleration until they are grouped by bunches and ready to collide. Bunches cross, a few tens of protons collide, and as a result, many particles are produced in the beam pipe; most of them decay. The remaining ones cross the devices of the detector: first the particles go through the tracker, aiming at reconstructing the trajectories of the charged particles. Secondly, the particles pass through the ECAL, which stops photons as well as positrons and electrons with electromagnetic interactions. Thirdly, the remaining particles go to the HCAL; the same procedure applies to the hadrons, but adapted to their nuclear interactions. Besides, the muons are the only particles left that can still be detected; this is achieved by the muon chambers.

With the L1 trigger, one can decide whether an event is interesting to keep or not only on the basis of the calorimeter and the muon chambers; if the L1 level triggers, the HLT refines the selection of calorimeters and reconstructs tracks. After the selection (a few tens per second at most) the data is sent to the computing centres of the CERN and its partners (`Tier 0`, `1`, `2`), and parsed in data files. After a careful check of data quality, most runs are certified for physics analyses.

References

[1] CMS Collaboration. *CMS public web site*. http://cms.web.cern.ch. Accessed: 2017-09-08.

[2] CERN. *Logo*. http://design-guidelines.web.cern.ch/outline-logo. Accessed: 2018=01-07.

[3] Christiane Lefèvre. "The CERN accelerator complex. Complexe des accélérateurs du CERN". Dec. 2008. URL: https://cds.cern.ch/record/1260465.

[4] CERN. *LHC the guide*. https://cds.cern.ch/record/2255762/files/CERN-Brochure-2017-002-Eng.pdf. Accessed: 2018-10-06.

[5] Lyndon Evans and Philip Bryant. "LHC Machine". In: *Journal of Instrumentation* 3.08 (2008), S08001. URL: http://stacks.iop.org/1748-0221/3/i=08/a=S08001.

[6] W.J. Stirling. *private communication*. http://www.hep.ph.ic.ac.uk/~wstirlin/plots/plots.html. Accessed: 2018-01-07. 2012.

[7] F. R. Elder et al. "Radiation from Electrons in a Synchrotron". In: *Physical Review* 71 (June 1947), pp. 829–830. DOI: 10.1103/PhysRev.71.829.5.

[8] CERN. *25 years of Large Hadron Collider experimental programme*. https://home.cern/about/updates/2017/12/25-years-large-hadron-collider-experimental-programme. Accessed: 2018-01-07.

[9] Django Manglunki. "Calcul, technique et réalisation des accélérateurs de particules". Course of 2nd master yead of civil engineering, ULB, given at CERN. 2013.

[10] Wikimedia Foundation. *Linear accelerator for elementary particles with drift tubes*. https://commons.wikimedia.org/wiki/File:Lineaer_accelerator_en.svg. Accessed: 2018-01-07.

[11] CERN. *LHC Machine Outreach*. http://lhc-machine-outreach.web.cern.ch/lhc-machine-outreach/. Accessed: 2018-01-07.

[12] K. Hanke. "Past and present operation of the CERN PS Booster". In: *Int. J. Mod. Phys.* A28 (2013), p. 1330019. DOI: 10.1142/S0217751X13300196.

[13] Pierre Freyermuth. *LHC Report: imaginative injectors*. https://cds.cern.ch/journal/CERNBulletin/2016/32/NewsArticles/2201549?ln=en. Accessed: 2017-09-08.

[14] *CMS glossary*. https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookGlossary. Accessed: 2017-04-17.

[15] P. V. Landshoff. "The Total cross-section at the LHC". In: *Acta Phys. Polon.* B39 (2008), pp. 2063–2094. arXiv: 0709.0395 [hep-ph].

[16] CMS Collaboration. "Measurement of the inelastic proton-proton cross section at $\sqrt{s} = 13$ TeV". In: (2016).

[17] CMS. *CMS Public TWiki page for luminosity*. https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults. Accessed: 2017-09-30.

[18] CMS Collaboration. "CMS Luminosity Measurements for the 2016 Data Taking Period". In: (2017).

[19] CERN. *CERN operation webtools*. https://op-webtools.web.cern.ch. Accessed: 2017-09-08.

[20] CMS Collaboration. "The CMS experiment at the CERN LHC". In: *Journal of Instrumentation* 3.08 (2008), S08004. URL: http://stacks.iop.org/1748-0221/3/i=08/a=S08004.

[21] V. Berardi et al. "TOTEM: Technical design report. Total cross section, elastic scattering and diffraction dissociation at the Large Hadron Collider at CERN". In: (2004).

[22] O. Adriani et al. "Technical design report of the LHCf experiment: Measurement of photons and neutral pions in the very forward region of LHC". In: (2006).

[23] James Pinfold et al. "Technical Design Report of the MoEDAL Experiment". In: (2009).

[24] A.L.S. Angelis et al. "CASTOR: Centauro and strange object research in nucleus-nucleus collisions at the LHC". In: *Nuclear Physics B - Proceedings Supplements* 97.1 (2001), pp. 227–230. ISSN: 0920-5632. DOI: https://doi.org/10.1016/S0920-5632(01)01270-1. URL: http://www.sciencedirect.com/science/article/pii/S0920563201012701.

[25] W. W. Armstrong et al. "ATLAS: Technical proposal for a general-purpose p p experiment at the Large Hadron Collider at CERN". In: (1994).

[26] *LHCb : Technical Proposal*. Tech. Proposal. Geneva: CERN, 1998. URL: https://cds.cern.ch/record/622031.

[27] "ALICE: Technical proposal for a large ion collider experiment at the CERN LHC". In: (1995).

[28] "CMS, the Compact Muon Solenoid: Technical proposal". In: (1994).

[29] Stefaan Tavernier. *Experimental Techniques in Nuclear and Particle Physics*. Berlin, Germany: Springer, 2010. DOI: 10.1007/978-3-642-00829-0.

[30] Siona Ruth Davis. "Interactive Slice of the CMS detector". In: (Aug. 2016). URL: https://cds.cern.ch/record/2205172.

[31] Serguei Chatrchyan et al. "Description and performance of track and primary-vertex reconstruction with the CMS tracker". In: *JINST* 9.10 (2014), P10009. DOI: 10.1088/1748-0221/9/10/P10009. arXiv: 1405.6569 [physics.ins-det].

[32] Rudolph Emil Kalman. "A New Approach to Linear Filtering and Prediction Problems". In: *Transactions of the ASME–Journal of Basic Engineering* 82.Series D (1960), pp. 35–45.

[33] Kenneth Rose. "Deterministic annealing for clustering, compression, classification, regression, and related optimization problems". In: *Proceedings of the IEEE* 86.11 (1998), pp. 2210–2239.

[34] R. Fruhwirth, W. Waltenberger, and P. Vanlaer. "Adaptive vertex fitting". In: *J. Phys.* G34 (2007), N343. DOI: 10.1088/0954-3899/34/12/N01.

[35] Serguei Chatrchyan et al. "Identification of b-quark jets with the CMS experiment". In: *JINST* 8 (2013), P04013. DOI: 10.1088/1748-0221/8/04/P04013. arXiv: 1211.4462 [hep-ex].

[36] C. Patrignani et al. "Review of Particle Physics". In: *Chin. Phys.* C40.10 (2016), p. 100001. DOI: 10.1088/1674-1137/40/10/100001.

[37] G L et al. Bayatian. "CMS Physics: Technical Design Report Volume 2: Physics Performance". In: *J. Phys. G* 34.CERN-LHCC-2006-021. CMS-TDR-8-2 (2007). revised version submitted on 2006-09-22 17:44:47, 995–1579. 669 p. URL: http://cds.cern.ch/record/942733.

[38] S Abdullin et al. "ERRATUM : The CMS barrel calorimeter response to particle beams from 2 to 350 GeV/c". In: 60 (Apr. 2009), pp. 353–356.

[39] Tancredi Carli, Klaus Rabbertz, and Steffen Schumann. "Studies of Quantum Chromodynamics at the LHC". In: *The Large Hadron Collider: Harvest of Run 1*. Ed. by Thomas Schörner-Sadenius. 2015, pp. 139–194. DOI: 10.1007/978-3-319-15001-7_5. arXiv: 1506.03239 [hep-ex]. URL: http://inspirehep.net/record/1375494/files/arXiv:1506.03239.pdf.

[40] Francesca Cavallari. "Performance of calorimeters at the LHC". In: *Journal of Physics: Conference Series* 293.1 (2011), p. 012001. URL: http://stacks.iop.org/1742-6596/293/i=1/a=012001.

[41] Serguei Chatrchyan et al. "Energy Calibration and Resolution of the CMS Electromagnetic Calorimeter in pp Collisions at $\sqrt{s} = 7$ TeV". In: *JINST* 8 (2013). [JINST8,9009(2013)], P09009. DOI: 10.1088/1748-0221/8/09/P09009. arXiv: 1306.2016 [hep-ex].

[42] A. M. Sirunyan et al. "Particle-flow reconstruction and global event description with the CMS detector". In: *JINST* 12.10 (2017), P10003. DOI: 10.1088/1748-0221/12/10/P10003. arXiv: 1706.04965 [physics.ins-det].

[43] Vardan Khachatryan et al. "The CMS trigger system". In: *JINST* 12.01 (2017), P01020. DOI: 10.1088/1748-0221/12/01/P01020. arXiv: 1609.02366 [physics.ins-det].

4

Monte Carlo techniques and Physics Generators

In HEP, Monte Carlo (MC) techniques are used in at least two different contexts: compute physics predictions, or simulate the interactions of particles with the detector. In this chapter, the basic techniques are presented. Then different physics generators, such as the ones used in Part II, are detailed.

4.1 Introduction

In this section, we want to illustrate the techniques used to produce predictions in HEP [1, 2]; in particular, we discuss:

— how to generate random numbers,

— and how to compute an integral such as the following:

$$I = \int_a^b f(u)\,\mathrm{d}u \tag{4.1}$$

Indeed, the typical problem in HEP is to integrate a cross section such as in Eq. 2.2. The dimension of such a problem is $d = N_{\text{FS particles}} \times 3 - 4$, where 3 corresponds to the three coordinates of the momentum and 4 to the conservation of energy and momentum. At LHC, the final state is expected to contain a few hundreds of stable particles. In ultra-high-dimension problems, MC techniques are the only hope to achieve integration.

In this section, a few definitions and results in probability theory are recapitulated. Then the efficiency of MC integration and sampling methods are discussed.

4.1.1 Reminder on probability theory

Probability density function. The random variable X has probability density function (p.d.f.) g (non-negative, integrable and normalised to unity) if, for the probability $\mathbb{P}[a \leq X \leq b]$ of X being in the interval $[a, b]$, the following condition is satisfied:

$$\mathbb{P}[a \leq X \leq b] = \int_a^b g(x)\,\mathrm{d}x \tag{4.2}$$

Expectation value and variance. Given f a function of a random variable X following a p.d.f. g, the *expectation value* $\mathbb{E}[f]$ and the *variance* $\mathbb{V}[f]$ are defined as follows:

$$\mathbb{E}[f] = \int_{-\infty}^{\infty} f(x)g(x)\,\mathrm{d}x \tag{4.3}$$

$$\mathbb{V}[f] = \mathbb{E}\left[(f - \mathbb{E}[f])^2\right] \tag{4.4}$$

Law of Large Numbers (LLN). Given independent and identically distributed realisations x_i of the random variable X, for $N \to \infty$:

$$\frac{1}{N}\sum_{i=1}^{N} f(x_i) \longrightarrow \mathbb{E}[f] \tag{4.5}$$

In the particular case of a uniform distribution $u_i \sim \mathcal{U}[a, b]$, we have the simple form of the LLN for $N \to \infty$:

$$\frac{1}{N}\sum_{i=1}^{N} f(u_i) \longrightarrow \frac{1}{b-a}\int_a^b f(u)\,\mathrm{d}u \tag{4.6}$$

4.1.2 Simple Monte Carlo integration

Solution. A basic solution to compute Eq. 4.1 numerically is obviously given by the simple form of the LLN (Eq. 4.6):

$$\mathbb{E}[f] = \frac{1}{b-a}\int_a^b f(u)\,\mathrm{d}u \tag{4.7}$$

$$\approx \frac{1}{N}\sum_{i=1}^{N} f(u_i) \tag{4.8}$$

Hence we have an estimate for our integral:

$$I_{\mathrm{MC}} = \frac{b-a}{N}\sum_{i=1}^{N} f(u_i) \tag{4.9}$$

Using Eq. 4.4, the precision is given by the following:

$$\sigma^2_{\mathrm{MC}} = \mathbb{V}[I_{\mathrm{MC}}] \tag{4.10}$$

$$= \frac{(b-a)^2}{N}\mathbb{V}[f] \tag{4.11}$$

$$= \frac{(b-a)^2}{N}\left(\frac{1}{N}\sum_{i=1}^{N}(f(u_i))^2 - \left(\frac{1}{N}\sum_{i=1}^{N} f(u_i)\right)^2\right) \tag{4.12}$$

Generation of uniformly distributed random numbers. The approach given in the previous paragraph assumes the existence of random generators. While such generators may be found in nature, based on random phenomena (abundant in quantum physics), they are intrinsically not reproducible. In numerical computation, one usually prefers pseudo-random generators, based on non-linear (but still deterministic) algorithms. Typically, a good generator is characterised by its capability to produce number very loosely correlated; but in general, assessing rigorously the quality of a generator is a non-trivial problem. A few examples of typical generators are illustrated in Fig. 4.1:

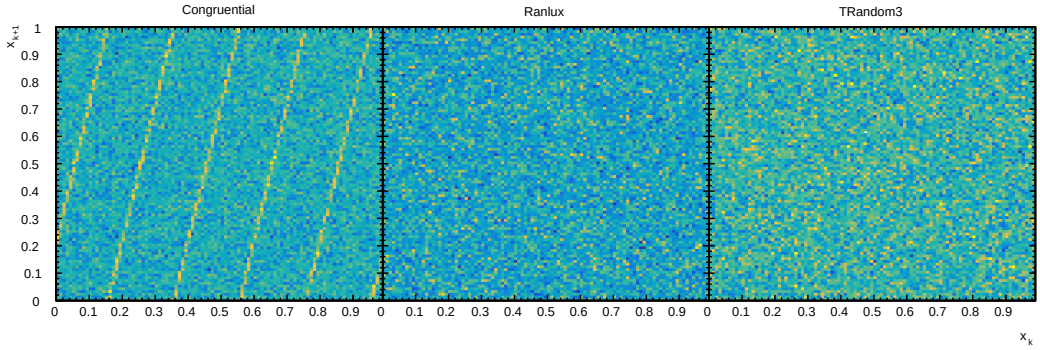


FIGURE 4.1 – Serial correlations in the generation of pseudo-random numbers with 10^7 entries. The congruential operator is given for $a = 205$, $c = 29573$, $m = 139968$ and $x_0 = 4711$. The pattern given by the congruential generator shows correlations, while the two other generators do not exhibit any.

Congruential generator Numbers are picked in a sequence of numbers u_i given by the following congruential relation:

$$x_{i+1} = ax_i + c \bmod m \tag{4.13}$$

A suboptimal choice of the *multiplier* a, the *increment* c and the *modulus* m can lead to serial correlations.

RANLUX Given the large amount of samples required in HEP, and given the increasing computing capabilities, specific random generators have been developed, such as `RANLUX` [3, 4]. This generator is based on the *Marsaglia-Zaman generator* [5] which can be considered as a congruential generator for a smart choice of the parameters (though with additional features), ensuring large periods of typically 10^{171}.

TRandom classes in ROOT Other generators have also been developed, implemented among other in the `ROOT` library [6]. The `TRandom3` class implements an algorithm [7] remarkable for its particularly long period of $2^{19937} - 1$.

An application will be shown in Chap. 7.

4.1.3 Importance sampling

A uniform sampling is not necessarily efficient when the function is peaked in a particular region of the domain. Therefore, one needs to modify the previous approach to improve the performance (Eq. 4.10 shows how the variance is related to the sample u_i).

Principle. Instead of the simple form, one introduces a p.d.f. g that mimics the function to integrate f, and uses the general form of the LLN (Eq. 4.5):

$$I = \int_a^b f(u)\,\mathrm{d}u \tag{4.14}$$

$$= \int_a^b \bar{f}(x)g(x)\,\mathrm{d}x \tag{4.15}$$

$$= \mathbb{E}\left[\bar{f}\right] \tag{4.16}$$

Now, in order to estimate the integral with MC, one considers $\bar{f} = f/g$ as the integrand with the sample x_i distributed according to g:

$$I_{\text{MC}} = \frac{1}{N} \sum \bar{f}(x_i) \tag{4.17}$$

Generation of non-uniformly distributed random numbers. Given $u_i \sim \mathcal{U}[0,1]$, the sample x_i can be described by the p.d.f. g if the following applies:

$$\int_{-\infty}^{x_i} g(x)\,\mathrm{d}x = u_i \int_{-\infty}^{+\infty} g(x)\,\mathrm{d}x \tag{4.18}$$

Intuitively, for u_i uniformly distributed, x_i will be mostly located where g is sharp. As an application, in the PS (see Sect. 2.3.3.2), the next branching is determined by generating a random number according to a Sudakov factor for each outgoing leg that has not reached the hadronisation scale yet: the random number corresponding to the leg with the highest scale is chosen to describe the next branching, and the other random numbers are discarded; the procedure is then iterated, as described previously.

4.2 Physics generators

MC techniques are widely used in HEP to produce theoretical predictions. From the large variety of different physics generators, we review some of them here.

4.2.1 Fixed-order parton-level calculations

Parton-level calculations are based on the factorisation (Eq. 2.37), and only include the PDFs the ME at fixed order (FO) in the perturbative regime. Typically, LO and NLO can be achieved, though some NNLO calculations can sometimes be found for certain processes. However, their use is somehow limited:

— they do not produce full event records similar to the data;

— missing orders can lead to sizeable discrepancies in the predictions for small-cone size jet clustering algorithms [1];

— finally, they do not include hadronisation.

At the time of writing this thesis, standard references of such generators are the following:

MadGraph — MadGraph (or MadEvent — both names shall here be used as synonymous) is a generator at LO and at NLO, including however more diagrams at tree level with an arbitrary number of particles in the final state [8, 9, 10]. In this thesis, we use `MadGraph5_aMC@NLO 2.2.2` at LO for $2 \to 2$, $2 \to 3$ and $2 \to 4$ processes.

powheg box — powheg is the method described in Refs. [11, 12, 13], and the name stands for *POsitive WeigHts Event Generator*. In this thesis, we use it for the production of QCD dijet processes with $2 \to 2$ and $2 \to 3$ at NLO. The additional radiation is treated with Sudakov factor using a splitting function at NLO. The powheg box corresponds to an implementation of the powheg method [14]. Usually, it is not used

1. As an illustration, in App. C, the CMS measurement of the inclusive jet production at $\sqrt{s} = 8\,\text{TeV}$ is presented. In particular, it was observed that the predictions for small cone size radii are significantly underestimating the measured cross section, while the large cone size were correctly describing it.

standalone but is interfaced with an event generator, as described in the next section.

NLOJET++ NLOJET++ is a "pure" fixed-order calculation, in the sense that it is only implement Eq. 2.37. No parton shower can be applied and the result is at parton-level.

4.2.2 Event generators

MC event generators simulate particle collisions on an event-by-event basis, similarly to what happens in the real detector. Generated events include the Matrix Element (ME) and the Underlying Event (UE). The latter can further decomposed as follows:

— extra ISRs and FSRs[2],

— Multi-Parton Interaction (MPI),

— hadronisation,

— hadron decays and soft photon radiations,

— and treatment of beam remnants.

This reflects the current understanding of an interaction, described in Sect. 2.3. The main difficulty consists in connecting the different steps in a consistent way, *i.e.* matching the ME and the UE.

4.2.2.1 General-purpose event generators

Today, three General-Purpose MC Event Generator (GPMC) make reference, including the simulation of different types of scatterings (proton-proton, electron-positron, electron-proton, etc.), different types of interaction (EW, QCD, Higgs processes, processes BSM, etc.):

PYTHIA Several versions of PYTHIA exist, the most recent being PYTHIA 8, written in C++ [15, 16], which is used in this thesis[3] (version 8.205). The hard process is hard-coded at the Born level, *i.e.* at LO. The generator includes p_T-ordered PS with *angular veto* and Lund string hadronisation. The MPI is accounted for with a smoothing factor with parameter p_{T0}. It is also interesting to note that PYTHIA makes use of the Marsaglia – Zaman algorithm presented above for the generation of random numbers.

HERWIG For this generator as well, two versions are maintained in parallel[4]: HERWIG++ and HERWIG 7, both written in C++ [17]. HERWIG 7 implements automated NLO calculations, while in HERWIG++, the hard process is hard-coded at the Born level. Unlike PYTHIA, HERWIG uses *angular-ordering* for PS and *cluster model* for hadronisation. The MPI is accounted for with a sharp cut-off with parameter p_{T0}. In this thesis, only HERWIG++ is considered.

SHERPA SHERPA includes automated calculation of MEs at LO and at NLO [18]. It includes its own PS and UE. Similarly to PYTHIA, the MPI is performed with a smoothing factor. It is here only mentioned and not further discussed in this thesis.

2. Depending on the context, the PS is sometimes considered separately from the UE.

3. One should mention PYTHIA 6, previous version written in FORTRAN, no longer actively maintained. However, PYTHIA 8 is largely based on PYTHIA 6, whose manual stays an important reference.

4. Here, one should also mention the previous version HERWIG 6, written in FORTRAN, but which is also no longer maintained.

General-Purpose MC Event Generator (GPMC) can also be interfaced with some of the fixed-order calculations [19], like MadGraph and powheg, as will be the case in Part II of this thesis, or with other event generators that only take care of a part of the full chain.

One interesting example is Cascade, an event generator using off-shell MEs with CCFM evolution [20]; it includes its own ISR but needs to be interfaced with, for instance, pythia for the FSR and for hadronisation.

4.2.2.2 Tuning

The phenomenological models used in the UE involved parameters, which need to be estimated, or *tuned*. A non-exhaustive list of the parameters is given in Tab. 4.1. A *tune* corresponds to a set of parameters of the UE.

model	parameter	signification
PS	$\alpha_S(M_Z)$	strong coupling
	p_{T0}	smoothing parameter
MPI	R	involved in *colour reconnection*
	p_{T0}	smoothing parameter
hadronisation	a, b	flavour-dependent, only for Lund string fragmentation
	r_Q	Bowler parameter, for HF
	Q_0	hadronisation scale
intrinsic k_T	σ	width of the Gaussian distribution

Table 4.1 – Non-exhaustive list of parameters in phenomenological models that need to be tuned (exhaustive list given in Chap. 9).

In order to determine the parameters, a fit of to data of several MB distributions (described later in this section) can be performed for different values of the parameters. A χ^2 is then computed and the parameters providing the best agreement is selected:

$$\chi^2(\mathbf{p}) = \sum_{\text{observables}} w_{\text{observable}} \sum_{\text{bins}} \left(\frac{f_{\text{bin}}(\mathbf{p}) - \mathcal{R}_{\text{bin}}}{\Delta^2_{\text{bin}}} \right)^2 \tag{4.19}$$

where

— $\mathbf{p}$ stands for the free parameters;

— $\mathcal{R}$ stands for the data (*reference*);

— f stands for the model (hadronisation, PS, etc.);

— Δ_{bin} stands for the experimental uncertainties;

— and w stands for the weight.

The weight is chosen more or less arbitrarily in order to increase the impact of a certain distribution with respect to others.

The naïve approach consists in performing a grid scan of all possible parameters; however, this method is extremely time- and resource-consuming. In the modern approach, the simulation is run in parallel for randomised sets of parameter [21]; the value of the parameters is then determined by interpolations. The current reference for this method in HEP is called Professor (PROcedure For EStimating Systematic errORs) [22, 23].

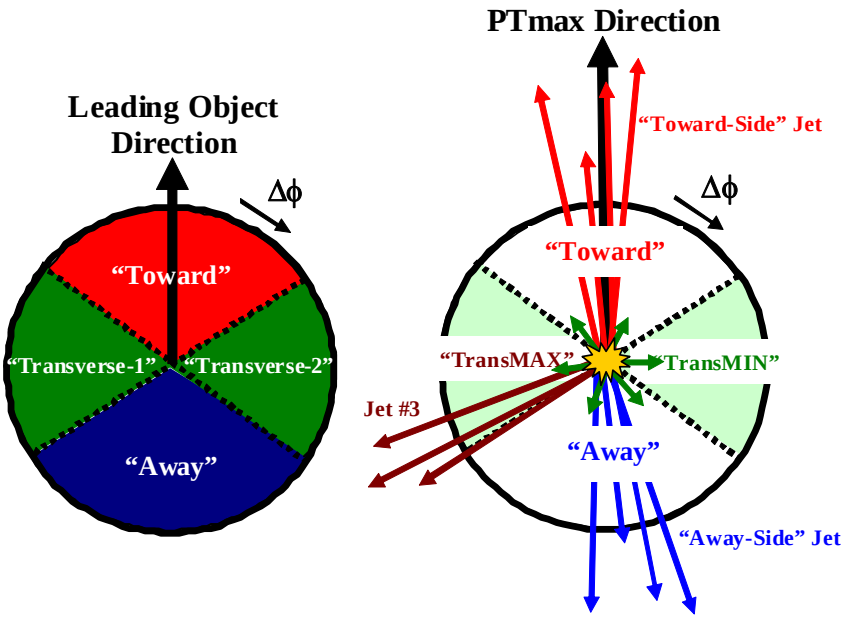


FIGURE 4.2 – Definitions of the different tuning regions in tuning techniques: the *toward* (*away*) region correspond to the direction (opposite direction) of the leading object; the tuning is performed in the transverse region (in green on the LHS). The *transMAX* and *transMIN* (in light green on the RHS) are defined by the presence of a third jet, to disentangle further the contributions to the UE. Figure from [10].

The activity coming from the UE — the *soft* activity — is disentangled from hard interaction by selecting the activity in the region transverse to the direction of the leading-p_T particle, as illustrated in Fig. 4.2. The considered observables are the multiplicity of the charged particles and the scalar sum of the transverse momenta of the charged particles. Sometimes, in order to disentangle contributions from the PS and from the UE, the transverse region is further separated into the *transMAX* region with the presence of a third jet and the *transMIN*.

In Part II, two tunes will be used [10]:

— `CUETP8M1` (CMS UE Tune PYTHIA 8 Monash 1)

— `CUETHppS1` (CMS UE Tune HERWIG++ Set 1)

They only differ in the physics generator that is used, but follow the same treatment:

— They are both based on the tunes obtained from previous experiments. For instance, the hadronisation parameters are still corresponding to parameters obtained at the LEP in $e^+e^- \to Z \to q\bar{q}$.

— For parameters related to hadronic collisions (like MPI or intrinsic k_T), MB data samples from Tevatron at $\sqrt{s} = 900\,\text{GeV}$ and $1.96\,\text{TeV}$ and from LHC at $\sqrt{s} = 7\,\text{TeV}$ are used.

The interest of b jet measurements will be discussed in Chap. 5.

4.3 Detector simulation

One major difference still remains between the event records in physics generators and in data: the former contains the exact kinematics of the full list of stable and unstable particles, while the latter contains lists of tracks and of energy deposits. In order to compare

data and simulation, one needs either to apply the effect of the detector (*i.e.* simulate the interactions of the particles with the different subdetectors) or to correct for it in the data. In both cases, the simulation of the detector is needed.

One software among all makes reference for the simulation of interactions of high-energy particles with media: GEometry ANd Tracking (GEANT4) [24, 25, 26]. It is used for the simulation of most of the modern experiments in HEP such as CMS, as well as in many experiments in space science, medical science, and engineering.

Given the description of the detector, which encompasses not only the active but also the passive parts of the detector, GEANT4 simulates the trajectories of all outgoing particles. In the description of the interaction of particles with the detector, random number generators are typically of use in the simulation of *multiple scattering* or of *decays*.

Although the simulation of the detector is given with very good precision, it is not perfectly correct. The *response* of the detector will present differences between the data and the simulation; for instance, in practice, the time dependence is not simulated.

Conclusion

MC techniques are heavily used in HEP at all levels. In Part II, MC simulations will be used in different contexts:

— First, the effect of the detector will be studied thanks to simulations based on MC samples (Chap. 7).

— Then, the effect of the detector on data will be removed thanks to the simulations (Chap. 8).

— Finally, the corrected data will be compared to predictions obtained from MC samples (Chap. 9).

However, it is important to stress that simulations only offer an approximative description of nature. Simulations are to be handled with care when comparing to measurements and when utilising them to correct the measurement from detector effects. An important part of the difficulty of a physics analysis relies in the assessment the description of the measurement by the simulation.

4.A Further methods of MC integration

Additional methods exist to compute efficiently integrals, by subdividing the region to integrate.

Stratified sampling. Here, the region is divided horizontally by considering successive intervals:

$$\int_a^b f(x)\,\mathrm{d}x = \int_a^c f(x)\,\mathrm{d}x + \int_c^b f(x)\,\mathrm{d}x \tag{4.20}$$

Each term can then be integrated with a different p.d.f.. In Part II, several examples will be encountered with the samples used to study the transverse momentum, which is a steeply falling spectrum, in order to have a large enough statistical sample also at high values:

— The PYTHIA-8 sample is produced in slices of the transverse momentum $\hat{p}_T$ of the outgoing partons, as illustrated in Fig. 4.3.

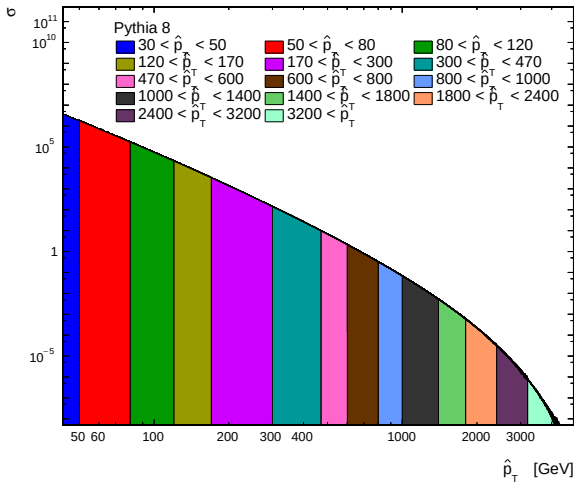


FIGURE 4.3 – The PYTHIA-8 samples are generated in 14 bins of $\hat{p}_T$, corresponding to the transverse momentum of the hard partons in the final state. A similar statistical precision can therefore be achieved over several orders of magnitude.

— Similarly, the MADGRAPH sample is produced in slices of scalar sum of the transverse momenta of the hard partons in the final state $H_T = \sum_i p_{Ti}$.

Subtraction method. Here, the region is divided vertically on the same interval by considering different p.d.f. f and g:

$$\int_a^b f(x)\,\mathrm{d}x = \int_a^b g(x)\,\mathrm{d}x + \int_a^b \left(f(x) - g(x)\right)\,\mathrm{d}x \tag{4.21}$$

This method is typically used in fixed-order calculations. For instance, while in MC@NLO, some events have negative weight, terms are rearranged in POWHEG such that all entries are positive (hence its name).

Hit-or-Miss integration. This integration method is an alternative to Eq. 4.17. The principle is to pick N couples of random numbers x_i, y_i in a rectangle:

$$a \leq x_i \leq b \tag{4.22}$$
$$0 \leq y_i \leq c \tag{4.23}$$

Determine the number M of couples such that $y_i \leq f(x_i)$. The integral is then estimated as follows:

$$I \approx c(b-a)\frac{M}{N} \tag{4.24}$$

A example of application of this algorithm can be found in the treatment of the PS, while dealing with the Sudakov factor. .

References

[1] F. James. "Monte Carlo theory and practice". In: *Reports on Progress in Physics* 43.9 (1980), p. 1145. URL: http://stacks.iop.org/0034-4885/43/i=9/a=002.

[2] Glen Cowan. *Statistical data analysis*. Oxford University Press, 1998.

[3] Martin Lüscher. *RANLUX*. http://luscher.web.cern.ch/luscher/ranlux/guide.pdf. 2016.

[4] Martin Luscher. "A Portable high quality random number generator for lattice field theory simulations". In: *Comput. Phys. Commun.* 79 (1994), pp. 100–110. DOI: 10.1016/0010-4655(94)90232-1. arXiv: hep-lat/9309020 [hep-lat].

[5] George Marsaglia and Arif Zaman. "A New Class of Random Number Generators". In: *Ann. Appl. Probab.* 1.3 (Aug. 1991), pp. 462–480. DOI: 10.1214/aoap/1177005878. URL: https://doi.org/10.1214/aoap/1177005878.

[6] Rene Brun and Fons Rademakers. "ROOT — An object oriented data analysis framework". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 389.1 (1997). New Computing Techniques in Physics Research V, pp. 81–86. ISSN: 0168-9002. DOI: https://doi.org/10.1016/S0168-9002(97)00048-X. URL: http://www.sciencedirect.com/science/article/pii/S016890029700048X.

[7] Makoto Matsumoto and Takuji Nishimura. "Mersenne Twister: A 623-dimensionally Equidistributed Uniform Pseudo-random Number Generator". In: *ACM Trans. Model. Comput. Simul.* 8.1 (Jan. 1998), pp. 3–30. ISSN: 1049-3301. DOI: 10.1145/272991.272995. URL: http://doi.acm.org/10.1145/272991.272995.

[8] Johan Alwall et al. "MadGraph 5 : going beyond". In: *JHEP* 06 (2011), p. 128. DOI: 10.1007/JHEP06(2011)128. arXiv: 1106.0522 [hep-ph].

[9] J. Alwall et al. "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations". In: *JHEP* 07 (2014), p. 079. DOI: 10.1007/JHEP07(2014)079. arXiv: 1405.0301 [hep-ph].

[10] Vardan Khachatryan et al. "Event generator tunes obtained from underlying event and multiparton scattering measurements". In: *Eur. Phys. J.* C76.3 (2016), p. 155. DOI: 10.1140/epjc/s10052-016-3988-x. arXiv: 1512.00815 [hep-ex].

[11] Simone Alioli et al. "Jet pair production in POWHEG". In: *JHEP* 1104 (2011), p. 081. DOI: 10.1007/JHEP04(2011)081. arXiv: 1012.3380 [hep-ph].

[12] Simone Alioli et al. "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX". In: *JHEP* 06 (2010), p. 043. DOI: 10.1007/JHEP06(2010)043. arXiv: 1002.2581 [hep-ph].

[13] P. A Nason. "Recent developments in POWHEG". In: *Proceedings, 9th International Symposium on Radiative Corrections: Applications of quantum field theory to phenomenology. (RADCOR 2009).* Vol. RADCOR2009. 2010, p. 018. arXiv: 1001.2747 [hep-ph].

[14] Tomáš Ježo and Paolo Nason. "On the Treatment of Resonances in Next-to-Leading Order Calculations Matched to a Parton Shower". In: *JHEP* 12 (2015), p. 065. DOI: 10.1007/JHEP12(2015)065. arXiv: 1509.09071 [hep-ph].

[15] Torbjörn Sjöstrand, Stephen Mrenna, and Peter Skands. "PYTHIA 6.4 physics and manual". In: *JHEP* 05 (2006), p. 026. DOI: 10.1088/1126-6708/2006/05/026. arXiv: hep-ph/0603175 [hep-ph].

[16] Torbjörn Sjöstrand, Stephen Mrenna, and Peter Skands. "A brief introduction to PYTHIA 8.1". In: *Computer Physics Communications* 178.11 (2008), pp. 852–867. ISSN: 0010-4655. DOI: https://doi.org/10.1016/j.cpc.2008.01.036. URL: http://www.sciencedirect.com/science/article/pii/S0010465508000441.

[17] M. Bahr et al. "Herwig++ physics and manual". In: *Eur. Phys. J. C* 58 (2008), p. 639. DOI: 10.1140/epjc/s10052-008-0798-9. arXiv: 0803.0883 [hep-ph].

[18] T. Gleisberg et al. "Event generation with SHERPA 1.1". In: *Journal of High Energy Physics* 2009.02 (2009), p. 007. URL: http://stacks.iop.org/1126-6708/2009/i=02/a=007.

[19] E. Boos et al. "Generic user process interface for event generators". In: *Physics at TeV colliders. Proceedings, Euro Summer School, Les Houches, France, May 21-June 1, 2001.* 2001. arXiv: hep-ph/0109068 [hep-ph]. URL: http://lss.fnal.gov/archive/preprint/fermilab-conf-01-496-t.shtml.

[20] H. Jung. "The CCFM Monte Carlo generator Cascade". In: *Computer Physics Communications* 143.1 (2002), pp. 100–111. ISSN: 0010-4655. DOI: https://doi.org/10.1016/S0010-4655(01)00438-6. URL: http://www.sciencedirect.com/science/article/pii/S0010465501004386.

[21] P. Abreu et al. "Tuning and test of fragmentation models based on identified particles and precision event shape data". In: *Z. Phys.* C73 (1996), pp. 11–60. DOI: 10.1007/s002880050295.

[22] Andy Buckley et al. “Systematic event generator tuning for the LHC”. In: *Eur. Phys. J.* C65 (2010), pp. 331–357. DOI: 10.1140/epjc/s10052-009-1196-7. arXiv: 0907.2973 [hep-ph].

[23] Andy Buckley et al. *Tutorial on* PROFESSOR. https://professor.hepforge.org/prof-tutorial.pdf. Accessed: 2018-01-01.

[24] S. Agostinelli et al. “Geant4—a simulation toolkit”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 506.3 (2003), pp. 250–303. ISSN: 0168-9002. DOI: https://doi.org/10.1016/S0168-9002(03)01368-8. URL: http://www.sciencedirect.com/science/article/pii/S0168900203013688.

[25] John Allison et al. “Geant4 developments and applications”. In: *IEEE Trans. Nucl. Sci.* 53 (2006), p. 270. DOI: 10.1109/TNS.2006.869826.

[26] J. Allison et al. “Recent developments in Geant4”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 835.Supplement C (2016), pp. 186–225. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2016.06.125. URL: http://www.sciencedirect.com/science/article/pii/S0168900216306957.

5

Discovery, overview and motivation of beauty physics

> *"Why is bottom quark physics so interesting? The cynic might argue that the labs are into bottom quark physics because it's affordable."*
>
> — Edward H. Thorndike [1]

In the first part of this chapter, the discovery and current knowledge of the properties of the b quark are reviewed. Then the current status and prospects for deepening this knowledge with the LHC experiments are detailed. Finally, tagging techniques and measurements of b quarks as a probe are discussed.

5.1 Discovery and first measurements

5.1.1 The E288 experiment at Fermilab

In the seventies, several experiments at Fermilab were ongoing to search for new resonances, especially the $W^{\pm}$ and Z^0 bosons of the electroweak theory. In this context, the *bottom* quark — or more poetically *beauty* quark — was discovered in 1977 with a fixed-target experiment at Fermilab led by Leon Lederman in *bottomonium* states $b\bar{b}$ [2, 3, 4, 5].

In the measurement of the invariant mass of dimuon systems from the outgoing particles of the collisions of 400-GeV protons on a nuclear target, they observed a double-peak structure around 9.5 GeV (see Fig. 5.1). The situation being very analogous to the discovery of J/ψ [7, 8], *i.e.* a very narrow resonance, the new discovery was similarly interpreted as bound systems of quark-antiquark pairs of a new generation, this time with charge $-1/3$, called Υ's [1].

The process of production and decay of $b\bar{b}$ states can be described as follows:

$$p + N \to \Upsilon \to b\bar{b} \to \gamma^* \to \mu^+\mu^- \tag{5.1}$$

where N stands for copper, platinum or beryllium; this corresponds to an electromagnetic decay. While the existence of a new quark seemed clear, its properties, such as electroweak

1. Two legends exist regarding the name of the particle. The first is that "Upsilon" stand for "up + psi", in reference to the J/ψ. The other legend is related to the spokesperson and leader of team's name: the Nobel-prized Leon Lederman. Observing first a resonance at 6 GeV in 1976, the members of the team agreed that its name would be Upsilon if it would be established, or *Oops-Leon* if it would not. The resonance at 6 GeV turned out to be nothing else than a fluctuation; however, using the same experimental set-up, they found a new resonance at 9 GeV: this time a real resonance. They decided to keep the name Upsilon.

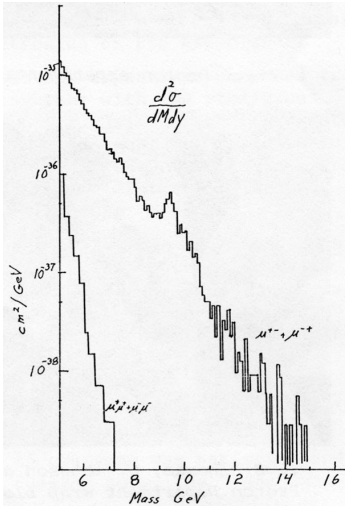


FIGURE 5.1 – "The plot showing an excess of events around 9.50 GeV, marking the discovery of the Υ. « There was no known object that could explain that bump, » Leon LEDERMAN said, E288 spokesman and Fermilab director emeritus." [6]

properties, could not be further determined; therefore this production mode of b quarks is called *hidden beauty* production. Confirming its existence and determining further its properties in the production of *open beauty* motivated further investigations on this new quark in later experiments.

But in addition, in contrast to the discovery in 1974 of the *charmonium* J/ψ that was expected (with the *GIM mechanism* [9, 10]) and understood as a pair of heavy quarks [11], the discovery in 1978 of the bottomonium Υ was rather a surprise, since a third generation of quarks was not yet evident. So far, only the famous publication on CP violation by Makoto KOBAYASHI and Toshihide MASKAWA [12] was assuming six quarks. The *tauon* was also not well established, which would have led to more serious speculations on a third generation of quarks. In addition, its discovery raised new questions, such as the existence of a partner to the fifth quark.

5.1.2 ISR

Situated at CERN, the Intersecting Storage Rings (ISR) was the first hadron collider in the world, composed of two rings with 150 m of diameter. It could deliver proton-proton and $p\bar{p}$ collisions at a centre-of-mass energy up to 62 GeV. It was active from 1971 to 1984 [13].

At the end of 1977, ISR confirmed the discovery of the Υ with proton-proton collisions in the di-electron invariant-mass spectrum [14]:

$$pp \to \Upsilon X \to e^+e^-X \tag{5.2}$$

The mass was estimated to $m_\Upsilon = 9.46 \pm 0.16$ GeV, in agreement with the findings of the E288 experiment.

5.1.3 DORIS

DORIS [2] was one of the first *storage rings*, allowing e^+e^- collisions, with a circumference of 289 m. Built from 1969 to 1974 at *Deutsches Electroknen-Synchrotron* (DESY), each beam was carrying 3.5 GeV of energy in the lab frame. If this was enough to study the production of charmonium, this was too low to reproduce the new resonance discovered at Fermilab. After undergoing an upgrade in 1978, beams could reach an energy of 5 GeV, which was then enough to confirm the findings of Fermilab `E288` experiment.

Three experiments at DORIS confirmed the resonance in the same year: Pluto, DASP2 and DHHM.

The Pluto collaboration [3] confirmed the existence of the resonance in 1978 [15] through its electromagnetic decay (Eq. 5.1) in hadronic production: whereas the decay into a dimuon system accounts for around 3% of the decays, hadronic decays account for 90%, and the most significant channel is the production of three hadrons:

$$e^+ + e^- \to \Upsilon \to b\bar{b} \to ggg \to hhh \tag{5.3}$$

The Pluto experiment measured $m_\Upsilon = 9.46 \pm 0.01$ GeV and confirmed from the measurement that, in analogy with J/ψ, it should be made of a pair of quarks of charge $-1/3$ [11].

In addition, Pluto confirmed the discovery of the tauon, first observed at the SPEAR [4] in the years 1974 to 1977. Indeed, the existence of a third generation of leptons seriously supports the existence of a third generation of quarks.

The Double Arm SPectrometer (DASP2) experiment published simultaneously as Pluto [16, 17]. The measurement was simpler, since it was only relying on the direction of the particles. In addition, the measurement was also performed through the electromagnetic decay:

$$e^+ + e^- \to \Upsilon \to \mu^+\mu^- \tag{5.4}$$

A few months later, the DESY-Hamburg-Heidelberg-München collaboration (DHHM) experiment also announced it, together with an excited state at 10.02 ± 0.02 GeV [18], similarly to `E288`, resolving then the double-peak structure previously observed.

The *Cornell potential*, already used to describe the $c\bar{c}$ system [19], is also tested to describe the $b\bar{b}$ system. This potential is composed of two contributions:

$$U(r) = \underbrace{-\frac{a}{r}}_{\text{Coulombic part}} + \underbrace{br}_{\text{confinement}} \tag{5.5}$$

where

— the Coulombic part describes the one-gluon interaction in analogy to the Coulomb interaction;

— the confinement part includes the non-perturbative, not well understood effects.

This empirical potential [5] is successfully used for charm and bottom.

2. *DOppel-RIng Speicher*
3. Named after the magnet of the detector.
4. *Stanford Positron Electron Asymmetric Rings*, situated at the SLAC.
5. One obvious limitation of this potential is that it does not allow the *fragmentation* of a pair of quarks. However, it successfully allowed to perform spectroscopy and describe lifetime.

Meson	Quark content	M/ MeV	τ/ps
$\boldsymbol{B^{\pm}}$	$u\bar{b}$, $\bar{u}b$	5279.29 ± 0.15	1.638 ± 0.004
$\boldsymbol{B^{0}}$	$d\bar{b}$	5279.61 ± 0.16	1.520 ± 0.004
$B_c^{\pm}$	$c\bar{b}$, $\bar{u}b$	6275.1 ± 1.0	0.507 ± 0.009
B_s^0	$s\bar{b}$	5366.79 ± 0.23	1.510 ± 0.005

TABLE 5.1 – B mesons' content, mass and lifetime. The $B^{\pm}$ and B^0 are from far the most common ones, and have almost the same mass and very similar lifetimes. B^+ and B^- are one another's respective antiparticles. Values are taken from PDG [24].

5.1.4 CESR

After the successes at `E288` and at DORIS, there could be no more doubt about the existence of the fifth quark. Yet it was still observed in its hidden form, *i.e.* in the form of Υ mesons. Another topic of interest was the spectroscopy: how similar to J/ψ was Υ? In particular, should one expect to find an excited state decaying in B mesons, analogously to D mesons? Finally, what about a sixth quark?

At the *Cornell's laboratory for Nuclear Physics*, a new, 768-meter long, symmetric e^+e^- collider was being built at the time of the discovery, the Cornell Electron-positron Storage Ring, pronounced "Caesar" (CESR), with two detectors: *Cleopatra* (for her/its proximity with Caesar) (CLEO) and Columbia University-Stony Brook (CUSB). Their contributions to HF physics would be important: a general review of Υ physics at CESR may be found in [20]; here, only a few key steps are mentioned.

Just a few months after the commissioning of the experimental set-up, the three first resonances were found before end of 1979 [21].

But more interestingly, in Spring 1980, a fourth resonance was found around 10.5 GeV by both experiments [22, 23]. The peak was much broader, indicating a faster decay; this was understood, in analogy to the decay of J/ψ in pairs of D mesons, as a decay of Υ in a new kind of mesons, named B mesons:

$$e^+e^- \to \Upsilon(4S) \to B + \bar{B} \quad \text{(strong int.)} \tag{5.6}$$

where B is a B meson. The characteristics of B mesons are summarised in Tab. 5.1: their content, mass and lifetime are given.

Then the B meson will further decay weakly:

$$B \to XW \to Xl\nu \quad \text{(weak int.)} \tag{5.7}$$

A peak being found in the outgoing-lepton spectrum at this energy, "bare bottom" was found. The door to a new area of physics was definitely open.

5.2 Further investigations

Since its discovery, the Υ and B mesons have kept physicists busy for several reasons:

— First, the electroweak properties related to the existence of the bare bottom, typically the CKM matrix and the CP violation, are heavily studied at CESR and in following experiments (Positron-Electron Project (PEP), DORIS-II, ...).

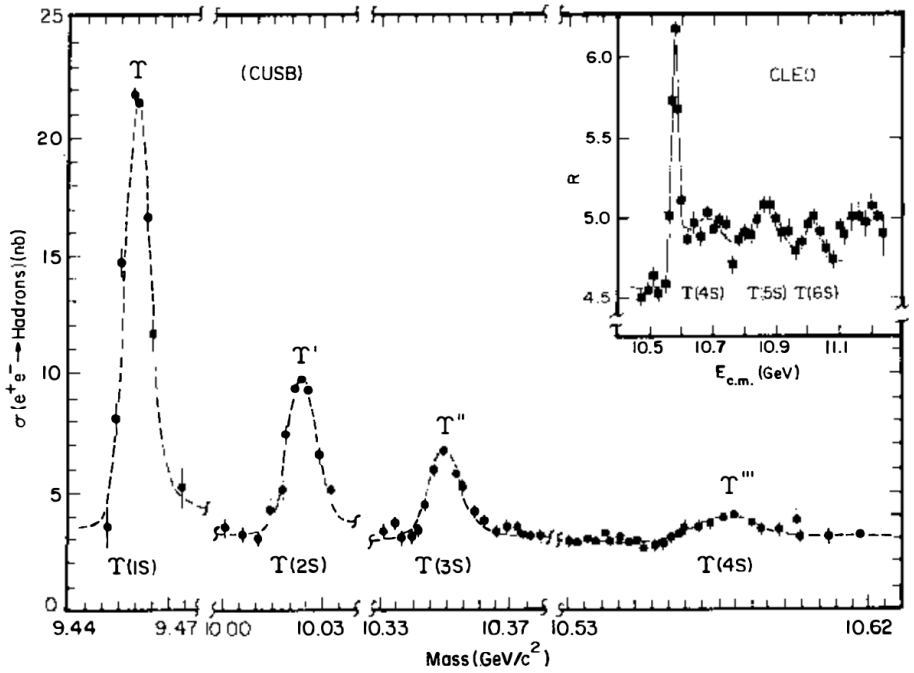


FIGURE 5.2 – Four Υ peaks at both CLEO and CUSB [25]. The three first peaks may be explained in terms of $b\bar{b} \to ggg$, $b\bar{b} \to \gamma gg$ or $b\bar{b} \to \gamma* \to q\bar{q}/l^+l^-$. In the fourth peak, one should also consider $b\bar{b} \to B\bar{B}$ where $B = b\bar{q}$ ($\bar{B} = \bar{b}q$).

— The partner of the beauty quark, the *top* — or *truth*[6] — is found only in 1995 at TEVATRON by the DØ[7] and Collider Detector at Fermilab (CDF)[8] collaborations, confirming many predictions obtained from measurements of b and B properties.

— Moreover, the study of the quarkonium $b\bar{b}$ (together with $c\bar{c}$) continued, giving rise to the Υ *spectroscopy*.

— B mesons are investigated, in particular for their remarkably long lifetimes. Also, other mesons, B_s and B_c, combining b with s or c quarks, are observed.

Since most of the modern particle accelerators have allowed the study of b quarks and B mesons, an exhaustive review is beyond the scope of this thesis; an overview of the discoveries in b physics to estimate size up its importance is given in Tab. 5.2, summarising the most notable contributions, experiments and colliders. Among all experiments, some of them were especially dedicated to the study of Υs and B mesons, called B *factories*[9], operating exactly at the mass of $\Upsilon(4S)$: PEP-II, Belle [26][10] and HERA-B[11].

5.3 Physics at hadron colliders with b's

At LHC, following up on the physics research at TEVATRON as a hadron collider, properties of b quarks are further investigated in order to refine previous results and validate the

6. This appellation has hardly gone beyond the joke "there is no truth in the SM" while physicists were struggling at finding it.
7. Named after the location of the detector.
8. Named Collider Detector at Fermilab.
9. One should maybe also mention an attempt of B factory that failed, B-TEV, because of termination by the U.S. government.
10. French word for beauty.
11. HERA Beauty.

Lab	Collider	Beams	$\sqrt{s}$	Experiment	Activity	Contributions	References
Fermilab		pN	$\leq 10\,\text{GeV}$	E288		Discovery	[4]
CERN	ISR	pp	$\leq 62\,\text{GeV}$	(only one)	1971-1984	Confirmation.	[13, 14]
DESY	DORIS-I	e^+e^-	$\sim 10\,\text{GeV}$	PLUTO		Confirmation.	[15]
				DASP2		Confirmation.	[16, 17]
				DHHM		Third resonance.	[18]
Cornell	CESR	e^+e^-	$\sim 10\,\text{GeV}$	CLEO	1979-2003	Υ and B spectroscopy and measurement of $\lvert V_{cb}\rvert$.	[27]
				CUSB	1979-1989	*idem*	[28]
SLAC	PEP-I		$\leq 29\,\text{GeV}$	(several)	1980-1990	First measurements of B lifetime.	
DESY	DORIS-II	e^+e^-	$\sim 10\,\text{GeV}$	ARGUS	1982-1992	$B^0 - \bar{B}^0$ mixing. Suspicions toward very massive top quark.	[29]
CERN	SpS̄pS	$p\bar{p}$	$546 - -630\,\text{GeV}$	UA1	1981-1990	$B^0 - \bar{B}^0$ mixing. Searches.	[30, 31, 32]
Fermilab	TEVATRON	$p\bar{p}$	$1.8\,\text{TeV}$	DØ	1983-2011	Production modes. Top quark.	[33, 34, 35, 36]
				CDF	1983-2011	Production modes. Top quark. $B_s - \bar{B}_s$ mixing. Secondary-vertex tagging.	[37, 38, 39]
				B-TEV	(aborted)		
SLAC	SLC	e^+e^-		Mark-II	1982-1985	b fragmentation	
			$\sim 90\,\text{GeV}$	SLD	1992-1998	b fragmentation	
	PEP-II	e^+e^-	$\sim 10\,\text{GeV}$	BaBar	1999-2008	CP violation. Secondary-vertex tagging.	
DESY	HERA	ep	$318\,\text{GeV}$	H1	1992-2007	b mass and running.	[40]
				ZEUS	1992-2007	b mass and running.	[40, 41]
				HERA-B	2002-2003	Novel method of production of B mesons. CP-violation	
CERN	LEP	e^+e^-	$45 - 209\,\text{GeV}$	(several)	1989-2000	B lifetime, b fragmentation.	[42, 43, 44]
KEK	KEKB	e^+e^-	$\sim 10\,\text{GeV}$	Belle	1999-2010	B physics. CP violation.	[45, 46]
CERN	LHC	pp	$7, 8, 13\,\text{TeV}$	CMS	2009-today	B physics. CP violation. Searches.	[47, 48, 49, 50, 51, 52]
				ATLAS	2009-today	*idem*	[47, 53, 54, 55]
				LHCb	2009-today	B physics. CP violation.	[47, 56, 57, 58, 59]
		Pb-Pb		ALICE	2009-today	B-physics in ion collisions.	[60]

TABLE 5.2 – Main experiments contributing to b physics are summarised. The experiments are ordered by collider and starting of activity (many experiments were contemporary). A non-exhaustive list of the contributions of each experiment is given.

current understanding of QCD.

Given its remarkable, long lifetime, b quarks or B mesons are in principle easy to detect in pp collisions. Assuming their properties to be known, they have become an interesting probe for other processes in and beyond the SM.

This section is organised in four subsections: in the first subsection, we define the flavour of jets and investigate tagging techniques; then the mechanism of production of b jets is detailed; finally, in the two last subsections, applications for some important processes where b quarks play a rôle are outlined. Throughout this section, we try to highlight relevant aspects for the analysis presented in Part II.

5.3.1 Heavy-flavour jets

5.3.1.1 Definition

At hadron colliders such as LHC, b quarks are mostly collimated inside of jets. The *flavour* of a jet can be defined as follows:

1. according to the *hadron-flavour definition*, b jets must contain a stable B meson;
2. similarly, c jets must contain a stable D meson, unless already accounted as b jets;
3. and all other jets are defined as light jets (or $udsg$).

(We do not discuss here the possibility to define t jets.) Whether a particle is considered as stable is sometimes matter of debate and the definition at generator level can change:

— At Tevatron and during LHC Run-I, the standard flavour definition was performed at parton level, with b and c quarks.

— During LHC Run-II (as for the present work), the definition was at hadron level, and *stable particles* were defined with respect to all strong and electroweak interactions.

— In other experiments (*e.g.* Zeus [12] and H1 [13]), the flavour is also defined at hadron level, with the difference that the stability is only required for strong and electromagnetic interactions but *not* for weak interactions [61, 62].

Despite these remarks, it is common to keep the notation c jet or b jet (*i.e.* with small letters) even with the hadron-flavour definition; since this is the convention at CMS, we shall follow it here.

5.3.1.2 Tagging

The object of tagging is to identify, or *tag*, b jets, *i.e.* how to discriminate b jets from c and light jets. Different types of discriminants exist, based on different properties of b jets.

Indeed, the intermediate lifetime and masses of B hadrons lead to distinctive kinematic signatures in the tracks produced within b jets. In addition, HF jets have an increased probability to contains leptons.

LHC experiments rely on similar approaches [63, 64, 65, 66], often combined with Multi-Variate Analysis (MVA).

A sketch of an event containing a b jet is shown in Fig. 5.3; the three tagging techniques are illustrated:

12. Named in reference to the relation of Zeus and Hera in the mythology.
13. HERA-1.

secondary vertex The long lifetime of the B mesons imply its decay to be significantly displaced with respect to the PV (but usually still inside of the beam pipe). For instance, at CMS, the mean free path for heavily boosted B's is of the order of $\lambda \approx 2\,\text{mm}$. Most modern taggers rely primarily on the presence of a SV.

impact parameter As a consequence of the presence of a SV, the impact parameter of the tracks with respect to the PV should be significantly larger. A jet quantity based on the impact parameter of its tracks should therefore be sensitive to the presence of a HF quark or hadron.

soft leptons The weak decay of a B meson may lead to a non-isolated lepton in the final state:

$$B \to WX \to Xl\nu \tag{5.8}$$

In practice, in the final state of pp collisions at LHC, Υ are much less likely to take place, but they could also lead to non-isolated leptons in the final state:

$$\Upsilon \to ll \tag{5.9}$$

In general, this technique of tagging is poor, since any weak decay may lead to soft lepton in the final state; however, used in combination with other taggers, a soft-lepton tagger may significantly increase the performance of the discrimination.

The specific taggers in CMS will be further detailed in Sect. 7.4, while performing the analysis.

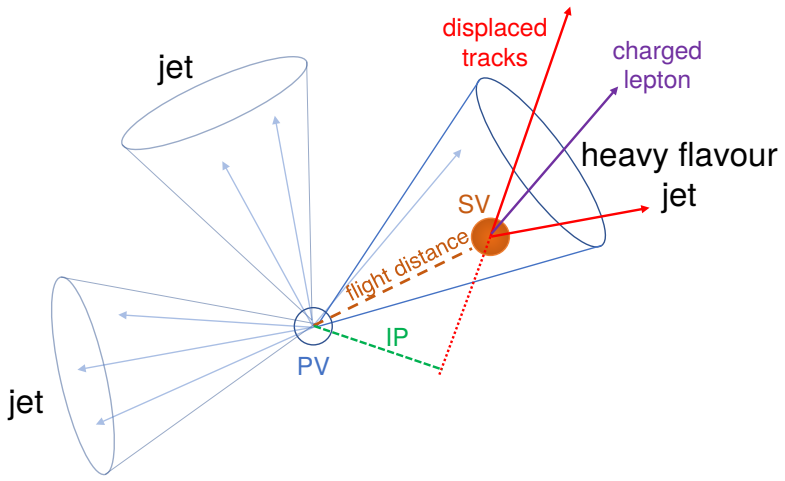


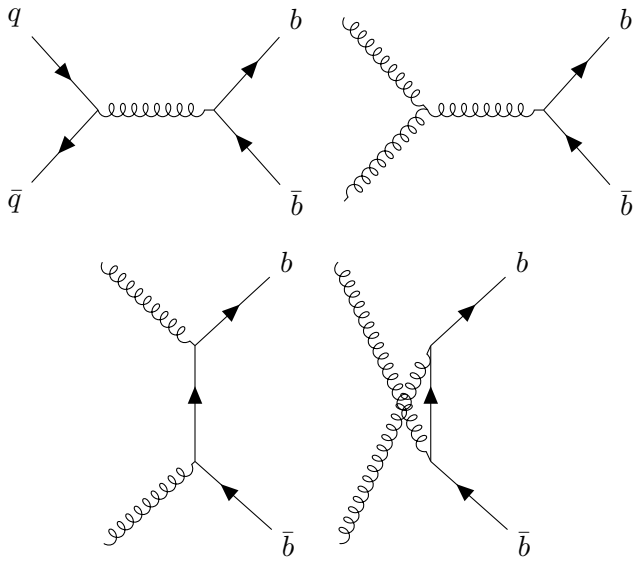
FIGURE 5.3 – Sketch of an event with b jets in pp collisions: the Primary Vertex (PV) is represented at the centre of the sketch with a small circle, while the Secondary Vertex (SV) is represented inside of the HF jet with a plain disc; jets are represented with cones pointing to the centre of the sketch, the HF jet with thicker lines; the tracks are represented with arrows. The three characteristic allowing the distinction of the jets with respect of the flavour are the interaction point (IP), a possibly non-isolated charged lepton in the jet, and the presence of a SV. From [66].

5.3.2 Mechanism of production of b's with QCD processes

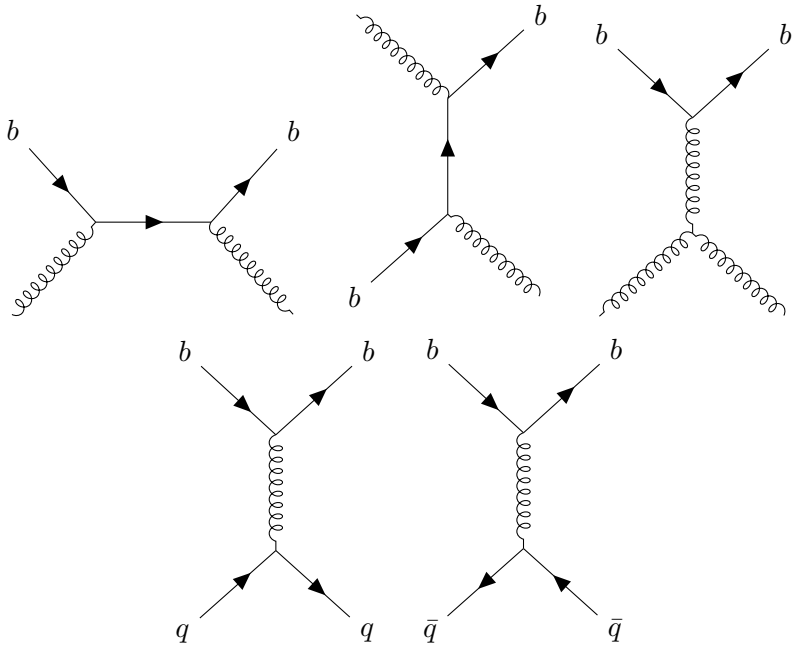
We review the mechanism of production of b quarks.

At Tevatron, three categories of production of b quarks were identified [67] (illustrated in Fig. 5.4):

Flavour Creation (FCR) where the pair of b quarks is produced in the final state of the hard process (Fig. 5.4a). In the initial state can be pairs of gluons or pairs of light (or charm) quarks; at tree-level, it corresponds to the following diagrams:

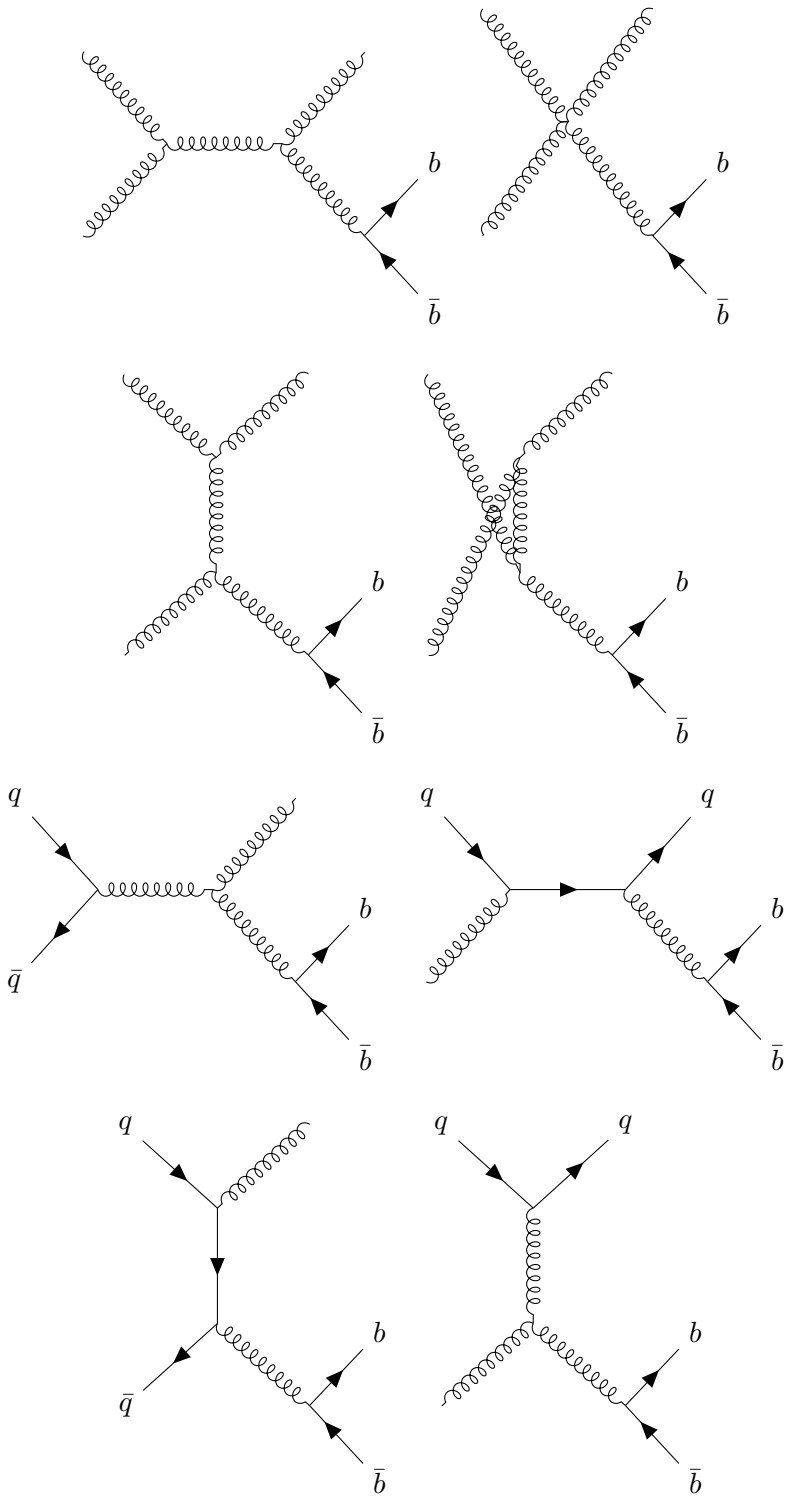


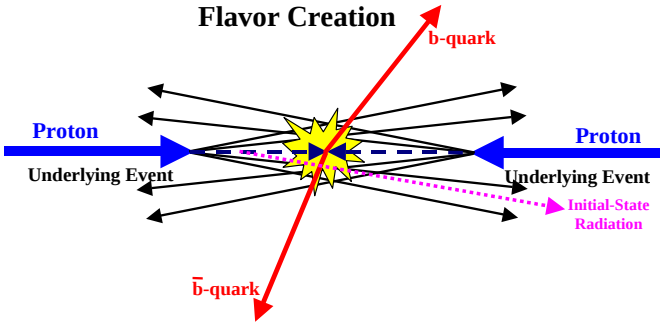
Flavour Excitation (FEX) where a b quark is present in the initial state of the hard process (Fig. 5.4b). A double FEX with two b quarks simultaneously present in the initial state is possible but negligible. The b quark in the initial state is accompanied with a gluon or with a light (or charm) quark; in the 5-flavour scheme the tree-level diagrams are:



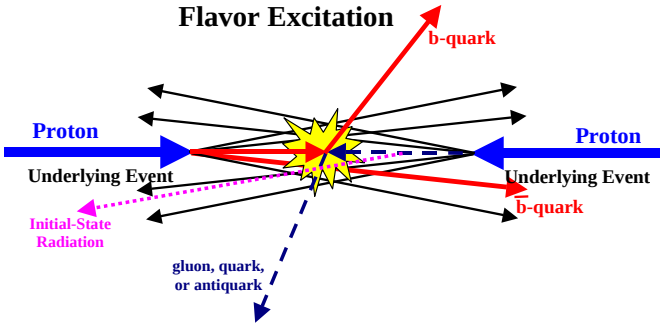
The b quark in the initial state must be issued from the evolution of the PDFs. The evolution is usually performed in the massless approximation, as will be the case in the analysis in Part II. For scales $Q^2 \gg m_b^2$, the mass is not expected not play any significant rôle.

Showering a.k.a. Gluon Splitting (GSP), where a pair of b quarks is issued from the branching of a gluon in the PS. The most relevant tree-level diagrams are the following:

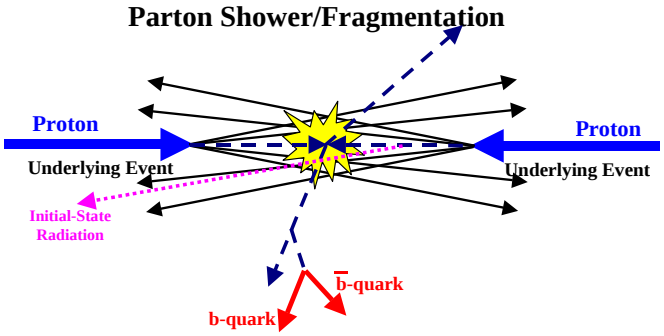




(A) FCR.



(B) FEX.



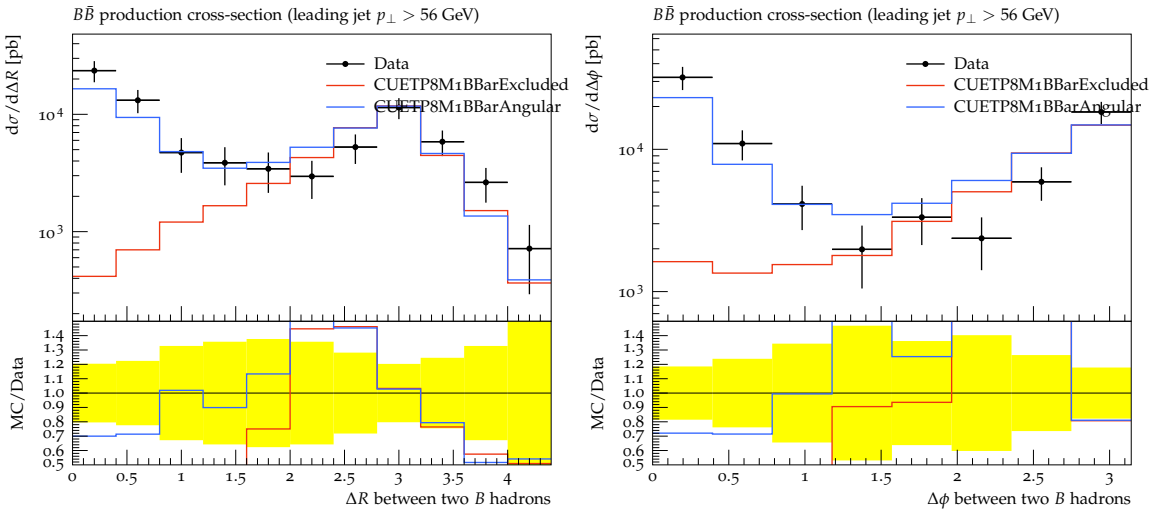
(C) Showering.

FIGURE 5.4 – Production modes of b quarks. The incident protons (or proton-antiproton) are represented horizontally; the hard partons (BBR) are shown with thick (thin) arrows. The dashed (continuous) thick arrows stand for gluons (b quarks). Figure modified from [67].

with q for light (or charm) quark. But in principle, GSP can take place on any gluon branch, for instance in combination with FEX. In the PSs, the mass usually only plays a threshold effect: for scales $Q^2 < (2m_b)^2$, no pairs of b quarks can be produced anymore. The contributions from GSP can be understood from the angular separation of $b\bar{b}$ systems; a simulation of the production of B mesons with and without GSP is compared to CMS data at 7 TeV in Fig. 5.5, and illustrates the important contributions from the PS.

5.3.3 b as a test for QCD

We briefly review three topics of QCD where b quarks could be of interest: the evolution, flavour democracy and tuning.



(A) ΔR (B) $\Delta\phi$

FIGURE 5.5 – "The differential $B\bar{B}$ cross section is measured as a function of the opening angle ΔR and $\Delta\phi$ using data collected with the CMS detector during 2010 and corresponding to an integrated luminosity of $3.1\,\text{pb}^{-1}$. The measurement is performed for three different event energy scales, characterized by the transverse momentum of the leading jet in the event (above 56 GeV, [...]). Simulated events are normalised in the region $\Delta R > 2.4$ and $\Delta\phi > \frac{3\pi}{4}$ respectively." [68] The present plots are performed with PYTHIA 8 through the RIVET interface [69]. The blue (red) line corresponds to the situation where gluon splitting in $b\bar{b}$ in the PS is included (excluded).

5.3.3.1 Evolution

As a significant fraction of the b jets are issued from the PS and from FEX, various measurements of the b jet production can be used to investigate the evolution. For instance, as shown in Fig. 5.6, attempts to use CCFM evolution to describe the inclusive b jet production were performed with data from TEVATRON [70, 71]. In addition to inclusive b jet measurements, measurements of the angular separation of $b\bar{b}$ pairs can help studying the extra radiations and test CCFM evolution.

5.3.3.2 Flavour democracy

Since QCD interactions are not sensitive to the flavour, an important verification of QCD consists in comparing the production of jets of different flavours. In regions of the phase space where the masses become negligible, the cross section should be the same for all flavours. This is illustrated in Fig. 5.7, where the partonic cross sections of b quark and t quark productions are compared in simulation for at $p_T \sim \mathcal{O}(1\,\text{TeV})$; in particular, from 1 TeV, even the mass of the top can be neglected.

Several jet measurements, such as the ones presented in App. C, can be repeated for b jets. First, the inclusive jet and b jet cross section can be compared — this will be presented in Part II of this thesis. In addition, the angular correlations of jets at high multiplicity can also be repeated for b jets, as well as the determination of the strong coupling from a ratio of cross section with three and two jets in the final state.

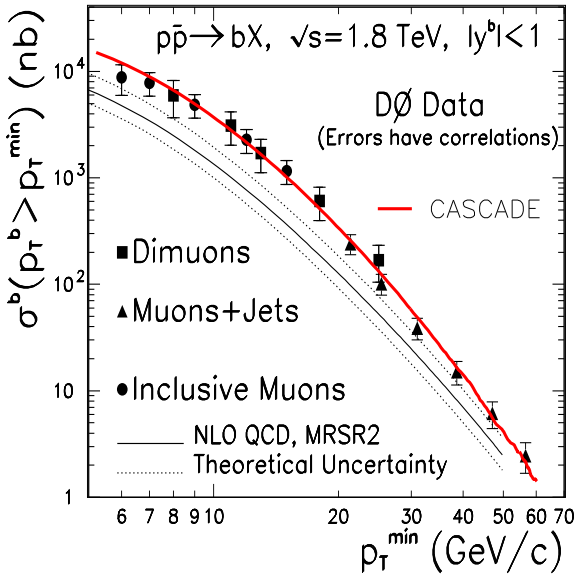


FIGURE 5.6 – "Cross section for $b\bar{b}$ production with $|y^b| < 1$ as a function of $p_T^{\min}$. Shown are the DØ [36] data points, the fixed-order NLO prediction, and the prediction of CASCADE." [71]

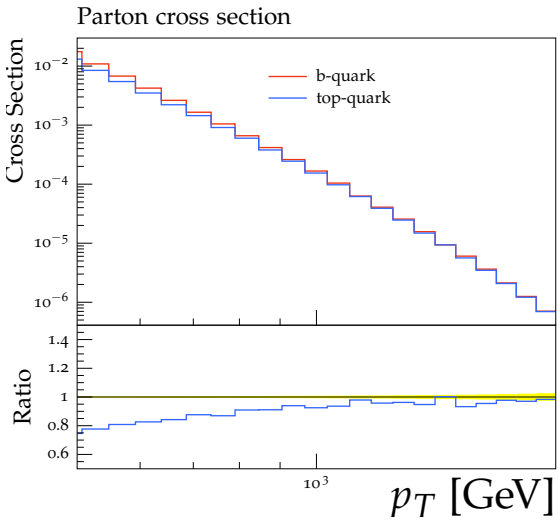


FIGURE 5.7 – Illustration of flavour democracy with b and t quarks. The cross sections are absolute and have been produced with PYTHIA 8; the plots are obtained with the Rivet interface [69].

5.3.3.3 Phenomenology

Finally, the phenomenology of the UE, as described in Sect. 2.3, has to be tested, or the parameters of the current tunes need to be verified or refined. For instance, as we saw in Sect. 2.3.3.4, fragmentation requires dedicated parameters for heavy quarks.

In principle, tuning (already introduced in Sect. 4.2.2.2) may be performed with any distributions that is sensitive to the soft activity. Typically, measurements of angular separation of $B\bar{B}$ pairs of hadrons would be crucial to improve the description of hadronisation, which, as we mentioned in Sect. 4.2.2.2, still relies on LEP measurements.

5.3.4 b as a probe for other processes

Although we are not concerned with other SM or BSM processes, the b jets often have a privileged rôle, since they are rather easy to detect. In order to illustrate it, we give here a short review of top physics, Higgs physics and searches.

5.3.4.1 Top physics

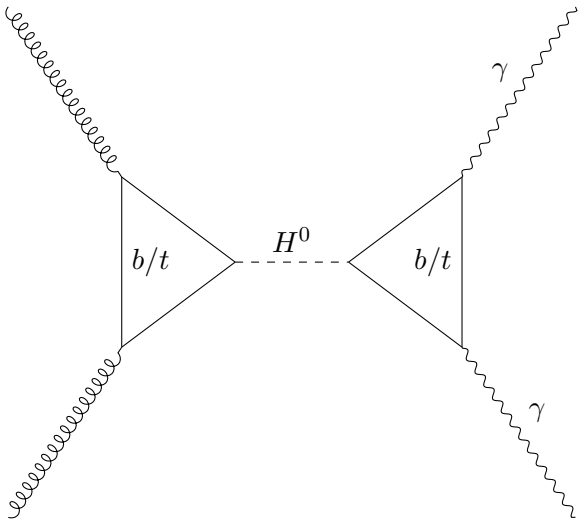
The top quark decays with a lifetime much shorter than the time scale for hadronisation. Thus, the top quark may be studied as a bare quark (*i.e.* without the complications related to hadronisation in the study of all other flavours). The main channel to measure it is the weak decay:

$$t \to Wb \tag{5.10}$$

It is therefore crucial to know well and detect efficiently b quarks (or B mesons) [72, 73, 74]. In addition, the W itself can decay hadronically, potentially giving another b quark.

5.3.4.2 Higgs physics

The Higgs boson was already mentioned in Chap. 2. Originally, the Higgs boson was discovered using the $H \to \gamma\gamma$ channel [75, 76]:



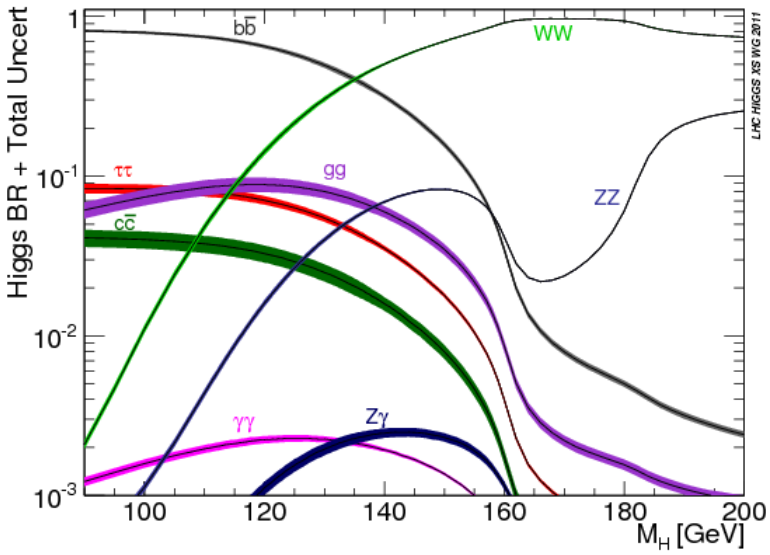
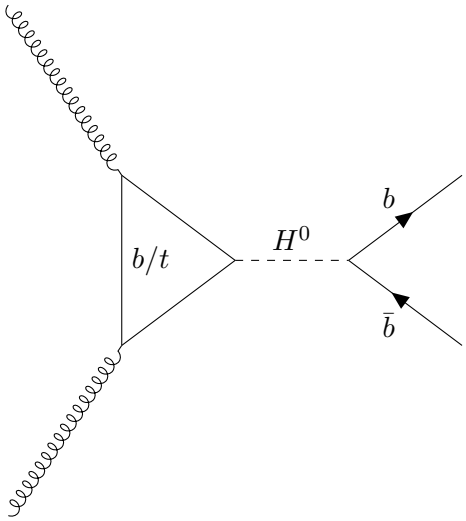


FIGURE 5.8 – Higgs branching ratios as a function of the mass of the Higgs boson [82]. The Golden Channel, involving b's, has an important branching ratio. This plot is historically important since it drove the design of LHC experiments in the quest of the Higgs boson.

However, from Fig. 5.8, one can see that the $H \to b\bar{b}$ channel has a much larger branching ratio:



Despite the much larger branching ratio, there is « only *evidence* » for this decay channel [14], which is due to overwhelming $b\bar{b}$ background.

Currently, many Higgs analyses rely on the presence of b jets in the final state. Only at CMS, one can cite many analyses [77, 78, 79, 80, 81].

5.3.4.3 Searches

b jets are heavily used in searches, either as a background or as part of the signal. In both cases, tagging techniques make b jets privileged objects since they are easy to manipulate, with a rather high efficiency. At CMS, one can mention for instance searches for heavy vector bosons [83], searches for light Higgs bosons [84] or searches for leptoquarks [85].

14. Observation conventionally means at least five sigma of significance.

This thesis does not treat on searches and extensions of the SM; therefore, we shall not detail them.

Conclusion

There can hardly be any doubt about the importance of the beauty quark in physics at LHC: top and Higgs physics, searches, hadronisation, calculations, PDFs. A deep understanding of its properties is therefore essential, and motivates precision measurements. At CMS for LHC Run-II, the b jet production is one of the first observables to investigate.

References

[1] Edward H. Thorndike. "Bottom quark physics: Past, present, future". In: *Probing luminous and dark matter. Proceedings, Symposium in Honor of Adrian Melissinos, Rochester, USA, September 24-25, 1999*. 1999, pp. 127–159. arXiv: hep-ex/0003027 [hep-ex].

[2] J. A. Appel et al. "A Study of Di-Lepton Production in Proton Collisions at NAL". In: (1974).

[3] Fermilab. "Discoveries at Fermilab - Discovery of the Bottom Quark". 1977. URL: http://lepewwg.web.cern.ch/LEPEWWG.

[4] S. W. Herb et al. "Observation of a Dimuon Resonance at 9.5 GeV in 400-GeV Proton-Nucleus Collisions". In: *Phys. Rev. Lett.* 39 (5 Aug. 1977), pp. 252–255. DOI: 10.1103/PhysRevLett.39.252. URL: https://link.aps.org/doi/10.1103/PhysRevLett.39.252.

[5] Fermilab History and Archives Project. "Discovery of the bottom quark, Upsilon". Accessed on 3 December 2017. URL: https://history.fnal.gov/botqrk.html.

[6] CERN Courrier. "Revisiting the b revolution". 2017. URL: http://cerncourier.com/cws/article/cern/68794.

[7] J. E. Augustin et al. "Discovery of a Narrow Resonance in e^+e^- Annihilation". In: *Phys. Rev. Lett.* 33 (1974). [Adv. Exp. Phys.5,141(1976)], pp. 1406–1408. DOI: 10.1103/PhysRevLett.33.1406.

[8] J. J. Aubert et al. "Experimental Observation of a Heavy Particle J". In: *Phys. Rev. Lett.* 33 (1974), pp. 1404–1406. DOI: 10.1103/PhysRevLett.33.1404.

[9] B.J. Bjørken and S.L. Glashow. "Elementary particles and SU(4)". In: *Physics Letters* 11.3 (1964), pp. 255–257. ISSN: 0031-9163. DOI: https://doi.org/10.1016/0031-9163(64)90433-0. URL: http://www.sciencedirect.com/science/article/pii/0031916364904330.

[10] S. L. Glashow, J. Iliopoulos, and L. Maiani. "Weak Interactions with Lepton-Hadron Symmetry". In: *Phys. Rev. D* 2 (7 Oct. 1970), pp. 1285–1292. DOI: 10.1103/PhysRevD.2.1285. URL: https://link.aps.org/doi/10.1103/PhysRevD.2.1285.

[11] Thomas Appelquist and H. David Politzer. "Heavy Quarks and e^+e^- Annihilation". In: *Phys. Rev. Lett.* 34 (1 Jan. 1975), pp. 43–45. DOI: 10.1103/PhysRevLett.34.43. URL: https://link.aps.org/doi/10.1103/PhysRevLett.34.43.

[12] Makoto Kobayashi and Toshihide Maskawa. "CP-Violation in the Renormalizable Theory of Weak Interaction". In: *Progress of Theoretical Physics* 49.2 (1973), pp. 652–657. DOI: 10.1143/PTP.49.652. eprint: /oup/backfile/content_public/journal/ptp/49/2/10.1143/ptp.49.652/2/49-2-652.pdf. URL: +%20http://dx.doi.org/10.1143/PTP.49.652.

[13] Christian W. Fabjan and Norman McCubbin. "Physics at the CERN Intersecting Storage Rings (ISR) 1978 – 1983". In: *Physics Reports* 403-404 (2004), pp. 165–175. ISSN: 0370-1573. DOI: https://doi.org/10.1016/j.physrep.2004.08.018. URL: http://www.sciencedirect.com/science/article/pii/S0370157304003308.

[14] J.H. Cobb et al. "The cross section for the production of massive electron pairs and the Υ (9.5 GeV) in proton-proton collisions at the CERN ISR". In: *Physics Letters B* 72.2 (1977), pp. 273–277. ISSN: 0370-2693. DOI: https://doi.org/10.1016/0370-2693(77)90720-1. URL: http://www.sciencedirect.com/science/article/pii/0370269377907201.

[15] Ch. Berger et al. "Observation of a narrow resonance formed in e+e annihilation at 9.46 GeV". In: *Physics Letters B* 76.2 (1978), pp. 243–245. ISSN: 0370-2693. DOI: https://doi.org/10.1016/0370-2693(78)90287-3. URL: http://www.sciencedirect.com/science/article/pii/0370269378902873.

[16] C.W. Darden et al. "Observation of a narrow resonance at 9.46 GeV in electron - positron annihilations". In: *Physics Letters B* 76.2 (1978), pp. 246–248. ISSN: 0370-2693. DOI: https://doi.org/10.1016/0370-2693(78)90288-5. URL: http://www.sciencedirect.com/science/article/pii/0370269378902885.

[17] C.W. Darden et al. "Study of the (9.46) meson in electron-positron annihilations". In: *Physics Letters B* 80.4 (1979), pp. 419–422. ISSN: 0370-2693. DOI: https://doi.org/10.1016/0370-2693(79)91204-8. URL: http://www.sciencedirect.com/science/article/pii/0370269379912048.

[18] J.K. Bienlein et al. "Observation of a narrow resonance at 10.02 GeV in e+e annihilations". In: *Physics Letters B* 78.2 (1978), pp. 360–363. ISSN: 0370-2693. DOI: https://doi.org/10.1016/0370-2693(78)90040-0. URL: http://www.sciencedirect.com/science/article/pii/0370269378900400.

[19] E. Eichten et al. "Charmonium: Comparison with Experiment". In: *Phys. Rev.* D21 (1980), p. 203. DOI: 10.1103/PhysRevD.21.203.

[20] P. Franzini and J. Lee-Franzini. "Upsilon physics at CESR". In: *Physics Reports* 81.3 (1982), pp. 239–291. ISSN: 0370-1573. DOI: https://doi.org/10.1016/0370-1573(82)90027-8. URL: http://www.sciencedirect.com/science/article/pii/0370157382900278.

[21] D. Andrews et al. "Observation of Three Upsilon States". In: *Phys. Rev. Lett.* 44 (17 Apr. 1980), pp. 1108–1111. DOI: 10.1103/PhysRevLett.44.1108. URL: https://link.aps.org/doi/10.1103/PhysRevLett.44.1108.

[22] D. Andrews et al. "Observation of a Fourth Upsilon State in e^+e^- Annihilations". In: *Phys. Rev. Lett.* 45 (4 July 1980), pp. 219–221. DOI: 10.1103/PhysRevLett.45.219. URL: https://link.aps.org/doi/10.1103/PhysRevLett.45.219.

[23] G. Finocchiaro et al. "Observation of the Upsilon"' at CESR". In: *Phys. Rev. Lett.* 45 (1980), p. 222. DOI: 10.1103/PhysRevLett.45.222.

[24] C. Patrignani et al. "Review of Particle Physics". In: *Chin. Phys.* C40.10 (2016), p. 100001. DOI: 10.1088/1674-1137/40/10/100001.

[25] D. Besson and T. Skwarnicki. "υ spectroscopy". In: *Ann. Rev. Nucl. Part. Sci.* 43 (1993), pp. 333–378. DOI: 10.1146/annurev.ns.43.120193.002001.

[26] T. Abe et al. "Belle II Technical Design Report". In: (2010). arXiv: 1011.0352 [physics.ins-det].

[27] David G. Cassel. "CLEO B physics". In: *ARGUS Fest, 20 years of B meson mixing 1987-2007. Proceedings, ARGUS-symposium, DESY, Hamburg, Germany, November 9, 2007.* 2007. DOI: 10.3204/DESY-PROC-2008-01/e307.

[28] Juliet Lee-Franzini. "Hidden and open beauty in CUSB". In: *AIP Conf. Proc.* 424.1 (1998), pp. 85–96. DOI: 10.1063/1.55115. arXiv: hep-ex/9709025 [hep-ex].

[29] H Albrecht et al. "Argus: A universal detector at DORIS II". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 275.1 (1989), pp. 1–48.

[30] C Albajar et al. "Beauty production at the CERN proton-antiproton collider". In: 186 (Mar. 1987), pp. 237–246.

[31] C. Albajar et al. "Beauty Production at the CERN Proton - anti-Proton Collider. 1." In: *Phys. Lett.* B186 (1987), pp. 237–246. DOI: 10.1016/0370-2693(87)90287-5.

[32] C. Albajar et al. "Measurement of the Bottom Quark Production Cross-Section in Proton - anti-Proton Collisions at s**(1/2) = 0.63-TeV". In: *Phys. Lett.* B213 (1988). [,473(1988)], p. 405. DOI: 10.1016/0370-2693(88)91785-6.

[33] V. M. Abazov et al. "b-Jet Identification in the D0 Experiment". In: *Nucl. Instrum. Meth.* A620 (2010), pp. 490–517. DOI: 10.1016/j.nima.2010.03.118. arXiv: 1002.4224 [hep-ex].

[34] F. Abe et al. "Inclusive jet cross section in $\bar{p}p$ collisions at $\sqrt{s} = 1.8$ TeV". In: *Phys. Rev. Lett.* 77 (1996), pp. 438–443. DOI: 10.1103/PhysRevLett.77.438. arXiv: hep-ex/9601008 [hep-ex].

[35] B. Abbott et al. "Cross section for b jet production in $\bar{p}p$ collisions at $\sqrt{s} = 1.8$ TeV". In: *Phys. Rev. Lett.* 85 (2000), pp. 5068–5073. DOI: 10.1103/PhysRevLett.85.5068. arXiv: hep-ex/0008021 [hep-ex].

[36] B Abbott et al. "Cross Section for b-Jet Production in p p Collisions at √ s= 1.8 TeV". In: *Physical Review Letters* 85.24 (2000), p. 5068.

[37] F. Abe et al. "Measurement of bottom quark production in 1.8 TeV $p\bar{p}$ collisions using semileptonic decay muons". In: *Phys. Rev. Lett.* 71 (1993), pp. 2396–2400. DOI: 10.1103/PhysRevLett.71.2396.

[38] F. Abe et al. "Measurement of the bottom quark production cross-section using semileptonic decay electrons in $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ TeV". In: *Phys. Rev. Lett.* 71 (1993), pp. 500–504. DOI: 10.1103/PhysRevLett.71.500.

[39] F. Abe et al. "Measurement of the B meson differential cross-section, $d\sigma/dp_T$, in $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ TeV". In: *Phys. Rev. Lett.* 75 (1995), pp. 1451–1455. DOI: 10.1103/PhysRevLett.75.1451. arXiv: hep-ex/9503013 [hep-ex].

[40] O. Behnke, A. Geiser, and M. Lisovyi. "Charm, Beauty and Top at HERA". In: *Prog. Part. Nucl. Phys.* 84 (2015), pp. 1–72. DOI: 10.1016/j.ppnp.2015.06.002. arXiv: 1506.07519 [hep-ex].

[41] S. Chekanov et al. "Measurement of beauty production from dimuon events at HERA". In: *JHEP* 02 (2009), p. 032. DOI: 10.1088/1126-6708/2009/02/032. arXiv: 0811.0894 [hep-ex].

[42] Yoram Rozen, François Pierre, and Giovanni Calderini. "B physics at LEP". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 446.1 (2000), pp. 37–52. ISSN: 0168-9002. DOI: https://doi.org/10.1016/S0168-9002(00)00016-4. URL: http://www.sciencedirect.com/science/article/pii/S0168900200000164.

[43] Yoram Rozen. "The legacy of LEP B program". In: *Nuclear Physics B - Proceedings Supplements* 120.Supplement C (2003). Proceedings of the 8th International Conference on B-Physics at Hadron Machines, pp. 83–90. ISSN: 0920-5632. DOI: https://doi.org/10.1016/S0920-5632(03)01885-1. URL: http://www.sciencedirect.com/science/article/pii/S0920563203018851.

[44] J. Abdallah et al. "A study of the b-quark fragmentation function with the DELPHI detector at LEP I and an averaged distribution obtained at the Z Pole". In: *Eur. Phys. J.* C71 (2011), p. 1557. DOI: 10.1140/epjc/s10052-011-1557-x. arXiv: 1102.4748 [hep-ex].

[45] K. Abe et al. "Observation of Large CP Violation and Evidence for Direct CP Violation in $B^0 \to \pi^+\pi^-$ Decays". In: *Phys. Rev. Lett.* 93 (2 July 2004), p. 021601. DOI: 10.1103/PhysRevLett.93.021601. URL: https://link.aps.org/doi/10.1103/PhysRevLett.93.021601.

[46] Y. Chao et al. "Evidence for Direct CP Violation in $B^0 \to K^+\pi^-$ Decays". In: *Phys. Rev. Lett.* 93 (19 Nov. 2004), p. 191802. DOI: 10.1103/PhysRevLett.93.191802. URL: https://link.aps.org/doi/10.1103/PhysRevLett.93.191802.

[47] T. Gershon and M. Needham. "Heavy Flavour Physics at the LHC". In: *Comptes Rendus Physique* 16 (2015), pp. 435–447. DOI: 10.1016/j.crhy.2015.04.001. arXiv: 1408.0403 [hep-ex].

[48] Serguei Chatrchyan et al. "Inclusive b-jet production in pp collisions at $\sqrt{s} = 7$ TeV". In: *JHEP* 04 (2012), p. 084. DOI: 10.1007/JHEP04(2012)084. arXiv: 1202.4617 [hep-ex].

[49] Vardan Khachatryan et al. "Measurement of the B^+ Production Cross Section in pp Collisions at $\sqrt{s} = 7$ TeV". In: *Phys. Rev. Lett.* 106 (2011), p. 112001. DOI: 10.1103/PhysRevLett.106.112001. arXiv: 1101.0131 [hep-ex].

[50] Vardan Khachatryan et al. "Inclusive b-hadron production cross section with muons in pp collisions at $\sqrt{s} = 7$ TeV". In: *JHEP* 03 (2011), p. 090. DOI: 10.1007/JHEP03(2011)090. arXiv: 1101.3512 [hep-ex].

[51] Serguei Chatrchyan et al. "Measurement of the B^0 production cross section in pp Collisions at $\sqrt{s} = 7$ TeV". In: *Phys. Rev. Lett.* 106 (2011), p. 252001. DOI: 10.1103/PhysRevLett.106.252001. arXiv: 1104.2892 [hep-ex].

[52] Serguei Chatrchyan et al. "Measurement of the Strange B Meson Production Cross Section with J/Psi ϕ Decays in pp Collisions at $\sqrt{s} = 7$ TeV". In: *Phys. Rev.* D84 (2011), p. 052008. DOI: 10.1103/PhysRevD.84.052008. arXiv: 1106.4048 [hep-ex].

[53] Georges Aad et al. "Measurement of the inclusive and dijet cross-sections of b^- jets in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector". In: *Eur. Phys. J.* C71 (2011), p. 1846. DOI: 10.1140/epjc/s10052-011-1846-4. arXiv: 1109.6833 [hep-ex].

[54] Georges Aad et al. "Measurement of the b-hadron production cross section using decays to $D^*\mu^- X$ final states in pp collisions at sqrt(s) = 7 TeV with the ATLAS detector". In: *Nucl. Phys.* B864 (2012), pp. 341–381. DOI: 10.1016/j.nuclphysb.2012.07.009. arXiv: 1206.3122 [hep-ex].

[55] Georges Aad et al. "Measurement of the differential cross-section of B^+ meson production in pp collisions at $\sqrt{s} = 7$ TeV at ATLAS". In: *JHEP* 10 (2013), p. 042. DOI: 10.1007/JHEP10(2013)042. arXiv: 1307.0126 [hep-ex].

[56] R. Aaij et al. "Measurement of $\sigma(pp \to b\bar{b}X)$ at $\sqrt{s} = 7$ TeV in the forward region". In: *Phys. Lett.* B694 (2010), pp. 209–216. DOI: 10.1016/j.physletb.2010.10.010. arXiv: 1009.2731 [hep-ex].

[57] R. Aaij et al. "Measurement of the $B^\pm$ production cross-section in pp collisions at $\sqrt{s} = 7$ TeV". In: *JHEP* 04 (2012), p. 093. DOI: 10.1007/JHEP04(2012)093. arXiv: 1202.4812 [hep-ex].

[58] R Aaij et al. "Measurement of B meson production cross-sections in proton-proton collisions at $\sqrt{s} = 7$ TeV". In: *JHEP* 08 (2013), p. 117. DOI: 10.1007/JHEP08(2013)117. arXiv: 1306.3663 [hep-ex].

[59] Roel Aaij et al. "Measurement of B_c^+ production in proton-proton collisions at $\sqrt{s} = 8$ TeV". In: *Phys. Rev. Lett.* 114 (2015), p. 132001. DOI: 10.1103/PhysRevLett.114.132001. arXiv: 1411.2943 [hep-ex].

[60] Alessandro Grelli and the ALICE Collaboration. "Heavy Flavour Physics with the ALICE detector at the CERN-LHC". In: *Journal of Physics: Conference Series* 316.1 (2011), p. 012025. URL: http://stacks.iop.org/1742-6596/316/i=1/a=012025.

[61] Achim Geiser. *Definition of stable hadron in ZEUS*. Private communication.

[62] Hannes Jung. *Definition of stable hadron in H1*. Private communication.

[63] Luca Scodellaro. "b tagging in ATLAS and CMS". In: *5th Large Hadron Collider Physics Conference (LHCP 2017) Shanghai, China, May 15-20, 2017*. 2017. arXiv: 1709.01290 [hep-ex]. URL: https://inspirehep.net/record/1621595/files/arXiv:1709.01290.pdf.

[64] Georges Aad et al. "Performance of b-Jet Identification in the ATLAS Experiment". In: *JINST* 11.04 (2016), P04008. DOI: 10.1088/1748-0221/11/04/P04008. arXiv: 1512.01094 [hep-ex].

[65] Serguei Chatrchyan et al. "Identification of b-quark jets with the CMS experiment". In: *JINST* 8 (2013), P04013. DOI: 10.1088/1748-0221/8/04/P04013. arXiv: 1211.4462 [hep-ex].

[66] CMS Collaboration. "Identification of b quark jets at the CMS Experiment in the LHC Run 2". In: (2016).

[67] R. D. Field. "The Sources of b quarks at the Tevatron and their correlations". In: *Phys. Rev.* D65 (2002), p. 094006. DOI: 10.1103/PhysRevD.65.094006. arXiv: hep-ph/0201112 [hep-ph].

[68] Vardan Khachatryan et al. "Measurement of Dijet Angular Distributions and Search for Quark Compositeness in pp Collisions at $sqrts = 7$ TeV". In: *Phys. Rev. Lett.* 106 (2011), p. 201804. DOI: 10.1103/PhysRevLett.106.201804. arXiv: 1102.2020 [hep-ex].

[69] Andy Buckley et al. "Rivet user manual". In: *Comput. Phys. Commun.* 184 (2013), pp. 2803–2819. DOI: 10.1016/j.cpc.2013.05.021. arXiv: 1003.0694 [hep-ph].

[70] H. Jung. "Heavy quark production at the TEVATRON and HERA using k_t factorization with CCFM evolution". In: *Phys. Rev.* D65 (2002), p. 034015. DOI: 10.1103/PhysRevD.65.034015. arXiv: hep-ph/0110034 [hep-ph].

[71] Hannes Jung. "Unintegrated parton densities applied to heavy quark production in the CCFM approach". In: *J. Phys.* G28 (2002), pp. 971–982. DOI: 10.1088/0954-3899/28/5/320. arXiv: hep-ph/0109146 [hep-ph].

[72] Albert M Sirunyan et al. "Cross section measurement of t-channel single top quark production in pp collisions at $\sqrt{s} = 13$ TeV". In: *Phys. Lett.* B772 (2017), pp. 752–776. DOI: 10.1016/j.physletb.2017.07.047. arXiv: 1610.00678 [hep-ex].

[73] Albert M. Sirunyan et al. "Measurement of the $t\bar{t}$ production cross section using events with one lepton and at least one jet in pp collisions at $\sqrt{s} = 13$ TeV". In: *JHEP* 09 (2017), p. 051. DOI: 10.1007/JHEP09(2017)051. arXiv: 1701.06228 [hep-ex].

[74] Vardan Khachatryan et al. "Measurement of differential cross sections for top quark pair production using the lepton+jets final state in proton-proton collisions at 13 TeV". In: *Phys. Rev.* D95.9 (2017), p. 092001. DOI: 10.1103/PhysRevD.95.092001. arXiv: 1610.04191 [hep-ex].

[75] Serguei Chatrchyan et al. "Search for the standard model Higgs boson decaying to bottom quarks in pp collisions at $\sqrt{s} = 7$ TeV". In: *Phys. Lett.* B710 (2012), pp. 284–306. DOI: 10.1016/j.physletb.2012.02.085. arXiv: 1202.4195 [hep-ex].

[76] M. Aaboud et al. "Evidence for the $H \to b\bar{b}$ decay with the ATLAS detector". In: *JHEP* 12 (2017), p. 024. DOI: 10.1007/JHEP12(2017)024. arXiv: 1708.03299 [hep-ex].

[77] CMS Collaboration. "Search for resonant pair production of Higgs bosons decaying to two bottom quark-antiquark pairs in proton-proton collisions at 13 TeV". In: (2016).

[78] *Search for resonant pair production of Higgs bosons decaying to bottom quark-antiquark pairs in proton-proton collisions at 13 TeV*. Tech. rep. CMS-PAS-HIG-17-009. Geneva: CERN, 2017. URL: http://cds.cern.ch/record/2292044.

[79] CMS Collaboration. "Evidence for the decay of the Higgs Boson to Bottom Quarks". In: (2017).

[80] *Search for Higgs boson pair production in the final state containing two photons and two bottom quarks in proton-proton collisions at* $\sqrt{s} = 13$ TeV. Tech. rep. CMS-PAS-HIG-17-008. Geneva: CERN, 2017. URL: http://cds.cern.ch/record/2273383.

[81] *Inclusive search for the standard model Higgs boson produced in pp collisions at* $\sqrt{s} = 13$ TeV *using* $H\to$ bb *decays*. Tech. rep. CMS-PAS-HIG-17-010. Geneva: CERN, 2017. URL: http://cds.cern.ch/record/2266164.

[82] A. Denner et al. "Standard Model Higgs-Boson Branching Ratios with Uncertainties". In: *Eur. Phys. J.* C71 (2011), p. 1753. DOI: 10.1140/epjc/s10052-011-1753-8. arXiv: 1107.5909 [hep-ph].

[83] Serguei Chatrchyan et al. "Search for a W ' boson decaying to a bottom quark and a top quark in pp collisions at $\sqrt{s} = 7$ TeV". In: *Phys. Lett.* B718 (2013), pp. 1229–1251. DOI: 10.1016/j.physletb.2012.12.008. arXiv: 1208.0956 [hep-ex].

[84] Vardan Khachatryan et al. "Search for neutral MSSM Higgs bosons decaying into a pair of bottom quarks". In: *JHEP* 11 (2015), p. 071. DOI: 10.1007/JHEP11(2015)071. arXiv: 1506.08329 [hep-ex].

[85] Serguei Chatrchyan et al. "Search for pair production of third-generation leptoquarks and top squarks in pp collisions at $\sqrt{s} = 7$ TeV". In: *Phys. Rev. Lett.* 110.8 (2013), p. 081801. DOI: 10.1103/PhysRevLett.110.081801. arXiv: 1210.5629 [hep-ex].

Part II

Physics analysis

6

General strategy and outline of the analysis

THE GENERAL STRATEGY FOR THE MEASUREMENT of the double differential cross section of inclusive b jet production using pp collisions with $\sqrt{s} = 13\,\text{TeV}$ recorded with the CMS detector is presented here.

First, the signal of the inclusive b jet production is studied in simulation. Second, previous inclusive b jet analyses at LHC are reviewed. Lastly, following on the conclusions from studies in simulation and previous analyses, the strategy of the measurement is outlined.

6.1 Monte Carlo studies

In this section, the contributions from QCD and from other processes is described from the simulation of the ME and of the PS with PYTHIA 8 in LHC conditions with `NNPDF 2.3` set [1] (shown and compared to other sets in Fig. 6.3).

6.1.1 Contributions from QCD processes

The QCD diagrams of the production of b quarks were discussed in Sect. 5.3.2.

First, the contributions from b jets to inclusive jet cross section is shown in Fig. 6.1. In the top row of this figure, the inclusive jet production is described in terms of partons in the initial state of the hard process: two quarks (blue), two gluons (green) or a pair quark-gluon (red). The contribution from b jets is shown in darker shades in the top row and alone in the bottom row; the contribution from gg in the initial state is significantly reduced in comparison to the two other initial states.

In Fig. 6.2, the contributions to inclusive b jets cross section are shown in terms of FCR, FEX, GSP. The FCR contributions are the less important one with around 10% of the signal, while the GSP contributions are getting more and more important at high transverse momentum of the b jets and represent $50 - 80\%$.

From this studies in simulation, one concludes that the contribution from PS is important to describe the inclusive b jet production.

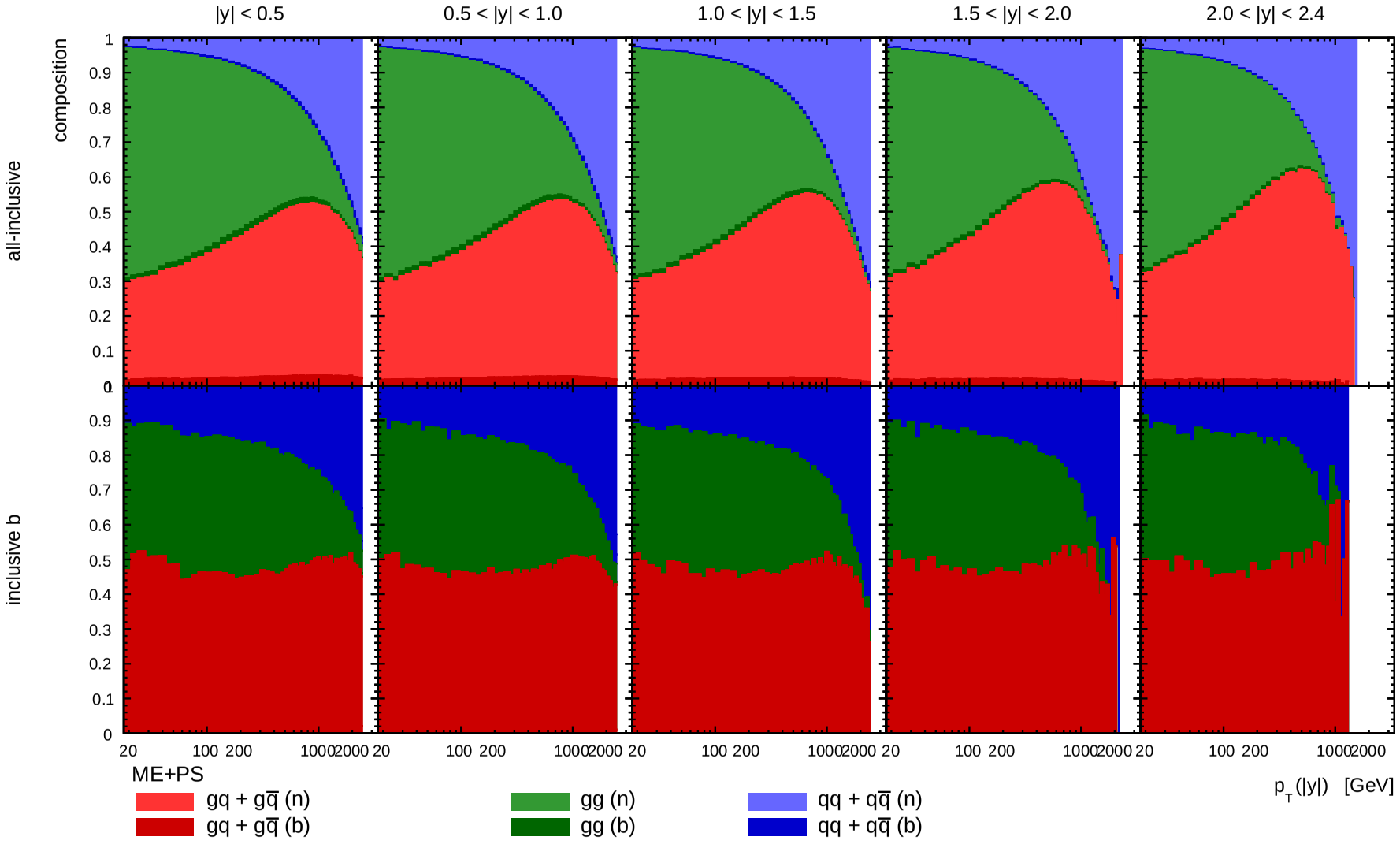


FIGURE 6.1 – Contributions to inclusive jet (top) and inclusive b jet (bottom) cross section at parton level as a function of the transverse momentum of the jets, as described by PYTHIA 8 (ME+PS) with `NNPDF 2.3` set [1]. Blue (green; red) corresponds to processes with two quarks (two gluons; one gluon and one quark) in the initial state; the dark (light) shade corresponds to b jets (non-b jets, or n jets). The five columns correspond to rapidity bins.

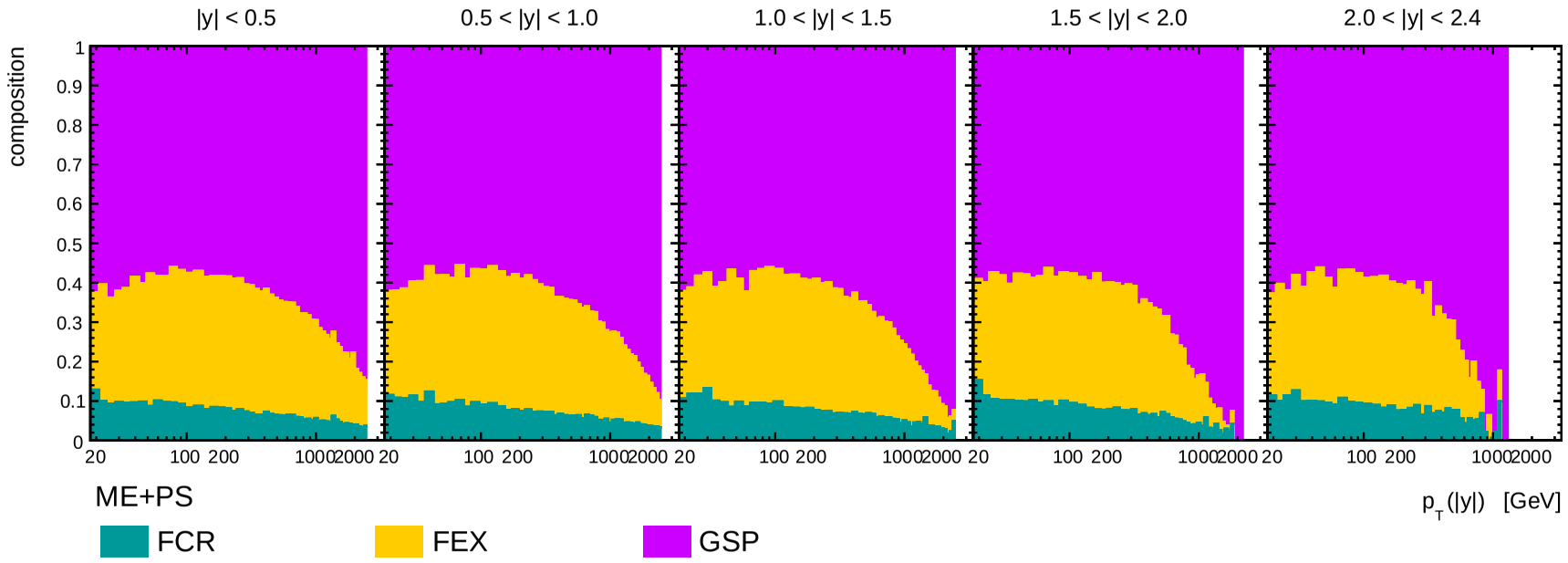


FIGURE 6.2 – Contributions to inclusive b jet cross section at parton level as a function of the transverse momentum of the jets, as described by PYTHIA 8 (ME+PS) with `NNPDF 2.3` set [1]. Violet (cyan; yellow) corresponds to FCR (FEX; GSP). The five columns correspond to rapidity bins.

6.1.2 Other processes

According to the SM, quarks have both strong and, in a lesser extent, electroweak interactions. In this subsection, we investigate the different channels contributing to the inclusive b jet production.

In addition to the QCD processes, the processes listed below can make significant contributions to the inclusive b jet cross section.

$$t\bar{t} \to b\bar{b}X \qquad (\mathrm{t\bar{t}}) \tag{6.1}$$
$$W \to bX \qquad (\mathrm{W}) \tag{6.2}$$
$$Z \to b\bar{b}X \qquad (\mathrm{DY}) \tag{6.3}$$

In principle, $t\bar{t}$ should be considered as part of QCD: however, given the specific properties of the top quark, it is considered separately here. Other signals like electroweak production of single top quarks, or Higgs production have negligible contributions and are not considered.

The cross sections for the different processes are shown in Fig. 6.4 as a function of the transverse momentum of the b jets in the central and forward bins of (absolute) rapidity at generator level. The predictions are calculated with PYTHIA 8 and jets are clustered with the anti-k_T algorithm with cone size radius $R = 0.4$.

Here, the predictions do not include any simulation of the detector (like resolution and efficiency). Up to 1 TeV, the standard QCD processes dominate.

6.2 Previous measurements of the inclusive b jet production at the LHC

The latest measurements of inclusive b jet production in ATLAS [6] and CMS [7] are now reviewed. In both cases, the double differential cross section in transverse momentum and absolute rapidity was measured.

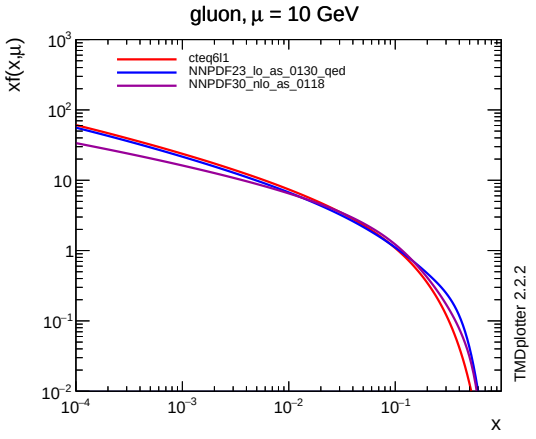
Both collaborations performed the analysis once with soft-lepton tagging and once with SV tagging; the four measurements are performed with anti-k_T jet clustering algorithm with $R = 0.5$ and corrected to parton level. The measurements are represented and compared with one another and with theoretical predictions from MC@NLO in Fig. 6.5; however, the measurements can only be compared in a single, inclusive bin of rapidity, since different rapidity binning schemes are used by each collaboration.

6.2.1 ATLAS at 7 TeV

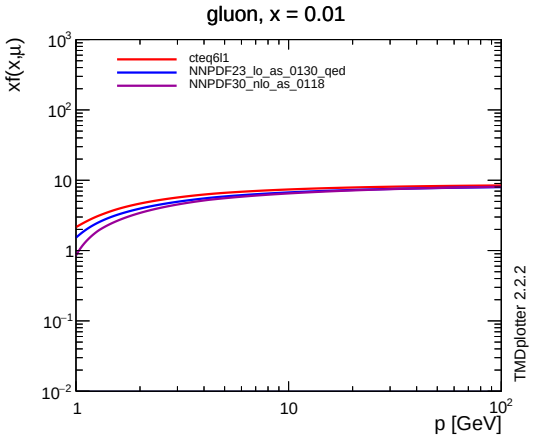
The comparison of the ATLAS [6] measurements to NLO parton-level predictions is shown in Figs. 6.5-6.6. The muon and vertex based analyses are compatible with one another within the systematic uncertainties. POWHEG +PYTHIA (green) shows a better agreement than MC@NLO +HERWIG (red).

6.2.2 CMS at 7 TeV

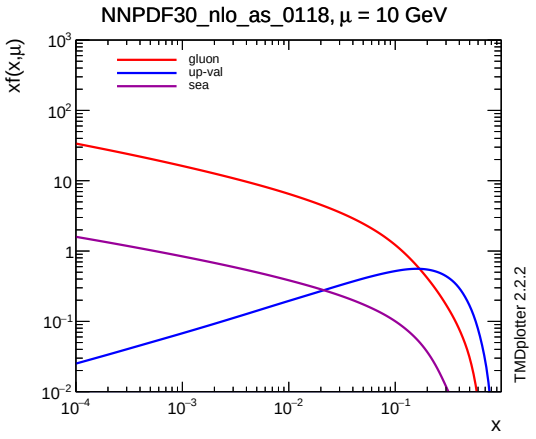
The measurements at CMS [7] are shown in Figs. 6.5-6.7. In addition to the ratio of the cross sections of the inclusive b jet production in data and simulation, the fraction of b jets in the inclusive jet production is also given.



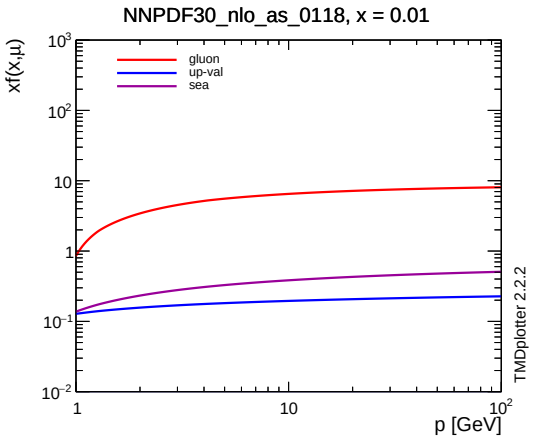
(A) Different sets compared on momentum fraction.



(B) Different sets compared on energy scale.



(C) Different partons compared on momentum fraction.



(D) Different partons compared on energy scale.

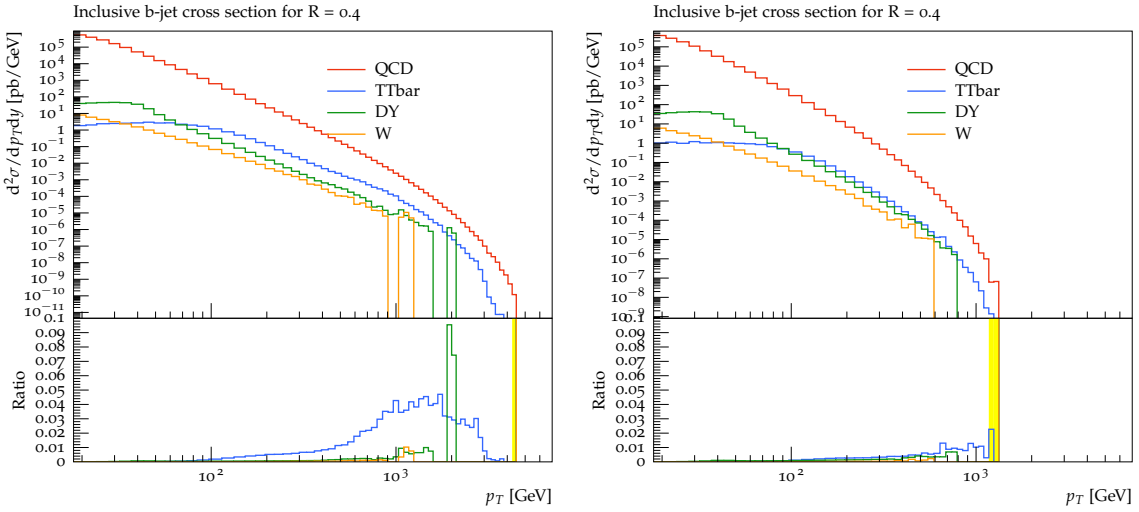
FIGURE 6.3 – Comparison of different PDF sets [2, 1, 3] in LHC conditions with different sets (top) and partons (bottom). The subfigures on the RHS (LHS) show the dependence on the scale μ (on the momentum fraction x). The curves come in pairs and show the uncertainties on the PDF sets. Performed with `TMDplotter 2.2.0` [4, 5]

Both measurements were published for $18\,\text{GeV} < p_T < 200\,\text{GeV}$. The conditions of low pile-up allowed to reach low transverse momentum. However, it is interesting to note that the upper boundary is not due to limited statistics but due to the reliability in the b tagging [8].

PYTHIA 6 (blue) exhibits a better description of the ratio with data (Fig. 6.7a), while MC@NLO (red) describes better the fraction of b jets in data (Fig. 6.7b).

6.3 Strategy

The aim of the present analysis is to perform a new *precision measurement* of the inclusive b jet cross section and fraction to the inclusive jet cross section, similarly to the measurements at $\sqrt{s} = 7\,\text{TeV}$ presented in Sect. 6.2. With the improvement in b tagging techniques from Run-I to Run-II and thanks to samples of much larger statistics, the



(A) $|y| < 0.5$ (B) $2.0 < |y| < 2.2$

FIGURE 6.4 – Double differential cross section for the most important modes of production of b jets in pp collisions at $\sqrt{s} = 13\,\text{TeV}$ in the central and forward regions, obtained from PYTHIA 8 (with `CUETP8M1` and `NNPDF 2.3` [1]).

measurable phase space is extended to the TeV scale.

The analysis will be conducted in two steps: first at detector level (Chap. 7), then at particle level (Chap. 8):

— At detector level, the effect of the detector is investigated in the simulation. Thanks to comparisons with data, corrections are applied to simulation to compensate for imperfect modelling. Moreover, the quality of the data is ensured and checked to be constant over time.

— Data distribution are then corrected to particle level. The procedure of correction from detector level to particle level is called *unfolding*.

The procedure of *unfolding* is crucial and underlies the whole analysis: the analysis at detector level ensures that the simulation describes optimally the data; then all detector and reconstruction effects are corrected together through the unfolding procedure.

Only after unfolding, conclusions from comparisons to theoretical predictions are drawn (Chap. 9). While comparisons to theoretical predictions (folded with the simulation of the detector) could be performed at detector level, unfolding is essential to allow comparisons with other measurements. It is also crucial for the extraction of parameters (*e.g.* α_S or tuning parameters) or of PDFs, as presented in App. C.

In this section, we first describe the phase space and the selection (Sect. 6.3.1), as well as the data and simulation samples that will be used to conduct the analysis (Sect. 6.3.2). Then we outline the different steps of the analysis in more detail (Sect. 6.3.3-6.3.4). Finally, we discuss the theoretical predictions to which the measurement will be compared.

6.3.1 Phase space definition and selection

The phase space definition and selection is summarised in Tab. 6.1.

— The phase space will be essentially limited by the pile-up at low transverse momentum and by the tracker acceptance for the rapidity coverage (including muon

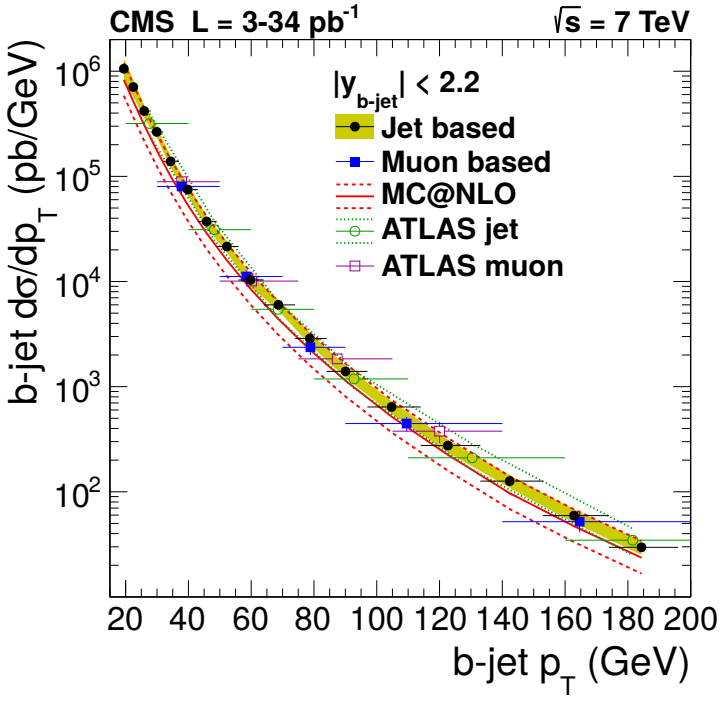


FIGURE 6.5 – Absolute differential cross section in transverse momentum at $\sqrt{s} = 7\,\text{TeV}$ of inclusive the b-quark jet production, measured in ATLAS (green and purple) and CMS (yellow and blue) and predicted with MC@NLO [7]. Four series of data points are shown, since both collaboration performed the analysis once with soft-lepton tagging (*muon*) and once with secondary-vertex tagging (*jet*). The CMS luminosity is $3\,\text{pb}^{-1}$ ($34\,\text{pb}^{-1}$) for the muon (jet) analysis, while ATLAS luminosity is $34\,\text{pb}^{-1}$ in both cases.

chambers).

— The selection of the events at detector level is motivated by the vertexing performance. Fake vertices due to a bad resolution and high contamination from pile-up are discarded. The standard definition of *good* event is performed centrally in the collaboration and corresponds to an event containing at least one *good* Primary Vertex (PV), defined as follow:

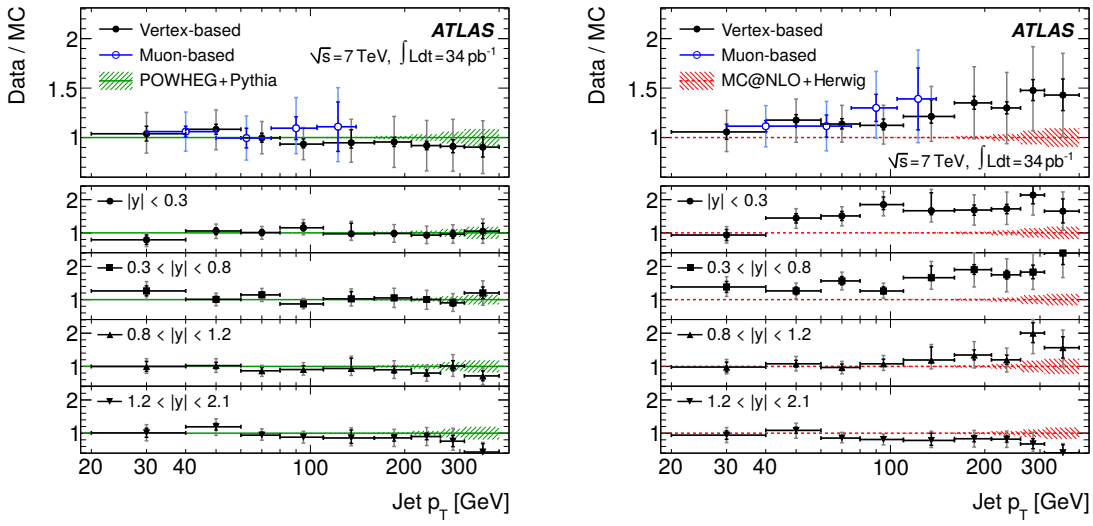
  — at least four tracks in the vertex fit;

  — the vertex position must satisfy $|z_{\text{PV}}| < 24\,\text{cm}$ around the IP, along the beam pipe and $|r_{\text{PV}}| < 2\,\text{cm}$ in the transverse plane to the beam pipe.

  The jet reconstruction at detector level aims at eliminating jets contaminated by pile-up; the *jet ID* is based on the stable particles entering the composition of the jet. In order to apply the calibration on jets, it is important to ensure that the jets are correctly described by the simulation.

— The b-tagging variable that shall be used by default in this analysis is primarily based on the detection of a Secondary Vertex (SV), namely the `CSVv2` tagger [9, 10, 11]. At the time when the analysis was started, the choice for the `CSVv2` algorithm was motivated by the fact that it was the best compromise between the size of the uncertainties and coverage of the calibration (*i.e.* up to $p_T = 1\,\text{TeV}$).

All the elements of the selection will be described in further details in Chap. 7. In general, since the analysis consists in a precision measurement, the tightest possible selection is considered.



(A) POWHEG (B) MC@NLO

FIGURE 6.6 – "Ratio of the measured cross-sections to the theory predictions of POWHEG and MC@NLO. In the region where the lifetime-based measurement overlaps with the muon p_T^{rel} measurement both results are shown. The top plot shows the full rapidity acceptance, while the four smaller plots show the comparison for each of the rapidity ranges separately. The data points show both the statistical uncertainty (dark colour) and the combination of the statistical and systematic uncertainty (light colour). The shaded regions around the theoretical predictions reflect the statistical uncertainty only. […]." [6]

	selection
reconstruction	good PV
	jet tight ID
kinematics	$p_T^{\text{jet}} > 74\,\text{GeV}$
	$\lvert y^{\text{jet}}\rvert < 2.4$
b tagging	tight selection

TABLE 6.1 – Phase space and selection of the analysis. The reconstruction criteria are centrally defined at CMS; the kinematics are restricted by the pile-up and by the tracker coverage; the b tagging is also define centrally.

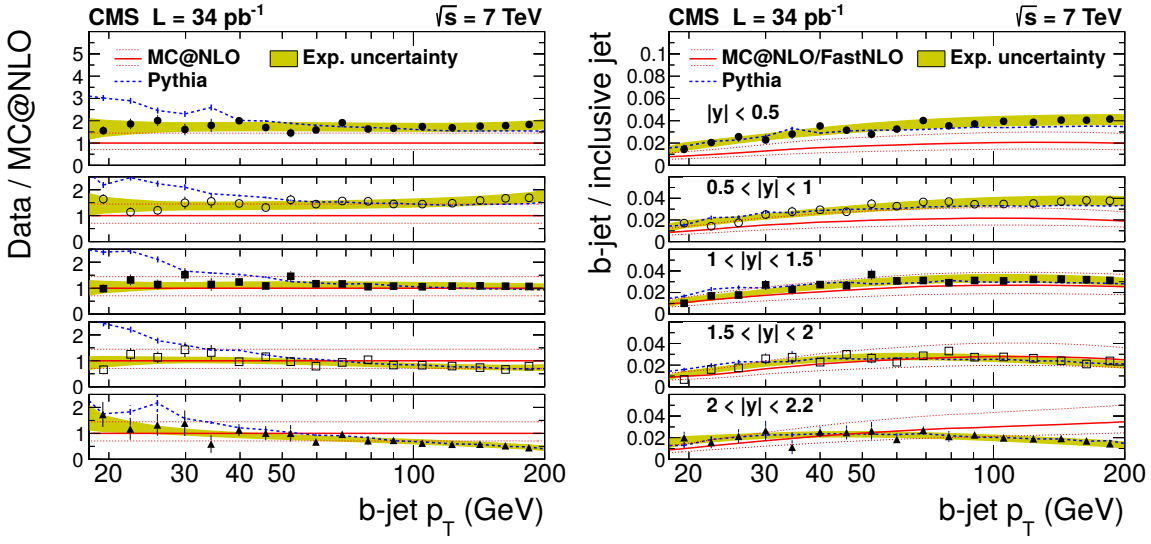
The discriminant is only applied on all jets to separate into non-b-tagged and b-tagged jets; in the end, the b jet cross section is extracted together with a non-b-jet cross section from the b-tagged and non-b-tagged cross sections.

Convention

In order to ease the discussions throughout the analysis, we use b and n ($\hat{b}$ and $\hat{n}$) for b true and non-b-true (b tagged and non-b-tagged) jets; even if the hadron definition is used, we shall use the convention at CMS and write b jet.

6.3.2 Samples

In this section, we give a description of the data and simulation samples.



(A) Ratio.

(B) Fraction.

FIGURE 6.7 – Measured b jet cross section from the jet analysis (yellow band and error bars for respectively systematics and statistical uncertainties), compared to the MC@NLO calculation (red lines, nominal and uncertainties are respectively plain and dotted) and to PYTHIA prediction (no uncertainty shown) [7].

6.3.2.1 Data

The data files are obtained after a long chain of certification, calibration and reconstruction that is not described here. The data taking is divided in different eras of various luminosities; the eras and respective luminosities are shown in Tab. 6.2. The different eras (or runs) correspond to different phases of data taking, and their exact definition is usually not relevant at the level of physics analyses. However, the calibration may be non-constant over time; in particular, the last era (`RunH`) was reconstructed with a rougher calibration [1]. In the analysis, we shall explicitly check that there is no dependence over time anymore after calibration.

Era	$\mathcal{L}/\,\mathrm{fb}^{-1}$	fraction
`Run2016BCD`	12.498	0.355
`Run2016EFearly`	6.589	0.187
`Run2016FlateG`	7.884	0.224
`Run2016H`	8.208	0.233
TOTAL	35.179	1.000

TABLE 6.2 – The luminosities are shown era per era, as well as their respective contribution to the total sample in terms of fraction. The number of events corresponds to the number of triggered events.

1. The reason for this is purely technical. At CMS, the size of the samples is such that the reconstruction with finer calibration can only be run once every few months. In particular, at the time of writing of the present thesis, one year after the end of the data taking, no fine calibration of `Run2016H` was yet available.

6.3.2.2 Simulation

Physics generators were already introduced in Sect. 4.2.2.1; here, we only describe details specific to samples used in the present analysis. Essentially, three generators will be considered: PYTHIA 8, MADGRAPH (interfaced with PYTHIA 8 for the PS, MPI and hadronisation), and HERWIG++. The simulations are compared in Tab. 6.3, and the PDF sets are shown in Fig. 6.3. Each sample contains a large number of event records at detector and particle levels; the description of the detector will be constructed by performing *matchings* between the two levels. In addition, all the samples include a simulation of the pile-up in data.

generator	PDF	ME	UE
PYTHIA 8	`NNPDF 2.3` [1]	LO hard-coded $2 \to 2$	`CUETP8M1`
MADGRAPH	`NNPDF 2.3` [1]	LO automated $2 \to 2, 3, 4$	`CUETP8M1`
HERWIG++	`CTEQ6L1` [2]	LO hard-coded $2 \to 2$	`CUETHppS1`

TABLE 6.3 – Simulation samples used to perform the analysis.

As a conclusion from Sect. 6.1 according to which QCD processes clearly dominate, it seems sufficient to consider only QCD processes in the simulation throughout the analysis. It is important here to stress that these samples will primarily be used to investigate the reconstruction, but not to draw any conclusions on physics. Whether other processes should be added while performing physics comparison depends on the resolution at high transverse momentum, but for the analysis itself, especially the unfolding, only QCD processes are considered.

Given the steeply falling character of the p_T spectrum, different strategies have been followed to provide sufficient statistics over the full phase space:

PYTHIA 8 — The sample is generated in slices of $\hat{p}_T$ (already mentioned in App. 4.A). The slices may be seen in Fig. 6.8a and the number of events per slice is summarised in Tab. 6.4a.

MADGRAPH — Similarly, the sample is generated in slices of the scalar sum of the transverse momenta of the outgoing partons in the ME:

$$H_T = \sum_{\text{jets}} p_T^{\text{jet}} \tag{6.4}$$

The slices may be seen in Fig. 6.8b and the number of events per slice is summarised in Tab. 6.4b.

HERWIG++ — Finally, the HERWIG++ sample is a smallest sample; the events are generated with a $\hat{p}_T$ uniformly distributed from 15 to 7000 GeV; the sample contains 9573, 938 events; the cross section is $1667 \cdot 10^6$ pb.

The HERWIG++ sample is a small sample with a slightly older calibration; its particular interest comes from the fact that it is generated via a completely independent sample, whereas the MADGRAPH sample uses the PYTHIA interface. Moreover, in addition to being a small sample, since the b calibration and Jet Energy Corrections (JECs) are derived for PYTHIA 8, we only use HERWIG++ as a cross-check in at the detector level.

6.3.3 Analysis at detector level

The analysis at detector level consists in checking in detail the data and simulation samples, and to correct the latter in order to improve the description of the former. The

inclusive b-tagged jet and inclusive jet productions are analysed simultaneously, in order to understand possible biases due to b tagging.

Formally, the double differential cross section in transverse momentum and rapidity can be written as follows:

$$\frac{\mathrm{d}^2\sigma^{\mathrm{rec}}_{(\hat{b})}}{\mathrm{d}p_T\ \mathrm{d}|y|} = \frac{N_{(\hat{b})}}{\epsilon\ \mathcal{L}\ \Delta p_T\ \Delta|y|} \tag{6.5}$$

where

— p_T (y) stands for the transverse momentum (rapidity),

— N ($N_{\hat{b}}$) stands for the count of the jets (b-tagged jets),

— $\mathcal{L}$ stands for the luminosity,

— $\epsilon(p_T, |y|)$ stands for the efficiencies (trigger, tracking, etc.),

— and Δp_T and Δy stand for the bin widths.

First, the data acquisition is detailed. Given the steeply falling character of the spectrum, the sample is split into different intervals of transverse momentum of the leading jet, corresponding to different acquisition rates (or *trigger rates*); the normalisation and combination of these different regions of the phase space is presented.

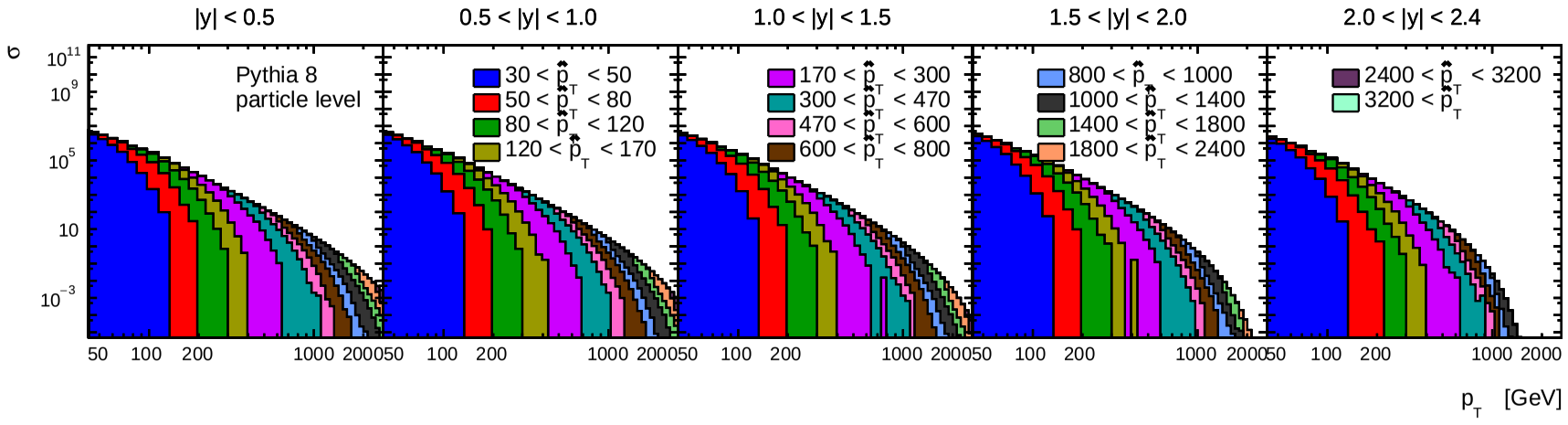
Then, various procedures are applied in order to improve the description of the data, and hence their agreement. Important aspects to treat are the pile-up simulation, the jet energy scale and resolution, and the calibration of b tagging. An estimation of each of these effects is given, together with their systematic uncertainties, without necessarily attempting to correct for them: this belongs to the unfolding procedure. Especially, at high transverse momentum, we shall see that large differences related to the b tagging in simulation and in data will be observed, together with a low tagging efficiency.

6.3.4 Unfolding to particle level

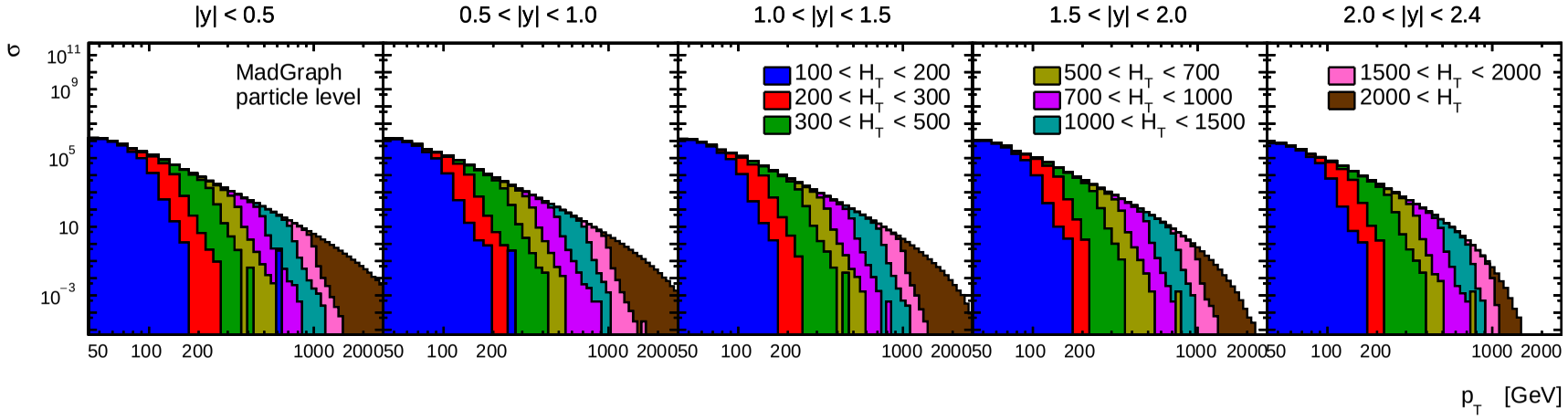
Once the simulation at detector level has been ensured to describe the measurement, the simulation samples can be used to revert the effect of the detector on the cross section. The *effect of the detector* usually refers to the smearing of the p_T spectrum: indeed, resolution effects translates into migrations of jets among neighbouring p_T bins, which, in case of a significant slope, implies a smearing. One can include the dependence in other variables, such as the rapidity.

Formally, the transition from the *true* (unknown) count N_b to the measured (known) count $N_{\hat{b}}$ can be expressed in terms of a matrix that can be constructed from the simulation; in this context, the transition matrix is often called Response Matrix (RM). However, the steeply falling character of the p_T spectrum can make additionally the inversion of the RM unstable. The specific treatment of the inversion of the RM is what was already mentioned as *unfolding* earlier in the chapter.

In addition to considering the migrations of jets among different regions of the phase space due to the effect of the detector, one can consider *flavour migrations* in the unfolding. Indeed, at first sight, the inclusive b jet measurement may be seen as a repetition of the all-inclusive jet measurement with an additional procedure of discrimination on the flavour of the jets; then, the inefficiency of discrimination has to be corrected for with a bin-to-bin correction extracted from the simulation. However, as we shall explain, in this thesis, we consider the b and n counts in parallel, and include in the unfolding the migrations among flavours from detector (or tag) level to particle (or true) level.



(A) PYTHIA 8



(B) MADGRAPH

FIGURE 6.8 – Contributions to the p_T spectrum of the different slices at generator level.

$\hat{p}_T$ slice		#events	σ / pb
30	50	9699558	140932000
50	80	9948791	19204300
80	120	7742665	2762530
120	170	5748730	471100
170	300	7838066	117276
300	470	11701816	7823
470	600	3959986	648.2
600	800	9628335	186.9
800	1000	11915305	32.293
1000	1400	6992746	9.4183
1400	1800	2477018	0.84265
1800	2400	1584378	0.114943
2400	3200	596904	0.00682981
3200	∞	391735	0.000165445

(a) PYTHIA

H_T slice		#events	σ / pb
100	200	69031923	27540000
200	300	18847246	1717000
300	500	17035890	351300
500	700	37645465	31630
700	1000	15534009	6802
1000	1500	4850746	1206
1500	2000	3970819	120.4
2000	∞	1967899	25.25

(b) MADGRAPH

TABLE 6.4 – Description of the sub-samples in PYTHIA 8 and MADGRAPH. The number of events and the cross section is given for each slice of $\hat{p}_T$ or H_T.

Indeed, in general, the advantage of performing a RM-based unfolding is to reduce the dependence on the simulation, in contrast to bin-to-bin corrections, especially when the bin-to-bin corrections factors are sensibly different from unity, which can lead to significant biases in the unfolded data [12]. Usually, the unfolding is performed for spectra involving different orders of magnitude, typically the p_T spectrum. Including the rapidity is less frequent, since migrations are not expected to be large; however, it may be preferable to include it in the unfolding to reduce effects at the edges of the phase space. Regarding the flavour, the performance of the b tagging will be such that the efficiency goes down to $20-30\%$ for $p_T \to 1$ TeV, with a significant contamination from other flavours; obviously, a dependence on the simulation would be inevitable with a bin-to-bin correction.

Such a treatment of the flavour migrations was not used in the past CMS or ATLAS publications, where only a bin-to-bin correction was performed. A major distinction that has to be mentioned is that differences between simulation and pseudo-data in toy experiments are rarely investigated in detail; however, it is crucial that the unfolding only corrects from detector effects but does not biases the physics result to the simulation. In the present analysis, the purity of the samples will be double checked (and, eventually, corrected) and the $p_T(y)$ spectrum will be modified in order to provide an optimal description of the smearing effect of the detector.

Finally, an additional, non-negligible advantage of using the RM-based unfolding is that statistical uncertainties can be correctly treated. Indeed, inclusive cross sections are multi-count observables, which means that jets measured in the same event are correlated. The importance of treating correlations correctly may be somewhat limited for the measurement of the inclusive b jet, where rarely more than one or two b jets are measured; but it is absolutely crucial to treat the all-inclusive jet measurement, where up to five or six jets may be measured in the same event, and, *in extenso*, to treat the fraction of b jets in the inclusive jet production.

6.3.5 Theoretical predictions

After the unfolding to particle level, the results can be confronted to theoretical predictions. In this analysis, we first present the comparison to the three LO predictions used to perform the analysis, *i.e.* PYTHIA 8, MADGRAPH and HERWIG++.

In addition, we also compare the data with NLO predictions from the POWHEG box, interfaced with PYTHIA 8 for the PS, MPI and hadronisation. For the PDF, the `NNPDF 3.0` [3] set has been considered (compared to the other PDFs in Fig. 6.3). Theoretical uncertainties include variations of the PDF, variations of the scales and variations of certain tune parameters.

References

[1] Richard D. Ball et al. "Parton distributions with LHC data". In: *Nucl. Phys.* B867 (2013), pp. 244–289. DOI: 10.1016/j.nuclphysb.2012.10.003. arXiv: 1207.1303 [hep-ph].

[2] J. Pumplin et al. "New generation of parton distributions with uncertainties from global QCD analysis". In: *JHEP* 07 (2002), p. 012. DOI: 10.1088/1126-6708/2002/07/012. arXiv: hep-ph/0201195 [hep-ph].

[3] Richard D. Ball et al. "Parton distributions for the LHC Run II". In: *JHEP* 04 (2015), p. 040. DOI: 10.1007/JHEP04(2015)040. arXiv: 1410.8849 [hep-ph].

[4] Patrick L.S. Connor, Francesco Hautmann, and Hannes Jung. *TMDplotter 2.2.0*. URL: http://tmdplotter.desy.de.

[5] Patrick L. S. Connor et al. “TMDlib 1.0.8 and TMDplotter 2.1.1”. In: *PoS* DIS2016 (2016), p. 039. DOI: 10.3204/PUBDB-2016-05953.

[6] Georges Aad et al. “Measurement of the inclusive and dijet cross-sections of b^- jets in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector”. In: *Eur. Phys. J.* C71 (2011), p. 1846. DOI: 10.1140/epjc/s10052-011-1846-4. arXiv: 1109.6833 [hep-ex].

[7] Serguei Chatrchyan et al. “Inclusive b-jet production in pp collisions at $\sqrt{s} = 7$ TeV”. In: *JHEP* 04 (2012), p. 084. DOI: 10.1007/JHEP04(2012)084. arXiv: 1202.4617 [hep-ex].

[8] Mikko Voutilainen. *b tagging in CMS during LHC Run-I*. Private communication.

[9] Serguei Chatrchyan et al. “Identification of b-quark jets with the CMS experiment”. In: *JINST* 8 (2013), P04013. DOI: 10.1088/1748-0221/8/04/P04013. arXiv: 1211.4462 [hep-ex].

[10] Luca Scodellaro. “b tagging in ATLAS and CMS”. In: *5th Large Hadron Collider Physics Conference (LHCP 2017) Shanghai, China, May 15-20, 2017*. 2017. arXiv: 1709.01290 [hep-ex]. URL: https://inspirehep.net/record/1621595/files/arXiv:1709.01290.pdf.

[11] Albert M Sirunyan et al. “Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV”. In: (2017). arXiv: 1712.07158 [physics.ins-det].

[12] Stefan Schmitt. “Data Unfolding Methods in High Energy Physics”. In: *EPJ Web Conf.* 137 (2017), p. 11008. DOI: 10.1051/epjconf/201713711008. arXiv: 1611.01927 [physics.data-an].

7

Analysis at detector level

In this chapter, the analysis is described at the level of the detector, *i.e.* without correcting the measurement from the artefacts of the detector. The selection is discussed, as well as several calibrations and their associated systematic uncertainties. At the end of this chapter, a global picture of the content of the sample at detector level is drawn.

The double differential cross section of the all-inclusive and b inclusive productions in data and simulation is shown in Fig. 7.1, as well as the ratio of simulation to data in Fig. 7.2. On the latter plot, one can observe that the ratios for the all-inclusive (in the top) and for the b-inclusive (in the bottom) differ more and more while reaching high p_T values.

Along this chapter, we explain the different procedures applied in order to obtain the distributions in Fig. 7.1, both data and simulation: first, the trigger strategy used to record the data is described. Then, the pile-up in data is described as well as the corrections applied in the simulation. This is followed by a description of the jet selection, and a first comparison of data and MC at detector level. Finally, the effect of b tagging is described. At the end, we conclude on the differences between the ratios of the all-inclusive and b-inclusive productions.

Some additional investigations and results complete the analysis in the appendix, as well as some investigations on the MET.

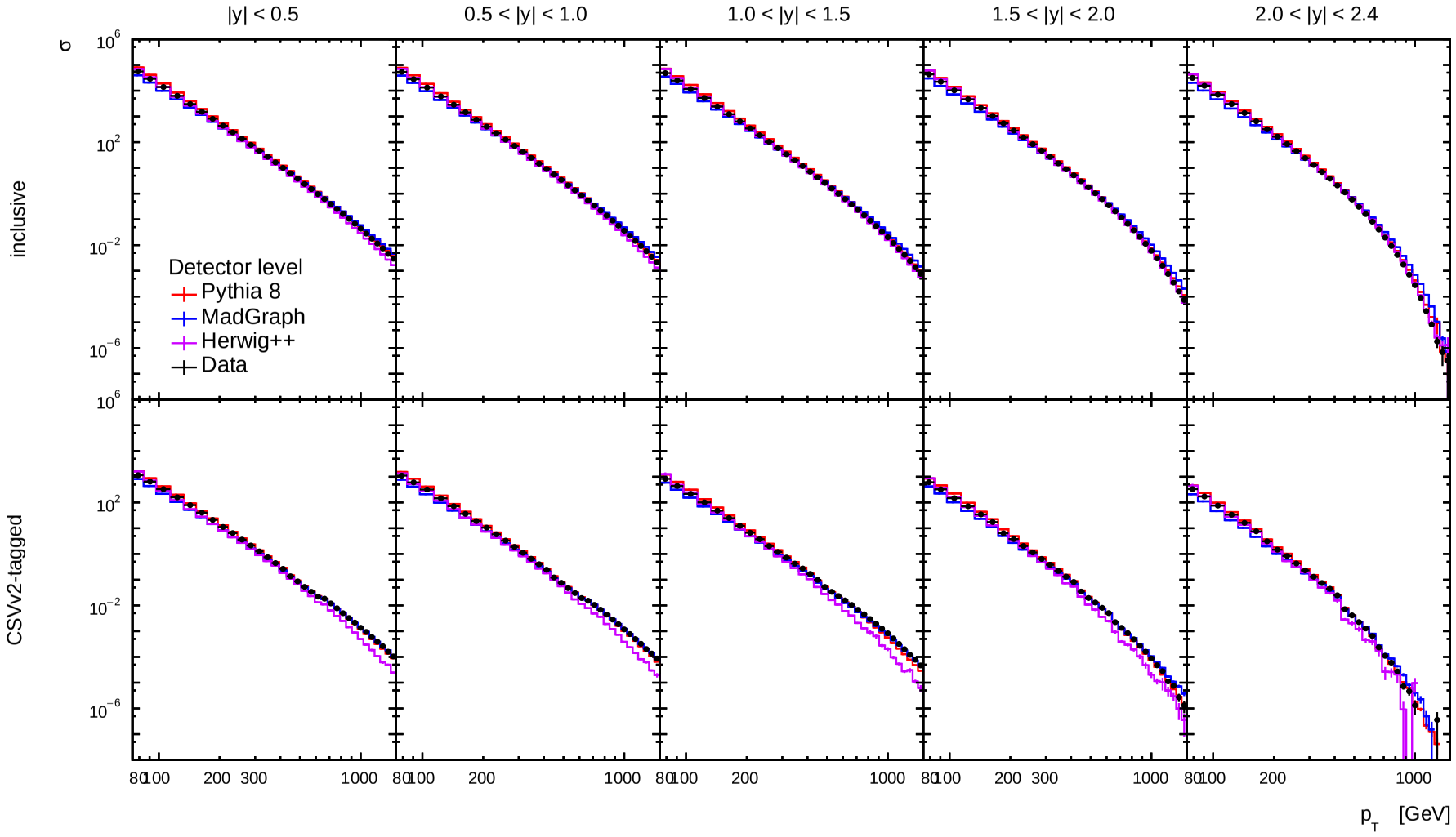


FIGURE 7.1 – Absolute double differential cross sections in transverse momentum and rapidity of the simulations (coloured curves) and of the data (black points with error bars). The columns correspond to the five rapidity bins; the top (bottom) row corresponds to the all-inclusive jet (inclusive $\hat{b}$ jet) selection.

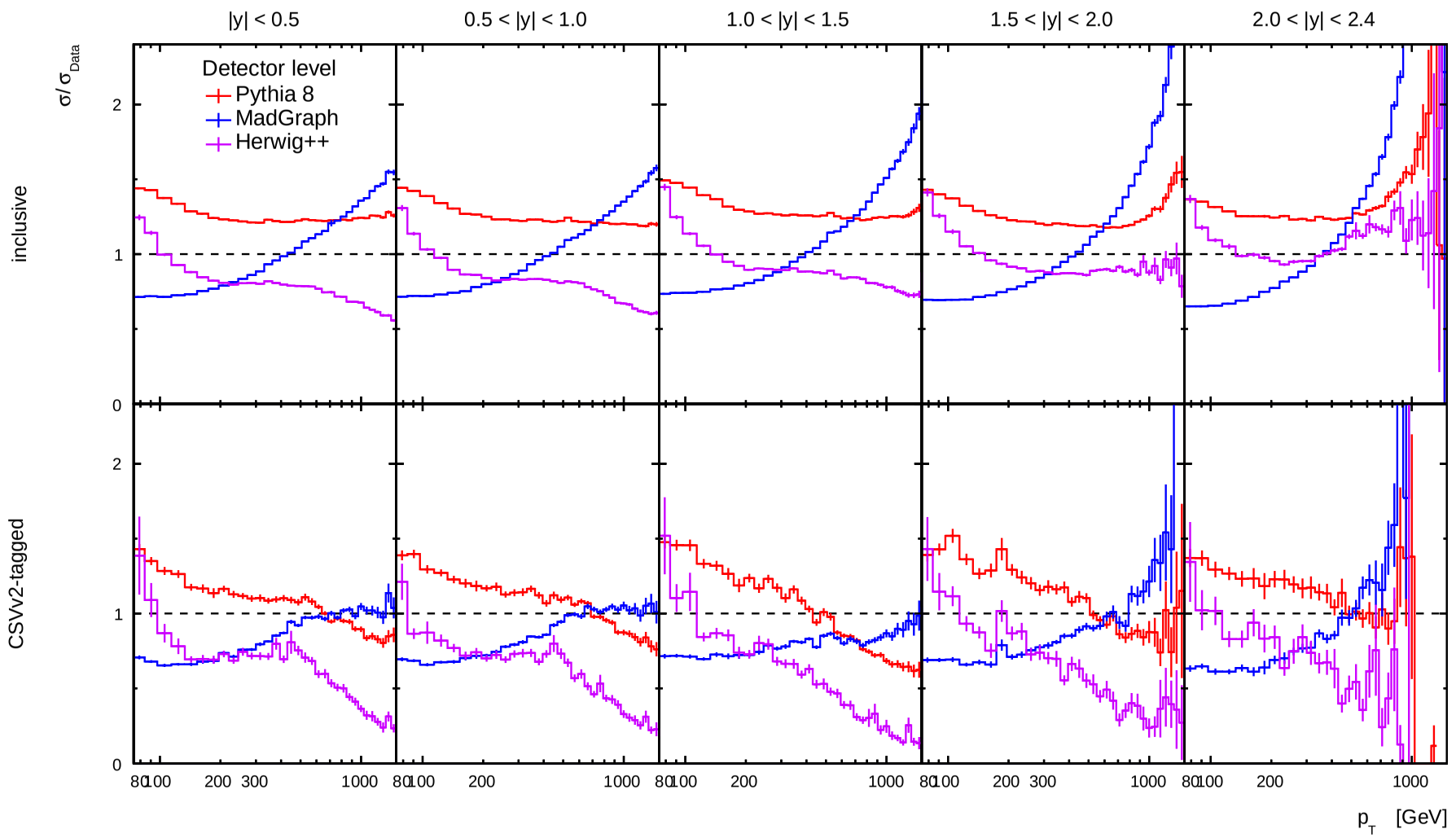


FIGURE 7.2 – Ratio of the double differential cross sections of the three simulation of the data. The columns correspond to the five rapidity bins; the top (bottom) row corresponds to the all-inclusive jet (inclusive $\hat{b}$ jet) selection.

7.1 Trigger

The general working principles of the trigger system has already been described in Sect. 3.2.2.5. The trigger strategy of the current analysis is now described.

The inclusive jet and inclusive b jet analyses use the single-jet triggers, defined only by the minimum transverse momentum of the leading jet. QCD processes being largely dominant, all these triggers are heavily *pre-scaled*, except the one of highest transverse momentum, in order to compromise between the extremely high event rate and the capability of the system to record the event. For instance, an event with a leading-jet p_T of 50 GeV will be triggered only once out of hundred thousands times, while an event with a leading-jet p_T of 800 GeV will be triggered each time.

The conditions of data taking may vary along the eras; the *trigger version* is regularly updated, corresponding to changes in the trigger pre-scales or in some correction applied on-line. Therefore, the stability of the conditions in different periods has to be checked: the four periods as defined for the Jet Energy Corrections (described later in Sect. 7.3) are `BCD`, `EFearly`, `FlateG` and `H`.

In this section, the strategy to use and combine triggers is described. First, since two triggers could be used for the same event, the strategy to combine the different triggers is given. Then the efficiencies of the different triggers are computed with a different method. Finally, the stability is checked on a run-by-run basis.

7.1.1 Trigger strategy

The *exclusive method* is used to combine triggers such that the total cross section is the sum of the cross sections obtained from each trigger:

$$\sigma_{\text{all triggers}} = \sum_{\text{all triggers}} \sigma_{\text{trigger}} \tag{7.1}$$

In this method, each trigger is associated to a different region of the phase space. Since the triggers are defined in terms of p_T of the leading jet of the event, the phase space is simply divided as a function of the leading-jet p_T; however, all subleading jets may have lower transverse momenta [1].

The different triggers are technically denoted as `HLT_JetX_vY`, where X (Y) stands for the p_T threshold in GeV (the version the trigger). Since from the trigger system a decision is needed very fast whether an event is worth being recorded, the trigger system has a very fast algorithm of reconstruction, not so precise as the PF reconstruction; therefore, p_T^{HLT}, corresponding to the HLT reconstruction, and p_T^{PF}, corresponding to the PF reconstruction need to be distinguished.

This distinction being made, the interval of p_T^{PF} (*i.e.* the phase space) corresponding to each trigger has to be defined. In general, the HLT algorithm being faster, it misses contributions and leads to an underestimation of the transverse momentum. This underestimation is not constant, and requires a determination of the transverse momentum from which they are fully efficient; this value is referred as *turn-on* point.

1. An alternative method exists, where the phase space is divided according to the different triggers not only for the leading jet but for all jets. In this method however, the statistical accuracy is not as good as in the here adopted method, since jets would not be considered with the lowest-pre-scaled trigger available.

7.1.1.1 Determination of the trigger efficiency

Three methods exist to determine the trigger efficiency:

1. the *reference trigger* method,
2. the *trigger emulation* method,
3. the Tag & Probe method.

These methods will be explained in this section.

Once the efficiency has been computed, a fit of the efficiency is performed with a transformed version of the error function [2]:

$$\epsilon(p_T) = a + 0.5 \times (1-a) \times \left(1 + \mathrm{erf}\left(\frac{p_T^{\mathrm{PF}} - \mu}{\sigma}\right)\right) \tag{7.2}$$

where a, μ and σ are the fit parameters. The *turn-on* corresponds to the value of p_T^{PF} where the efficiency is 99%. A given trigger can then be used from its turn-on up to the turn-on of the next trigger.

Reference trigger method. This method is the easiest in terms of methodology. Given a trigger of a certain p_T threshold known to be fully efficient, one tests another trigger of higher p_T threshold. The efficiency is therefore obtained according to the following formula:

$$\epsilon = \frac{N(\text{test fired}|\text{ref fired})}{N(\text{ref fired})} \tag{7.3}$$

This method has the drawback of very low statistics, especially for the trigger of lowest p_T where a minimum-bias or zero-bias trigger should be used. Indeed, the reference trigger intrinsically fires less than the test trigger. Therefore, in practice, the method has not been used, but is only mentioned in order to motivate the second method.

Trigger emulation method. The second method is an improved version of the reference trigger method, with the difference that the test trigger is emulated instead of directly used. Indeed, in the reference trigger method, the statistics is limited because of the condition that both the test and the reference triggers must have fired. By reproducing the conditions in which the test trigger would have fired, one allows a larger statistics.

$$\epsilon = \frac{N(\text{test emulated}|\text{ref fired})}{N(\text{ref fired})} \tag{7.4}$$

The turn-on points are extracted with this method and are given in Tab. 7.1. The only remaining drawback is that the turn-on point of the trigger of lowest p_T threshold can still not be computed [3].

Tag & Probe method. The third method allows to determine the turn-on of the trigger of lowest p_T and to cross-check the result obtained from the emulation method. The principle of the Tag & Probe method is not restricted to the determination of the

2. It is sometimes called *sigmoid function*, in reference to the S-shaped curve. However, sigmoid function refers sometimes to the *logistic function* $S(x) = \frac{1}{1+\exp(-x)}$, which is of similar shape. In comparison, the error function is defined by $\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp(-t^2)\,\mathrm{d}t$. In order to avoid any confusion, the term *sigmoid function* shall be avoided.

3. In principle, one could use a MB trigger, but the statistics could be too low.

Trigger	Turn-on	$\mathcal{L}/\text{pb}^{-1}$
HLT_PFJet_40	74	0.262
HLT_PFJet_60	84	0.711
HLT_PFJet_80	114	2.70
HLT_PFJet_140	196	23.7
HLT_PFJet_200	245	102
HLT_PFJet_260	330	580
HLT_PFJet_320	395	1730
HLT_PFJet_400	468	5070
HLT_PFJet_450	507	35200

TABLE 7.1 – Summary of the triggers turn-on values for 100% efficiency and luminosities per trigger.

trigger thresholds: it is a general method to determine the efficiency of reconstruction of a given objet from situations with two such objects are expected in an event. In the present case, it consists in using events with a di-jet final state and checking when only one or both should fired the trigger [4]. First, PF jets are matched jets with HLT objects and di-jet topologies are defined as follows:

— The di-jet final state is defined such that

1. both leading jets are back-to-back: $\Delta\phi_{12} > 2.4$;
2. and all other jets have significantly lower p_T: $p_T^i < 0.3 \times \frac{p_T^1 + p_T^2}{2} \; \forall \; i > 2$.

— The matching between PF and HLT jets is defined in $\Delta R < 0.5$.

The values of parameters are empirical but safe from mismatching, in the sense that no procedure was applied in order to determine optimal values. Then the efficiency is computed as follows:

$$\epsilon = \frac{N(\text{probe}|\text{tag})}{N(\text{tag})} \tag{7.5}$$

— the *probe* jet defines whether the event has fired;

— the *tag* jet tests the trigger.

In practice, the efficiency depends here also on the definition of the di-jet final state; therefore the obtained value for the turn-on with the trigger emulation method should be cross-checked. The trigger efficiency for each of the four aforementioned eras is determined in bins of rapidity. The result may be found in App. 7.B. In general, the turn-on values do not change over time; however, they are significantly lower in the forward region, where the activity is more intense.

An example of the measurement and fit of the trigger efficiencies is shown in Fig. 7.3 for the Tag & Probe method.

7.1.1.2 Final subdivision of the phase space

The choice of the turn-on points is obtained by taking the maximum given from each method, except for the first trigger (HLT_PFJet40), whose turn-on can only be computed

4. Another typical example of Tag & Probe method is the determination of the efficiency of reconstruction for muons. In that case, candidate muons pairs with $M_{\mu\mu} \approx M_Z$ are considered: as soon as one of them has been reconstructed, it is defined as the tag; then the efficiency is determined from the rate of reconstruction of the second muon.

with the Tag & Probe method. In addition, a correction is applied for the trigger inefficiency in the attempt of extending the measurement down to $p_T = 43\,\text{GeV}$.

We perform a last test in order to check the turn-on points and the extension down to 43 GeV. In order to test the efficiency of the trigger, one can look for steps or irregularities in the p_T spectrum. Such an effect is expected to take place at the linear scale, and therefore cannot be seen in the logarithmic scale. A solution to find any discrepancy consists in dividing the (*a priori* smooth) spectrum with another smooth curve of the same order. A fit the p_T spectrum in bins of rapidity and era is performed with the following function [5]:

$$f(x) = \exp\left(a + b\log x + c\,(\log x)^2\right) \tag{7.6}$$

The ratio of the cross section to the fit is shown in Fig. 7.4: the fact that the ratio is not at one simply means that the function is only an approximation of the shape of the spectrum; the ratio is rather smooth at high p_T, which was expectable since conservative choices have been done, except a step that can be seen at 74 GeV, where only the Tag & Probe method could be used. The step means that the correction of the trigger efficiency is not optimal, which is most likely related to fake di-jet topologies in high pile-up conditions. Eventually, the low p_T boundary of the phase space is defined to 74 GeV, as already mentioned in Chap. 6.

The combination of the triggers to form the spectrum can be found in App. 7.B.

7.1.1.3 Run-by-run stability

A crucial test for the trigger efficiency is to check the integrated cross section for each trigger on a run-by-run basis. The cross sections are simply obtained by counting the number of jets in the events fired by the trigger and dividing it by the effective luminosity:

$$\sigma^{\text{run}}_{\text{trigger}} = \frac{N^{\text{run}}_{\text{trigger}}}{\mathcal{L}^{\text{run}}_{\text{trigger}}} \tag{7.7}$$

(One can also define the run-by-run average pre-scales, as shown in App. 7.B.)

The run-by-run cross section of (fraction of `CSVv2`-tagged jets in) the inclusive jet production is shown in Fig. 7.5 (Fig. 7.6). Apart of a few outliers, the cross section and the fraction of $\hat{b}$ jets are rather constant.

7.1.2 Effective luminosity and average pre-scales

The only time dependence of the triggers is related to the conditions of data taking such as the pile-up; the pile-up decreases with time, because of the dispersion of the beam in the transverse plane. However, the hard process of interest should be of same nature at any time. The pre-scales decrease with the pile-up (already shown in Chap. 3 in Fig. 3.8). Since we are not interested in reproducing in MC the exact count as in data, we only need an average pre-scale over the whole period per trigger.

5. This function is well suited for this, since it is indeed typically used for peak or step hunting, for instance in dark matter searches.

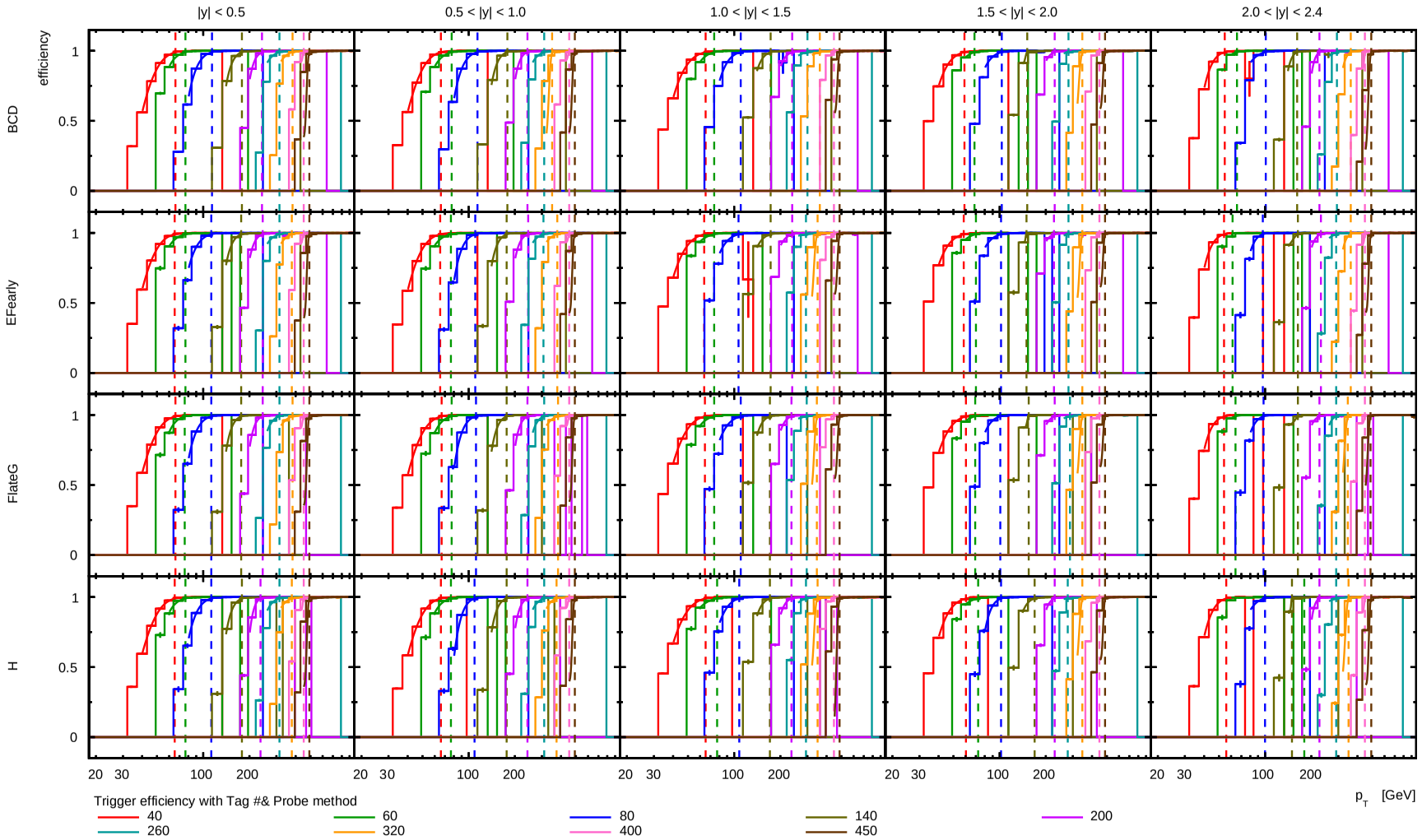


FIGURE 7.3 – Trigger efficiencies are obtained in bins of rapidity and fitted (continuous curves) to obtain the turn-on (vertical, dashed lines) with the Tag & Probe method. Each colour corresponds to a trigger; the row (columns) corresponds to the rapidity bins (period of data acquisition).

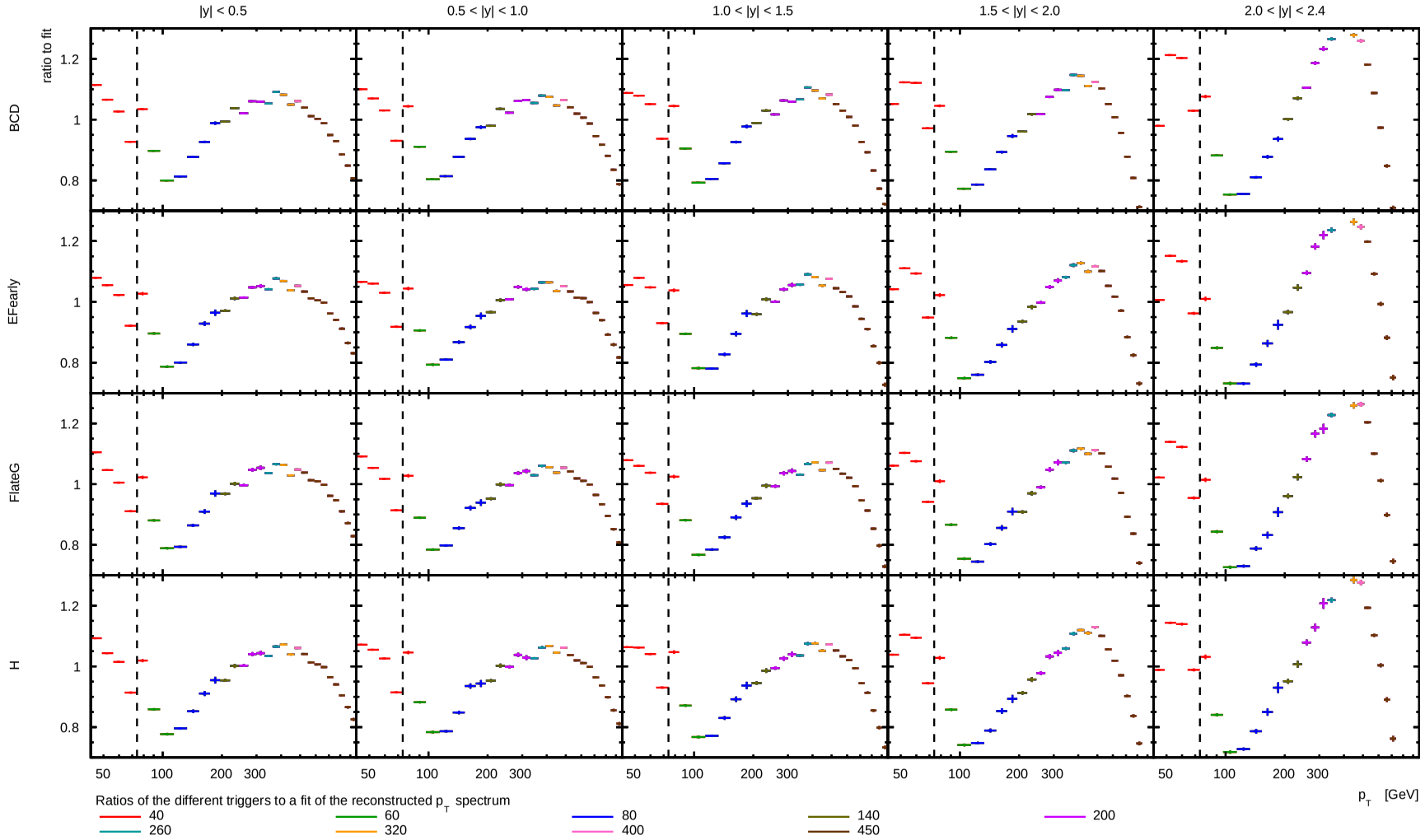


FIGURE 7.4 – The different contributions from the different triggers to the total cross section are divided by a smooth fit of the inclusive p_T spectrum. Each colour corresponds to a trigger; the row (columns) corresponds to the rapidity bins (period of data acquisition). The vertical, dashed, black line at 74 GeV correspond to the minimum p_T considered in the analysis.

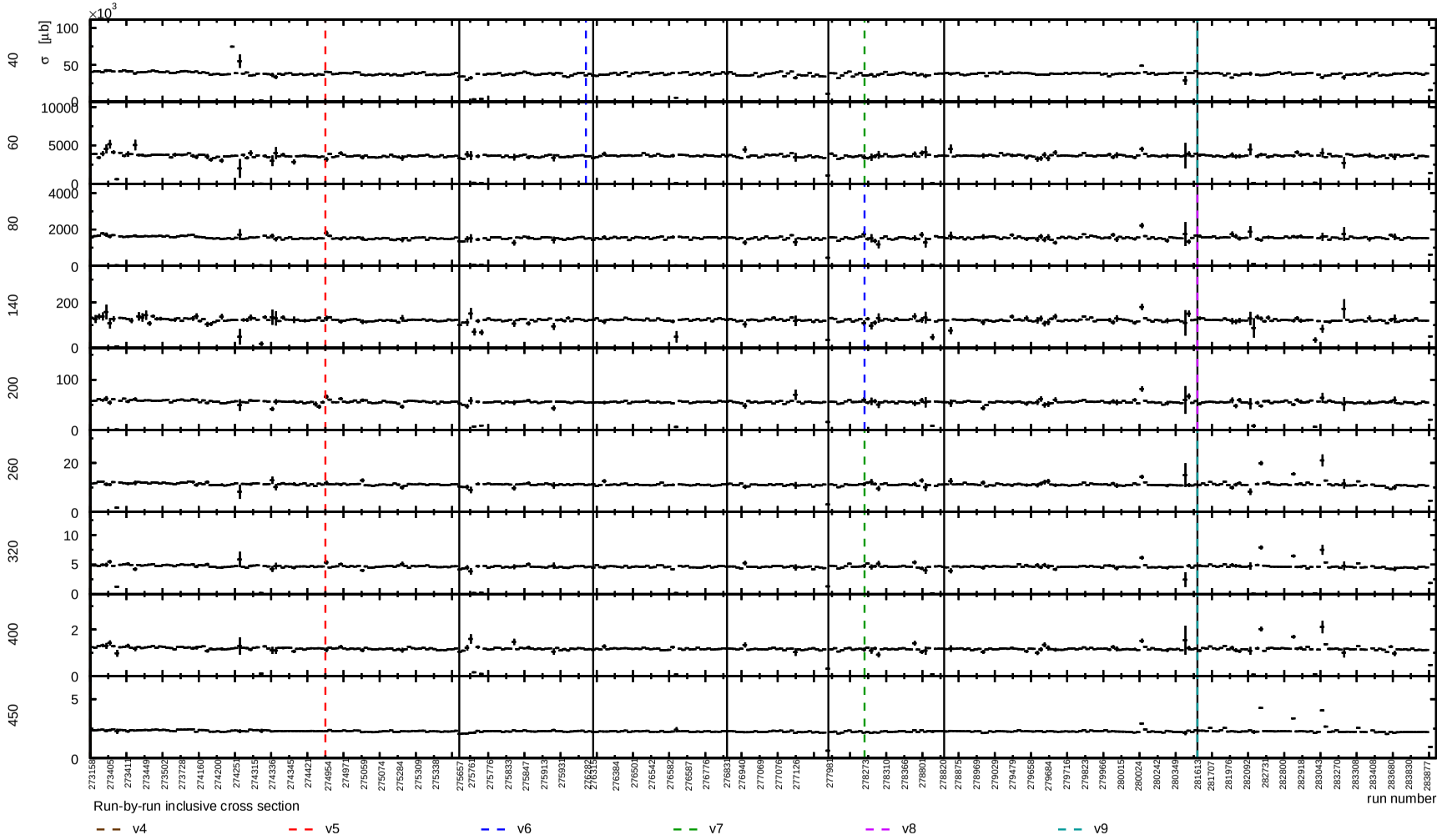


FIGURE 7.5 – Run-by-run inclusive jet cross section for each trigger. Each row correspond to a trigger (whose threshold is indicated on the very left). The vertical, continuous black (dashed coloured) lines correspond to a change of era (trigger version). 372 runs are considered in the analysis (only a limited number of them are explicitly labelled for readability).

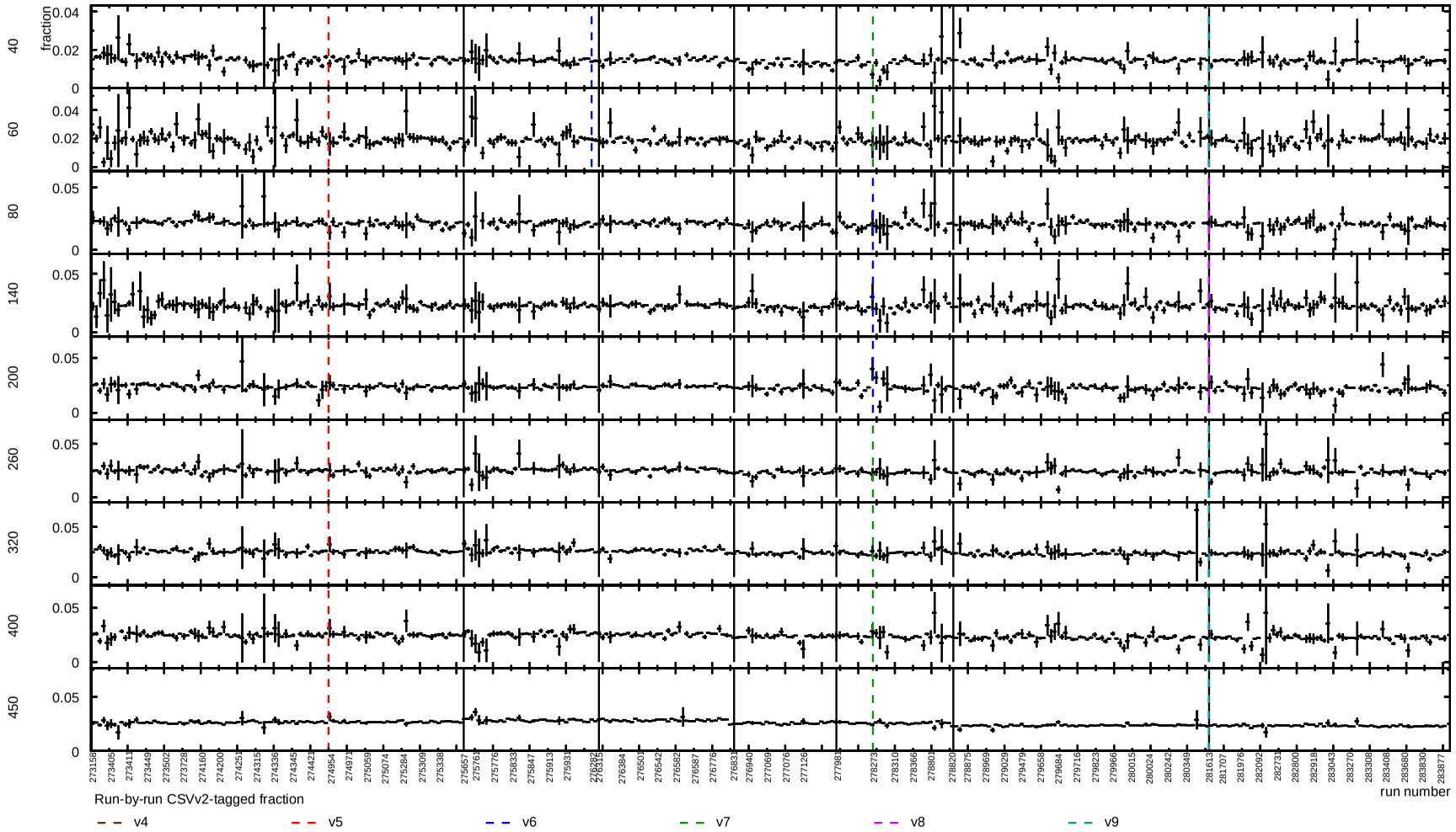


FIGURE 7.6 – Run-by-run fraction of $\hat{b}$ jets in the inclusive jet production for each trigger. Each row correspond to a trigger (whose threshold is indicated on the very left). The vertical, continuous black (dashed coloured) lines correspond to a change of era (trigger version). 372 runs are considered in the analysis (only a limited number of them are explicitly labelled for readability).

The global average pre-scale factor is given by the ratio of the effective luminosity of a trigger with the total luminosity of the sample:

$$f^{\text{av}}_{\text{trigger}} = \frac{\mathcal{L}_{\text{total}}}{\mathcal{L}_{\text{trigger}}} \tag{7.8}$$

$$= \frac{\sum_{\text{all runs}} \mathcal{L}^{\text{run}}_{\text{total}}}{\sum_{\text{all runs}} \mathcal{L}^{\text{run}}_{\text{trigger}}} \tag{7.9}$$

Eventually, the cross section of a given trigger is simply given by the following formula:

$$\sigma_{\text{trigger}} = \frac{\sum_{\text{all runs}} N^{\text{run}}_{\text{trigger}}}{\sum_{\text{all runs}} \mathcal{L}^{\text{run}}_{\text{trigger}}} \tag{7.10}$$

In fact, in addition to being simpler to apply, this procedure with average pre-scale factors has the advantage of reducing the impact of events of higher pile-up, since these usually require higher pre-scales.

7.2 Pile-up

Pile-up, already mentioned in 3.1.3, is a consequence of the increase of the delivered luminosity by LHC: on the one hand, increasing the pile-up increases the probability of seeing an interesting event; on the other hand, it increases the number of tracks in the event and may contaminate the measurement. Therefore, it is important to include a good simulation of the pile-up in the MC samples. Eventually, the effect of the pile-up on the measurement will be corrected while performing the unfolding.

In this section, general considerations about the pile-up are given. Then the two procedures applied in this analysis are described: first the reduction of overweighted pile-up events in the MC samples and secondly the reweighting procedure of the pile-up profile.

7.2.1 General considerations on the pile-up

The value of the pile-up corresponds to the *number of pp interactions* per LHC bunch crossing. It has to be distinguished from the *number of vertices*:

— a *pile-up event* is a *pp* interaction,

— a *vertex* is a reconstructed point in space from which a collection of reconstructed tracks seem to come from.

The second is indeed affected by possible track inefficiencies [6] or by the reconstruction of fake vertices. In practice, for 25 interactions, one typically expects around 17 vertices [2].

From the point of view of physics, the contamination from the pile-up is two-fold:

1. additional tracks may be taken inside the jets coming from the hard process of interest;
2. additional jets may be clustered in the events and wrongly associated to the hard process of interest.

The first contamination is mitigated in the PF reconstruction by the *charged-hadron subtraction*, which removes tracks from a jet if they are associated to another vertex [3, 4];

6. In particular, the data collected until mid-August 2016 (*i.e.* at the end of `RunF`) is affected by a dynamic inefficiency in the track reconstruction [1]. This has not been simulated in the MC samples.

in addition, the JEC also account for pile-up effects [5] (discussed later in Sect. 7.3). The second contamination is not crucial in the present analysis, since the measurement is done at relatively high transverse momentum; therefore, contamination from pile-up jets is not expected to be significant.

On the point of view of the treatment, we need to ensure two points:

1. the pile-up simulation has to be performed correctly;
2. the *in-time pile-up*[7] profile in simulation has to be corrected to the one in data.

These two points will be the objects of two reweighting procedures that will be described later in this section.

7.2.1.1 Measurement of pile-up in data

The method to estimate the pile-up in data consists in exploiting the relation between cross section and luminosity $\sigma = N/\mathcal{L}$ for the MB inelastic cross section[8]. Using this method for each LS successively, one can estimate the number of interactions per bunch crossing, *i.e.* the pile-up. Eventually, given the instantaneous luminosity, the numbers of vertices roughly corresponds to around 70% of the number of interactions.

All CMS analyses in 2016 use the same MB inelastic cross section [6, 7]:

$$\sigma_{\text{MB}} = 69.2 \pm 3.2\,\text{mb} \tag{7.11}$$

7.2.1.2 Simulation of the pile-up

The technique consists in generating QCD events following a Poisson distribution $\mathcal{P}(\lambda)$:

$$p(k) = \frac{\lambda^k}{k!}\exp(-\lambda) \tag{7.12}$$

However, a pure Poisson distribution can be too approximative; in practice, it is implemented as follows:

1. Several Poisson distributions with different λ parameters may be added to simulate better the data.
2. The mean of the Poisson distribution is taken a bit higher than the expected average pile-up in data; in other words, the MC samples are produced with slightly overestimated scenarios, and corrected later if need be.

In the current analysis, the PYTHIA 8 and MADGRAPH sample share the same pile-up simulation (with double Poisson distribution), while HERWIG++ has a slightly older one (with only one Poisson distribution); indeed, PYTHIA 8 and MADGRAPH account for the change of pile-up conditions after the tracker dynamics inefficiency at the end of `RunF`, while HERWIG++ does not [1]. In both cases, a reweighting procedure, described later in this section, is applied to improve the description of the pile-up in the simulation in agreement with the data.

7. An *out-of-time pile-up* exists as well, coming from the overlay of successive bunch crossings.
8. At detector level, the MB trigger is simply defined by a minimal amount of energy deposit in the HF.

7.2.2 Removal of overweighted events in simulation

As already explained in Sect. 6.3.2, in order to get a high statistics for all p_T values, the PYTHIA 8 (MADGRAPH) samples are generated in slices of $\hat{p}_T$ (H_T). However, the pile-up simulation is performed in addition without slices in p_T values, therefore, high-p_T jets can be produced even for the low-p_T slices. Therefore, when rescaling to the cross section, events with high-p_T jets will largely dominate the population. To fix this intrinsic problem of the simulation, events with a "too high" reconstructed leading jet p_T are rejected; the exact procedure is described in this subsection. As a result, this procedure modifies the count N in Eq. 6.5.

Being generated in slices of H_T rather than in slices of $\hat{p}_T$, the slices of MADGRAPH are wider with respect to the p_T spectrum. Nevertheless, a similar cut-off is applied on the leading generated jet (the $\hat{p}_T$ can be defined but does not allow to define exclusive samples, and is therefore not useful in the MADGRAPH samples).

The same issue of outliers in the simulation happens in HERWIG++. However, even though the sample is not generated in slices, they are more problematic, as the events are generated uniformly with respect to $\hat{p}_T$, and as the statistics is much smaller.

The cut-off is defined as follows: if the transverse momentum of the leading reconstructed jet is a few times higher than the transverse momentum of the leading generated jet (or than the scale of the ME), then the reconstructed jet is considered as a *bad jet* and has to be removed from the simulation:

$$X \times p_T^{\text{gen,lead}} < p_T^{\text{det,lead}} \tag{7.13}$$

$$X \times \hat{p}_T < p_T^{\text{det,lead}} \tag{7.14}$$

Unfortunately, with this cut-off purely based on the transverse momentum, it is not possible to remove these bad jets without removing as well *good jets*, *i.e.* without changing the shape of the spectrum. Therefore, an additional uncertainty is estimated by varying the value of the cut-off; in practice, it seems that $X = 3.5 \pm 1.0$ allows to remove the bad events with limited impact on the simulation.

It was also checked whether only the jet outliers should be removed or the full event containing the jet outlier. It turned out that this choice has a much lower effect than the choice of the numerical value of the cut-off.

In Fig. 7.7a, the effect of the procedure is shown on the absolute cross section from PYTHIA 8; from the colours, it can be seen that the bad jets are removed from the spectrum. The effect of the cut-off together with the associated uncertainties can be seen on Figs. 7.7b for the three MC samples. PYTHIA 8 and MADGRAPH are affected only at small p_T with around one percent; however, HERWIG++ is extremely contaminated.

7.2.3 Procedure of pile-up reweighting

The procedure of reweighting to the observed pile-up is applied to reproduce the pile-up conditions in the MC samples. The distribution on which this procedure is applied is the so-called *pile-up profile*, which describes the pile-up as a function of the average number of bunch crossings. At CMS, the pile-up profile is computed with an independent framework from the one used to perform the analysis [7]; however, it was taken care to extract it from the same data sample (*i.e.* from the full 2016 period of data taking).

Three quantities are shown in Fig. 7.8a before and after the reweighting procedure:

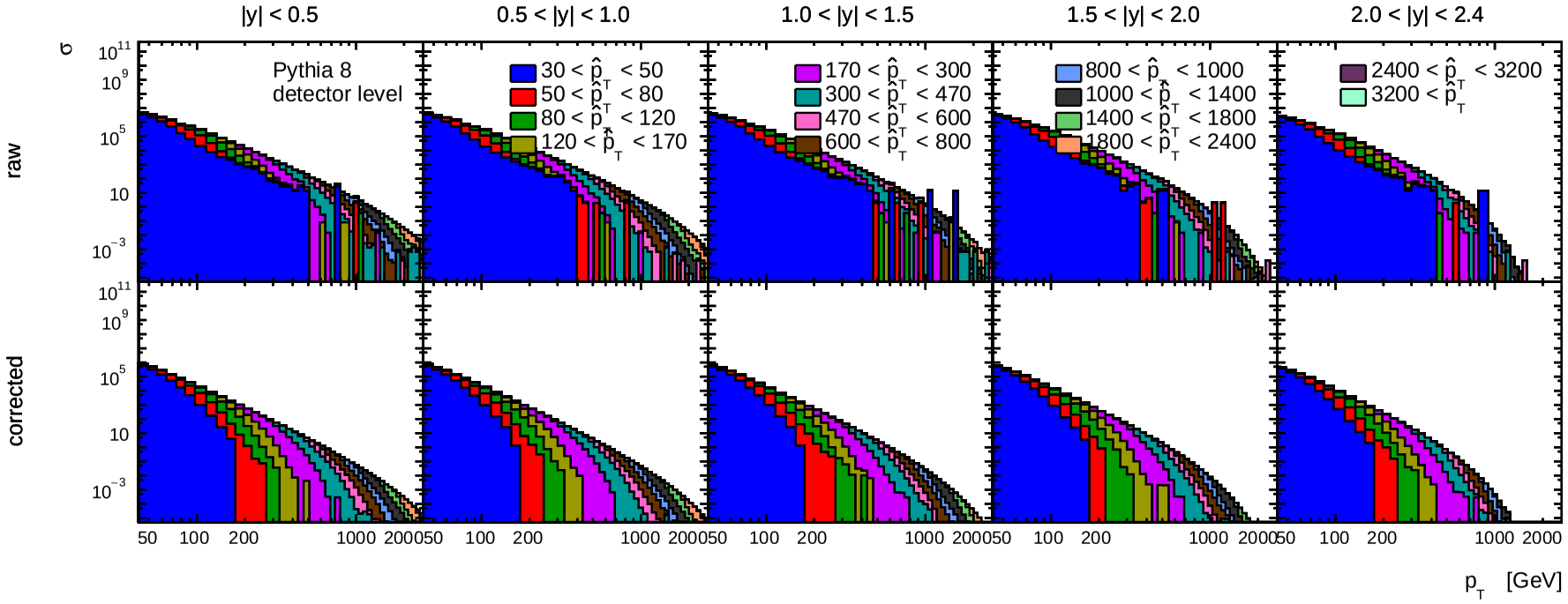
1. The pile-up profile is shown in the first column; the agreement after calibration is perfect as expected.
2. The number of (good) vertices is shown in the second and third columns; the agreement after calibration cannot be perfect as expected (since for instance, as mentioned earlier in the section, the tracker dynamics inefficiency is not simulated), but is improved.
3. Finally, the additional soft activity due to the pile-up is well described by the ρ variable [8]:

$$\rho = \text{median}\left[\left\{\frac{p_{tj}}{A_j}\right\}_{j=1,\ldots,N_{\text{jets}}}\right] \tag{7.15}$$

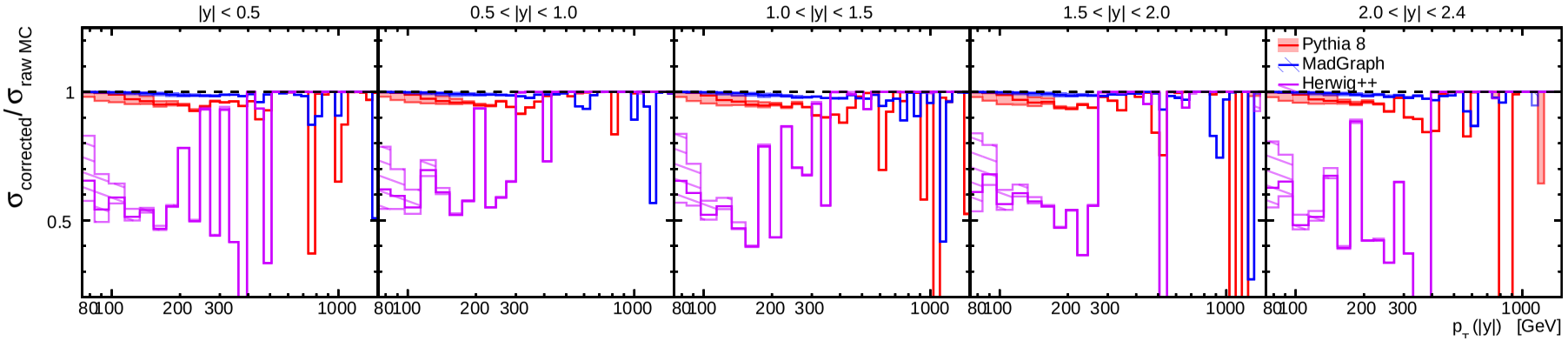
   where A corresponds to the *jet area* [9]. Since it is a median, ρ is not sensitive to the hard activity, and estimates the UE, the electronics noise, and the pile-up. Though not perfect, the agreement is improved, and the uncertainties reduce the difference between data and simulation.

The uncertainty band corresponds to the uncertainty on the MB measurement (Eq. 7.11).

The pile-up profile is then reweighted to correct the simulation to the data. The effect on the double differential cross section can be seen in Fig. 7.8b. Since the simulation of the pile-up in PYTHIA 8 and MADGRAPH (HERWIG++) is close to (different from) the real pile-up in 2016, the reweighting has a small (large) effect on the cross section; the width of the uncertainty band is related to the statistics of the sample, therefore larger for HERWIG++.

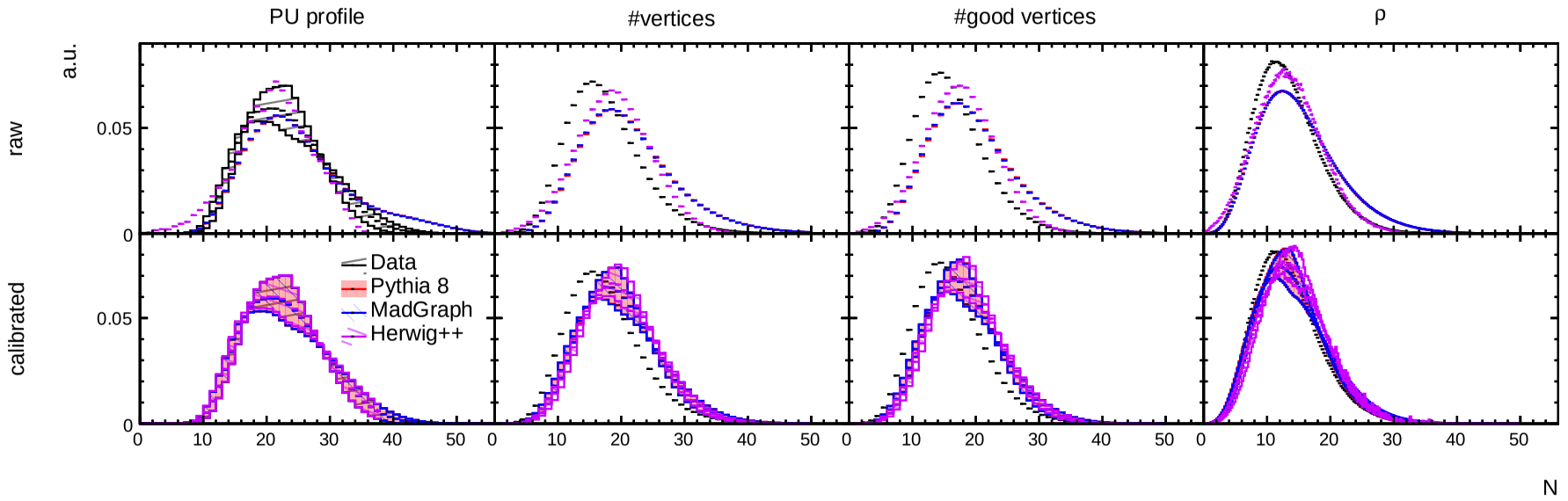


(A) The differential cross section simulation is shown in stacks corresponding to the different slices before and after the removal.

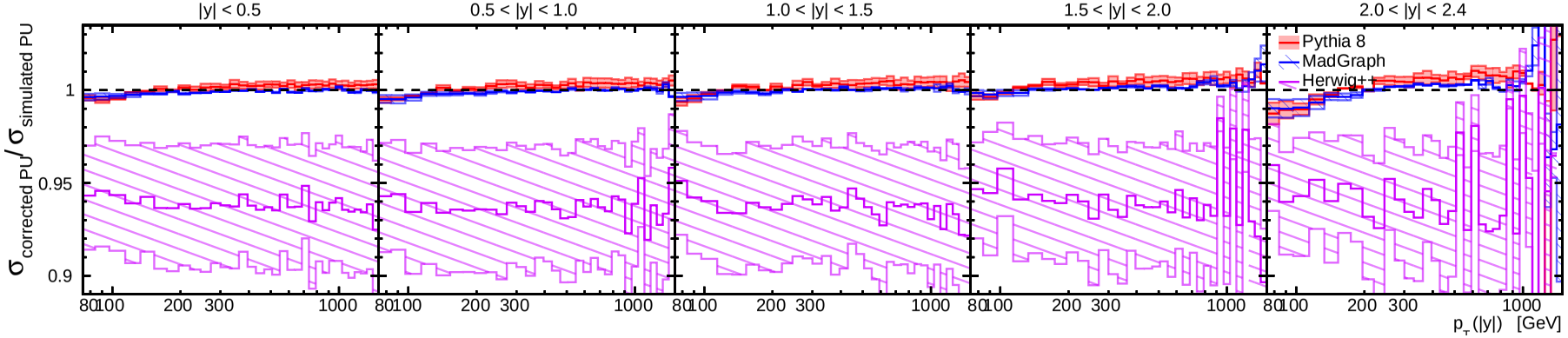


(B) Effect of the cut-off on the simulated cross section. The uncertainty corresponds to variations of the cut-off. The removal procedure has small (large) effect and uncertainties on MadGraph and pythia 8 (herwig++).

Figure 7.7 – Effect of the removal of the overweighted pile-up jet outliers and bins of rapidity. The different columns correspond to the rapidity bins.



(A) The pile-up profile, number of (good) vertices and soft-jet activity ρ are shown before (top) and after (bottom) the procedure of pile-up reweighting in the different columns. The curves are normalised to area one. The pile-up profile is perfectly corrected in simulation.



(B) Effect of pile-up reweighting with associated uncertainties on the spectrum. The reweighting procedure has small (large) effect and uncertainties on MadGraph and pythia 8 (herwig++). The different columns correspond to the rapidity bins.

Figure 7.8 – Effect of the reweighting of the pile-up profile of the simulation to the pile-up profile in the data. The uncertainty bands correspond to the uncertainty on the measurement of the MB correction (Eq. 7.11) and is red/shaded for pythia 8, blue/hashed for MadGraph and purple/hashed for herwig++.

7.3 Jets

Jets were introduced in Sect. 2.3.2.2. The interest in jets relies in that they can be defined at both particle and detector level. Elements of reconstructions were given in Sect. 3.2.2.2.

At CMS, the performance of the jet reconstruction is studied centrally [10, 11]. In this section, we discuss the calibration of jets both in data and simulation: in data, the dependence on time needs to be compensated; in simulation, it needs to describe the response of the detector in data. Then, we discuss the resolution on the jet kinematics in data and simulation.

In App. 7.E, we check the jet constituents by investigating multiplicity and the energy fraction of the different stable particles entering into the jet composition.

7.3.1 Jet Energy Scale Correction

The purpose of Jet Energy Corrections (JECs) is to correct the measured to the true energy of jet in the form of a global multiplicative factor to the four-momentum:

$$p_\mu^{\text{true}} = C_{\text{JEC}} \times p_\mu^{\text{raw}} \tag{7.16}$$

This correction factor can be further divided into several components, applied in a chain.

$$C_{\text{JEC}} = C_{\text{offset}}(p_T^{\text{raw}}) \times C_{\text{MC}}(p'_T, \eta) \times C_{\text{relative}}(\eta) \times C_{\text{absolute}}(p''_T) \tag{7.17}$$

The corrections are computed at three different *levels* in sequence, as illustrated by a diagram in Fig. 7.9, in the following order:

1. The *offset correction* C_{offset} removes everything that is not related to the pp collision of the hard interaction, *e.g.* pile-up and electronic noise ($p_T^{\text{raw}} \to p'_T$).
2. The *MC calibration* C_{MC} corrects for the main non-uniformities in pseudorapidity and non-linearities in transverse momentum ($p'_T \to p''_T$); typically, the calorimeters have a non-linear response.
3. The *residual corrections* $C_{\text{relative}} \times C_{\text{absolute}}$ accounts for finer corrections between data and MC ($p''_T \to p_T^{\text{true}}$):
   - relative energy scales are corrected by investigating dijet topologies, where the same energy is expected from both jets in opposite direction.
   - absolute energy scales are corrected by investigating Z + jets topologies, where the energy is measured accurately from the Z decay into electrons in the ECAL or into muons in the tracking system; complementarily, multijet events are also used.

An additional correction for HF jets can also be included, to account for differences with light jets, but the effect is mostly relevant for low p_T and is not considered here.

At CMS, the JEC are provided centrally [5, 11], together with an estimation of the associated uncertainties. The current JEC are computed with PYTHIA 8, therefore suited for the PYTHIA 8 and MADGRAPH samples, but not totally adequate for the HERWIG++ sample.

Each level contributes to the uncertainties:

1. The pile-up offset is mainly important at low p_T, and its uncertainty is below 0.1% from $p_T \sim 100\,\text{GeV}$.
2. The time stability matters especially at high p_T, where it reaches 0.5% of the uncertainty.

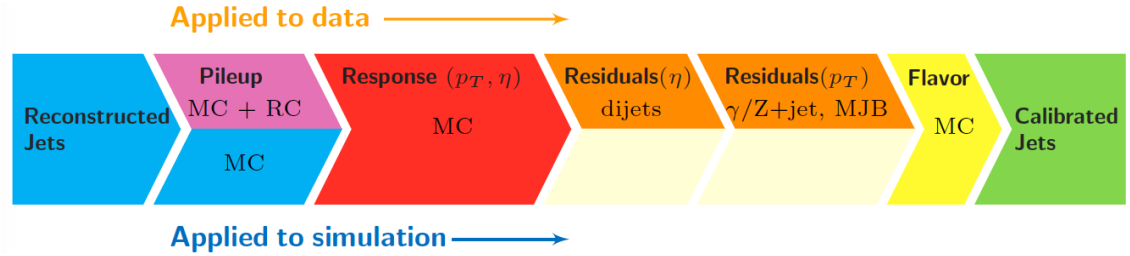


FIGURE 7.9 – Diagrammatic description of the application of JECs. More corrections are applied to data than to MC. Additional flavour corrections are not considered in this analysis. RC stands for *Residual Corrections*. Taken from [11].

3. The absolute (relate) scale contributes for around slightly less than 1% (slightly more than 0.1%) of the uncertainty.

Here we only show the global uncertainties due to JEC on the $p_T(y)$ spectrum in Fig. 7.10a.

7.3.2 Jet Energy Resolution

The choice of the binning scheme is related to the Jet Energy Resolutions (JERs); in QCD measurements at CMS, the binning is standard for all jet analyses in order to ease the comparison of measurements. The resolution on the transverse momentum is usually finer in MC than in data; therefore the transverse momentum has to be smeared, which is crucial for the unfolding. In the current subsection, we explain the procedure of smearing of the spectrum of transverse momentum in MC and check that the standard binning is reasonable. The binning is given in App. 7.C.

7.3.2.1 JER smearing

In MC samples, given a jet generated (reconstructed) with a transverse momentum p_T^{gen} (p_T^{rec}), we define the *resolution* as follows:

$$\Delta = \frac{p_T^{\text{rec}} - p_T^{\text{gen}}}{p_T^{\text{gen}}} \tag{7.18}$$

For a given p_T^{gen}, it is a Gaussian-like curve with a core and tail, but with two characteristics:

1. it is slightly asymmetric, since it is more probable for a jet to be reconstructed at lower values, when components (tracks) are missed by the reconstruction;
2. the left tail is more important because of various reconstruction effects (*e.g* situations where a jet is reconstructed into two jets, or situations where a pile-up jet is considered by mistake).

Analytically, the curve is usually fitted with a simple Gaussian curve (or with a Crystal-Ball curve, the latter being a modified version of the Gaussian curve to take into account the behaviour of the left tail). Therefore, the choice of this function may be discussed, since is does not describe the deviation to the Gaussian curve from the top; however in practice, a simple Gaussian fit appears to be enough.

Sometimes, the term *resolution* refers specifically to the width σ_{JER} of this Gaussian-like curve.

The resolution in data is measured and released centrally at CMS. Given these resolutions, one has to make sure that they are similar in data and MC by applying smearing on the spectrum of transverse momentum.

In principle, given the resolution Δ, the reconstructed and generated transverse momenta is related by the following formula (which is a rewritten version of Eq. 7.18):

$$p_T^{\text{rec}} = p_T^{\text{gen}} \times (1 + \Delta_{\text{MC}}) \tag{7.19}$$

Two methods exist to correct the resolution [12]:

Scaling method

This method assumes that the following matching can be done:

— $\delta R < R_{\text{cone}}/2$ where the R_{cone} is the cone size radius of the jet clustering algorithm (here $R_{\text{cone}} = 0.4$), and $\delta R = \sqrt{\delta y + \delta\phi}$ is the angular separation;

— $|\Delta_{\text{MC}}| < 3\sigma_{\text{JER}}$ where σ_{JER} is the measured resolution in data.

Then the resolution obtained from the MC value of p_T^{rec} has to be corrected with Scale Factors (SFs):

$$\Delta_{\text{data}} = \text{SF} \times \Delta_{\text{MC}} \tag{7.20}$$

Given this correction to the resolution, the value of p_T^{rec} can be corrected in turn:

$$p_T^{\text{rec}} = p_T^{\text{gen}} \times (1 + \Delta_{\text{MC}}) \tag{7.21}$$

$$\longmapsto p_T^{\text{rec}} = p_T^{\text{gen}} \times (1 + \text{SF} \times \Delta_{\text{MC}}) \tag{7.22}$$

These *smearing scale factors* are extracted from the measurement of the resolution in data, and are provided centrally as well [9].

Stochastic method

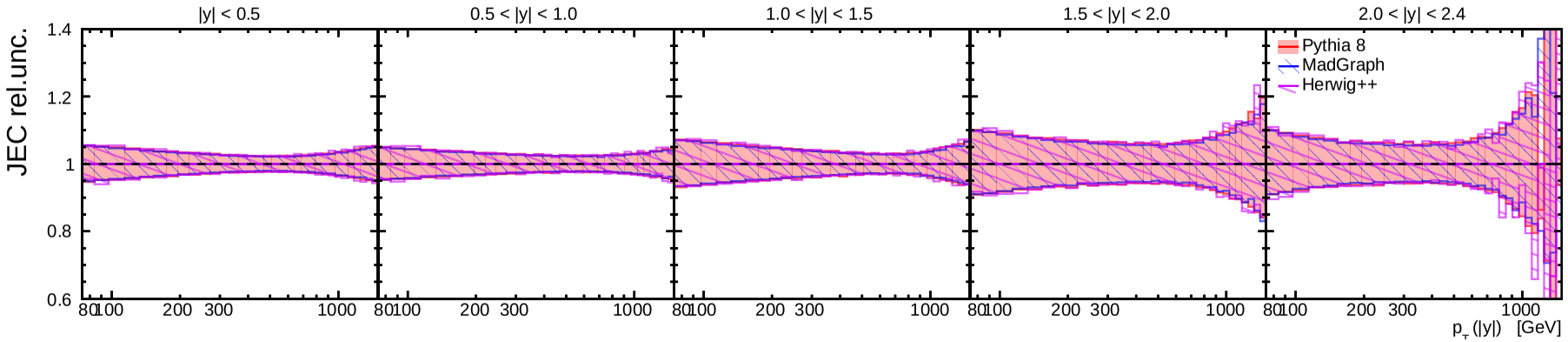
This alternative method is intended to be used in the case no matching can be done; in this case, one resorts to random numbers. One picks a number from the data resolution according to a centred Gaussian distribution with width equal to the resolution σ_{JER}, *i.e.* one picks $\Delta \sim \mathcal{N}(0, \sigma_{\text{JER}})$; then the transverse momentum is smeared accordingly:

$$p_T^{\text{rec}} \longmapsto p_T^{\text{rec}} \times \left(1 + \Delta \times \sqrt{\max\left(\text{SF}^2 - 1, 0\right)}\right) \tag{7.23}$$

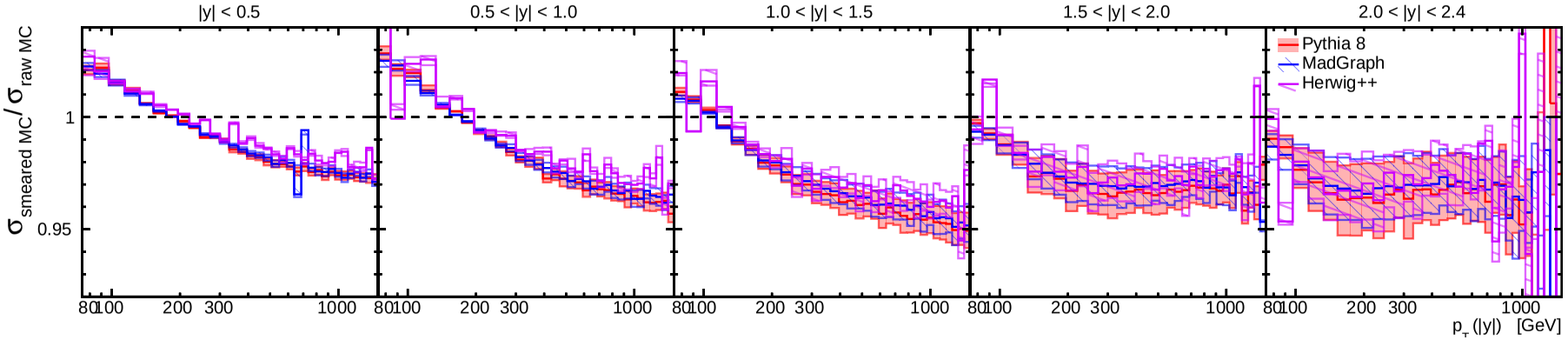
In practice, a so-called *hybrid method* is applied, according to whether the matching may be performed or not.

The effect on the spectrum can be seen on Fig. 7.10b: all simulations give similar results; the resolution has much larger uncertainties in the forward region than in the central, which is related to the contamination from the pile-up in the forward region; however, the smearing is more important in the central region than in the forward region.

9. In practice, it may sometimes be easier to apply $p_T^{\text{rec}} \longrightarrow p_T^{\text{rec}} \times (1 + (\text{SF} - 1) \times (p_T^{\text{rec}} - p_T^{\text{gen}})/p_T^{\text{rec}})$. The two formulations are entirely equivalent — one just has to play with the definition of the resolution.



(A) JEC uncertainties.



(B) Effect of JER correction on the spectrum. The bands correspond to the systematic uncertainties from the JER smearing.

FIGURE 7.10 – Uncertainties of JEC and effect of smearing of the transverse momentum on the spectrum for all three simulations. The columns correspond to the rapidity bins; the bands correspond to the systematic uncertainties.

7.3.2.2 Differential resolution and binning

After the JER smearing, the resolution can be checked. The differential resolution is shown in Fig. 7.11a for PYTHIA 8 in the five rapidity bins.

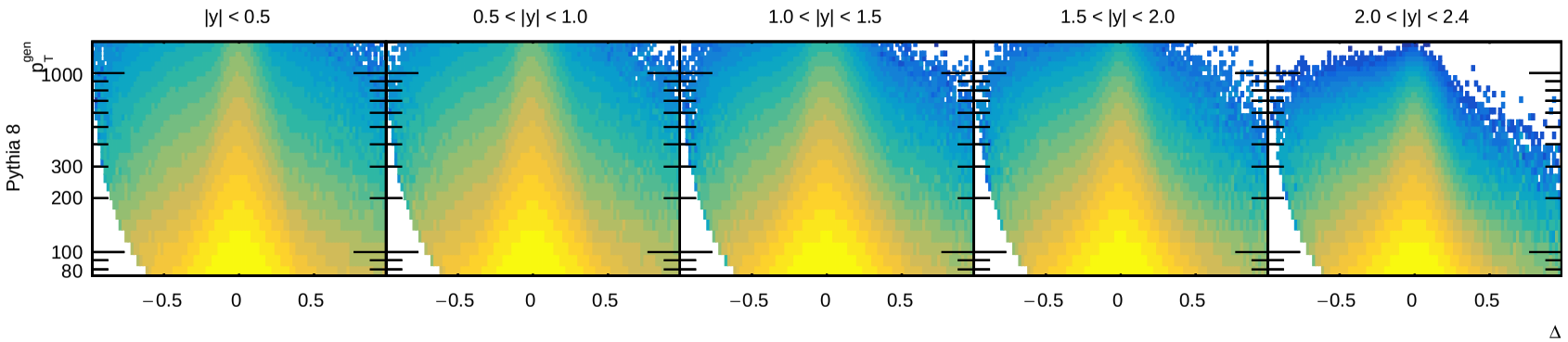
The profile of the resolution is extracted from the differential resolutions in order to check the binning, as shown in Fig. 7.11b. For each p_T bin, a Gaussian fit is performed, and the mean $\langle\Delta\rangle$ (top row) and width $\sigma\,[\Delta]$ (bottom row) are extracted. The resolution defined by the bin width is shown with a black curve and correspond to the following:

$$\frac{\delta p_T}{p_T} \approx \frac{p_T^{\text{up}} - p_T^{\text{down}}}{p_T^{\text{center}}} \tag{7.24}$$

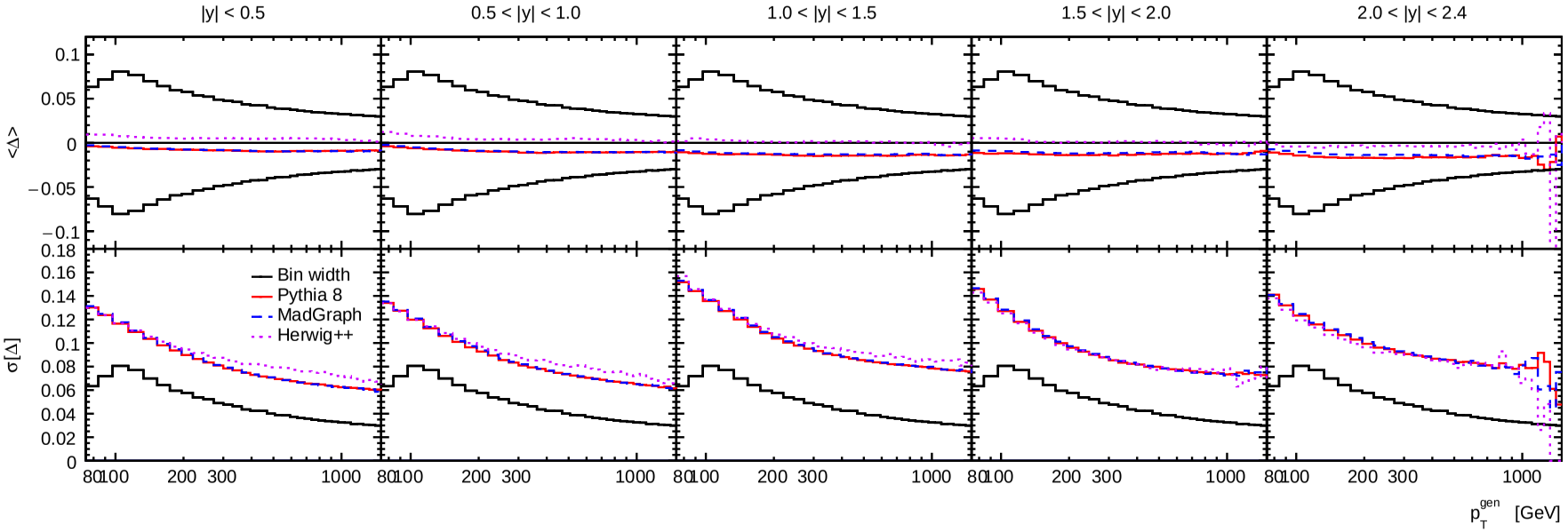
where up, down and center are defined for each bin respectively.

In average, the jets are reconstructed in the same bin and the jet migrations follow the bin widths; the width of the profile is a multiple of the bin width, except for low values where the bin width is less regular. The systematic shift of HERWIG++ with respect to PYTHIA 8 and MADGRAPH can be explained by the fact that there is no proper corrected to HERWIG++ (*i.e.* the SFs are determined only for PYTHIA 8).

It is also shown in App. 7.C for the systematic uncertainties of JER and JEC and for b and $\hat{b}$ spectra.



(A) Differential resolution shown in bins of rapidity for the PYTHIA 8 sample. At fixed p_T^{gen} values, the resolution has a Gaussian profile. The white area corresponds to impossible values according to the definition of the resolution in simulation.



(B) Resolution profile for PYTHIA 8, MADGRAPH and HERWIG++. Above (below) is the mean (width) of the Gaussian profile, compared to the resolution corresponding to the bin width (black line).

FIGURE 7.11 – Differential resolution and resolution profile, shown in bins of rapidity (columns) as a function of the transverse momentum of the generated jets.

7.4 Tagging

The principles of b tagging were already introduced in Sect. 5.3.1.2. We saw that different properties of b jets can be exploited to identify (or *tag*) them: the presence of a SV, the impact parameters of the tracks or the presence of a non-isolated lepton in the jet.

In this section, we describe the taggers at CMS in more detail and explain the calibration of the b tagging in the simulation.

7.4.1 Taggers at CMS

At CMS, different taggers are defined centrally [13, 14, 15]; in this section, we present three of them: Combined-Secondary-Vertex (`CSVv2`), Jet Probability (`JP`) and combined-Multi-Variate-Analysis (`cMVAv2`).

7.4.1.1 `CSVv2`

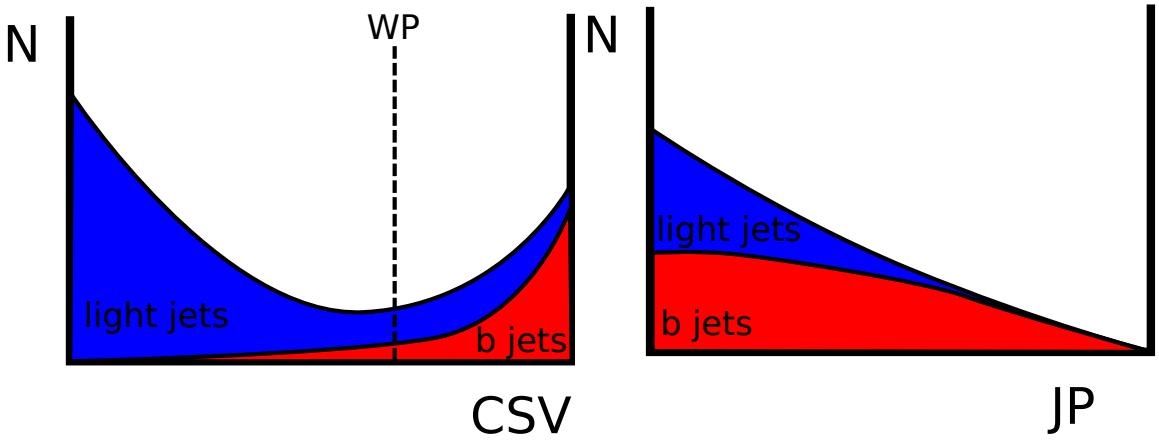
The `CSVv2` tagger primarily exploits the presence of a SV with an invariant mass $M_{\text{SV}} < 6.5\,\text{GeV}$ *i.e.* around 1.5 GeV above the mass of the B hadron (the invariant mass is computed from the tracks associated with the SV). To the difference of the PV that is reconstructed with the Adaptive Vertex Reconstruction (AVR) fitter (see Sect. 3.2.2.1), the SV is by default based on the IVF: indeed, the AVR fitter is only based on tracks already clustered in jets, while the IVF takes all tracks into account. Pairs of tracks compatible with a long-lived K^0_S are rejected. Finally, the flight direction of the SV must satisfy an angular separation $\Delta R < 0.4$ with the axis of the jet.

In its first version, during Run-I, the `CSV` tagger was also *combined* (hence its name) with additional variables, in order to increase the performance; for instance, it could be combined with track-based information in case no SV could be reconstructed, in a similar way as the `JP` tagger (described in the next paragraph). In its second version, intended for Run-II, many additional variables have been added in order to increase the power of discrimination: for instance a correction to the SV mass is considered, as well as various variables describing the kinematics of the tracks belonging to the SV and its geometry. All the variables are combined into a neural network.

In the present version, in this work, the `CSVv2` tagger is trained in 19 $(p_T, |y|)$ bins, which are given in Tab. 7.2. Its shape is sketched in Fig. 7.12a: it can be seen that it takes the shape of a valley, where light jets (b jets) are concentrated in the left (right). It is the main tagger for this analysis.

p_T/GeV	$\lvert\eta\rvert$
$15 - 40$	$0 - 1.2 - 2.1 - 2.4$
$40 - 60$	$0 - 1.2 - 2.1 - 2.4$
$60 - 90$	$0 - 1.2 - 2.1 - 2.4$
$90 - 150$	$0 - 1.2 - 2.1 - 2.4$
$150 - 400$	$0 - 1.2 - 2.1 - 2.4$
$400 - 600$	$0 - 1.2 - 2.4$
$600 - \infty$	$0 - 1.2 - 2.4$

TABLE 7.2 – Binning of the training of the `CSVv2` tagger.



(A) CSVv2. (B) JP.

FIGURE 7.12 – Sketches of discriminant variables. The red (blue) corresponds to bottom (light) jets. The discriminant takes different shapes according to the flavour. The Working Point (WP) is an arbitrary point separating flavours. Charm jets are neglected in this picture.

7.4.1.2 Jet Probability

The SV is sometimes too close to the PV to be identified with a SV tagger. Alternatively, the JP tagger is based on the impact parameter of the displaced tracks with respect to the PV. This technique is inherited from LEP [16, 17], with a slight difference in the definition in order to be suitable for pp collisions.

The construction of the JP tagger can be summarised as follows: given a jet, a probability for each track of the PV to belong to the jet is first defined from the resolution; indeed, tracks coming from a SV will rather populate the tail of the resolution function. The track probabilities are then combined into a probability for a jet; the combination is defined so as to be sensitive to the presence of tracks with low probability to below to the PV, *i.e.* to be sensitive to the presence of B mesons. (Despite its name, the JP tagger itself is rather the (negative) logarithm of the jet probability; therefore its value can be greater than one.)

The distribution takes the shape of a descending slope, where light jets mostly peak to 0 while b jets are spreader, as illustrated in Fig. 7.12b.

7.4.1.3 cMVAv2

The cMVAv2 tagger combines several taggers with MVA techniques, in order to exploit the advantages of all the techniques. It combines two versions of the CSVv2 tagger, two versions of the JP tagger and two additional soft-lepton taggers; the correlations among the taggers are shown in Fig. 7.13. In this analysis, it will be considered as a cross-check.

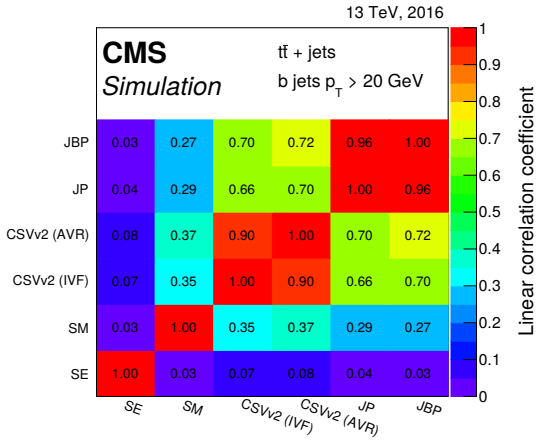


FIGURE 7.13 – "The correlation between the different input variables for the cMVAv2 tagger for b jets in $t\bar{t}$ events. (...)" [15] `JBP` is a variant of `JP`; AVR and IVF are two vertex fitters; `SM` (`SE`) stands for *Soft Muon* (*Soft Electron*).

7.4.2 Performance of the taggers

As already mentioned in Sect. 6.3.1, different levels of discrimination can be applied, corresponding to different compromises between the efficiency to tag a b jet and the rate of (mis)tagging a light or c jet; three standard *Working Points (WPs)* are usually defined for jets with $p_T > 30\,\text{GeV}$ as follows:

tight (T) for 0.1% of misidentified light jets,

medium (M) for 1% of misidentified light jets,

loose (L) for 10% of misidentified light jets.

The misidentification rates increase significantly at higher p_T, as will be shown in this section. The reason for defining these parameters is to compute standard calibration within the CMS collaboration. In this analysis, since the aim is to measure b jets with high p_T, the tight WP shall be considered; the two other WPs will be used for cross checks.

In this subsection, we explain the calibration of the MC samples, then we show the performance of tagging techniques in data.

7.4.2.1 Calibration

The calibration consists in correcting the MC efficiency and misidentification rates to the data [18]; the effect is illustrated in Fig. 7.14. The corrected quantity is the count N in Eq. 6.5; however, it only rearranges the contributions from the different flavours but does not change the all-inclusive double differential cross section, *i.e.* , it changes $N_{\hat{b}}$ and $N_{\hat{n}}$ but not their sum.

Procedure. The calibration is described only as a function of p_T and of the true flavour; in principle, it should be also a function of η, but this was not done at CMS in 2016. It is formulated in terms of SFs, given for each flavour; *b-tag SFs* denote the SFs for tagging (true) b jets, while *mistag SFs* denote the SFs for tagging light and c jets. A reweighting factor w_{entry} is computed for each entry from the SFs and from the efficiencies

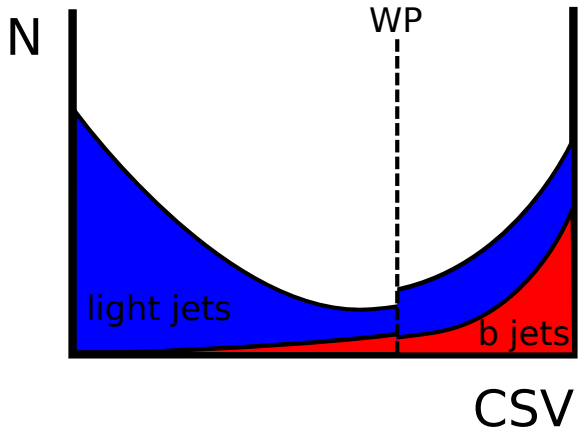


FIGURE 7.14 – Sketch of the calibration on a discriminant variable. The red (blue) corresponds to bottom (light) jets. The calibration rescales the different contributions to the total cross section in bins of transverse momentum, correcting the efficiency of the simulation to the efficiency of the data; the total cross section does not change. Charm jets are neglected in this picture.

and misidentification rates:

$$w_{\text{entry}}(p_T,\eta,f) = \frac{\prod\limits_{i\in\text{tagged}} \text{SF}_i(p_T,f)\epsilon_i(p_T,\eta,f) \prod\limits_{j\in\text{non-tagged}} (1-\text{SF}_j(p_T,f)\epsilon_j(p_T,\eta,f))}{\prod\limits_{i\in\text{tagged}} \epsilon_i(p_T,\eta,f) \prod\limits_{j\in\text{non-tagged}} (1-\epsilon_j(p_T,\eta,f))} \tag{7.25}$$

where f stands for the flavour. The inclusive b jet analysis is a multi-count observable, *i.e.* each jet corresponds to one single entry; therefore, the computation of the weight simplifies to the following [10]:

— if the jet is tagged: $w_{\text{jet}}(p_T,\eta,f) = \text{SF}(p_T,f)$

— else: $w_{\text{jet}}(p_T,\eta,f) = (1-\text{SF}(p_T,f)\epsilon(p_T,\eta,f))/(1-\epsilon(p_T,\eta,f))$

(In practice, the singularity at 1 is always avoided, since the efficiency hardly reaches 0.6.) The point of Eq. **??** is simply that the inclusive jet spectrum remains unchanged by the procedure: the flavours are only rebalanced inside of each $(p_T, |y|)$ bin.

Scale factors. The SFs correspond to ratios of efficiencies in MC and data, and are obtained from specific processes where the efficiency can be measured in data; the SFs are shown in Fig. 7.15 for the tight WP (see App. 7.D for the other WPs). The correction applied to b jets is close to one, as well as for c jets since they are applied the same SFs with larger uncertainties; however, the correction applied to light jets gets larger and larger at higher p_T, corresponding to boosted topologies where the discrimination is more difficult.

Uncertainties. The procedure of calibration comes with uncertainties related to the SFs. The determination of the SFs is performed by performing template fits [11]; the templates are obtained from the simulation, and all the uncertainties on the simulation that may affect the shape of the templates are propagated to determine uncertainties on the SFs. Variations of the shape of the templates are the following:

10. For instance, in the case of the measurement of the mass of the $b\bar{b}$ pair, two b jets would enter the same bin; therefore the computation of the weight would be more sophisticated.

11. A similar procedure is performed in Chap. 8.

— scale variation,

— background subtractions,

— JEC and JER variations,

— pile-up reweighting,

— fragmentation and Gluon Splitting (GSP).

The two last ones are the largest sources of uncertainties in the b calibration.

Application. The application of the procedure on the $p_T(y)$ spectrum can be seen in Fig. 7.16 on the inclusive b jet; it was also checked that the global effect of this reweighting procedure indeed does not change the inclusive jet cross section, since it only corrects the tagging efficiency in the simulation to the one in data.

7.4.2.2 Performance

In order to assess the performance of a tagger and to compare the MC samples, one needs to define the following quantities:

fraction	fraction tagged jets among all jets, *i.e.* $f = N(\text{tagged})/N(\text{all})$
fraction ratio	ratio of the respective fractions in MC and data, *i.e.* $f_{\text{MC}}/f_{\text{data}}$
efficiency	fraction of tagged jets among the true jets, *i.e.* $N(\text{tagged}\vert\text{true})$
mistag	fraction of tagged jets among the non-true jets, *i.e.* $N(\text{tagged}\vert\text{non-true})$
purity	fraction of true jets among the tagged jets, *i.e.* $N(\text{true}\vert\text{tagged})$
contamination	fraction of non-true jets among the tagged jets, *i.e.* $N(\text{non-true}\vert\text{tagged})$

These quantities will be essential for the discussion; in particular, it is crucial to understand the fraction. Indeed, it allows to estimate whether the calibration of b tagging in the simulation is correct.

The performance is shown for the PYTHIA 8 and MADGRAPH samples after tagging with `CSVv2` at the tight WP (`CSVv2T`) in Fig. 7.17, including uncertainties from SFs; the other taggers and WP can be seen in App. 7.D. One can observe a disagreement especially at high p_T, particularly marked in the region $1.0 < |y| < 1.5$. In the different appendices of this chapter are extensive studies conducted in order to investigate possible reasons. Finally, the disagreement will be treated in Chap. 8.

A small discrepancy is also seen in all rapidity bins for $p_T = 400 - 600\,\text{GeV}$; it occurs in all quantities except the fraction ratio where it cancels out; it corresponds most likely to a problem in the training of the tagger [12]; the discrepancy will be solved in the unfolding procedure.

Note that as the calibration is provided assuming PYTHIA 8 hadronisation, an additional calibration would be needed to include HERWIG++. This has not been done here; however, the effect of the calibration not being too large (though still present), including HERWIG++ in the investigations help assessing the agreement, since it is completely independent from PYTHIA 8 and MADGRAPH; this is shown in App. 7.D.

12. Unfortunately, as it can be seen in App. 7.D, it reflects also in the performance of the `cMVAv2` tagger. The `cMVAv2` can therefore not be used as a substitute.

Conclusions

In this chapter, we have investigated in detail the samples of data and simulation. The trigger efficiency has defined the low-p_T boundary of the phase space; the pile-up in the simulation has been corrected to the real pile-up profile; the response of the detector in the jet reconstruction has been described and corrected; finally, the tagging has been investigated. Uncertainties have been associated to all corrections in the simulation.

The difference in the simulation-to-data ratios (Fig. 7.2) for the inclusive jet and inclusive b jet production is confirmed in all situations, regardless of the tagger or of the WP, and does not seem to be related to the sample, since the HERWIG++ sample has confirmed tendencies from PYTHIA 8 and MADGRAPH.

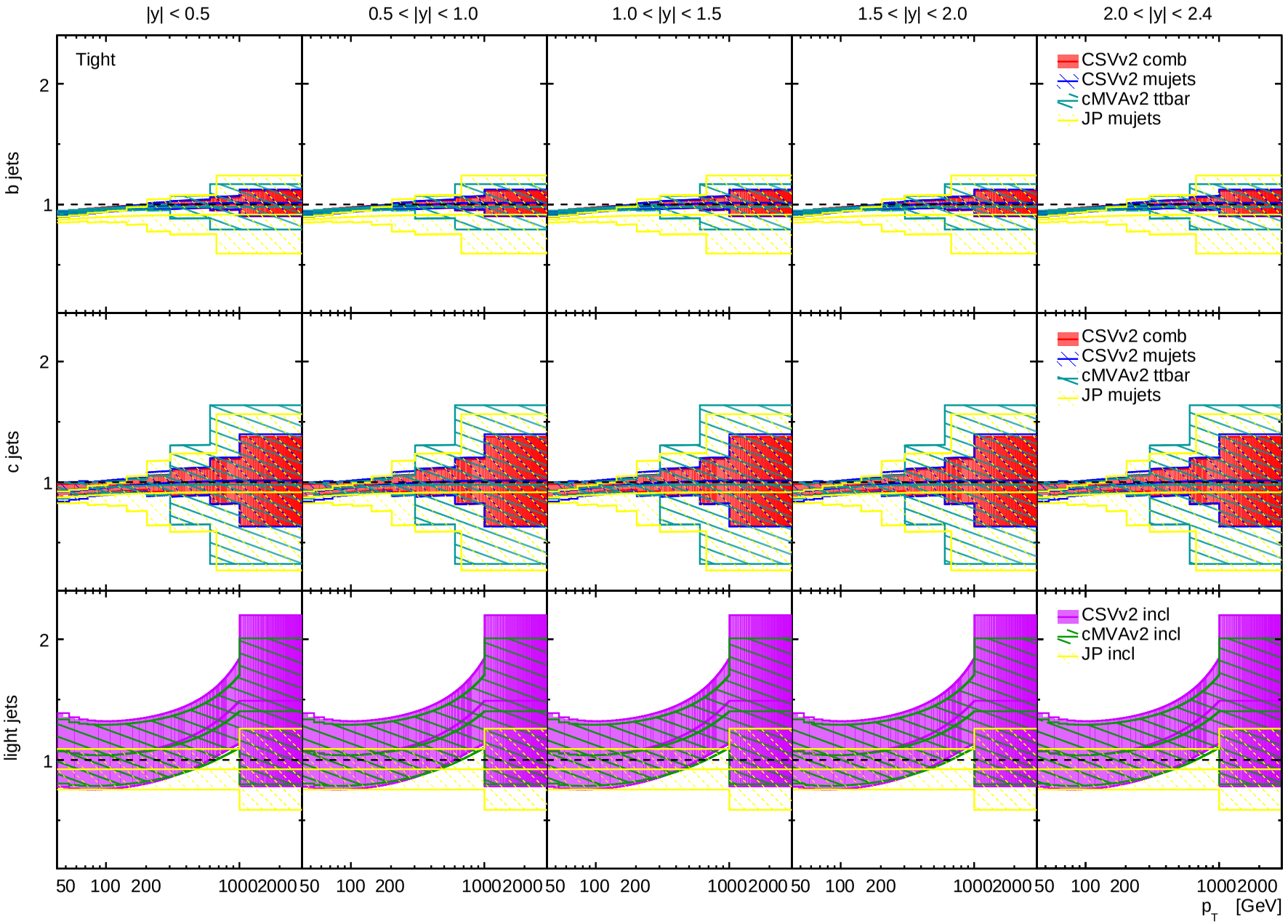
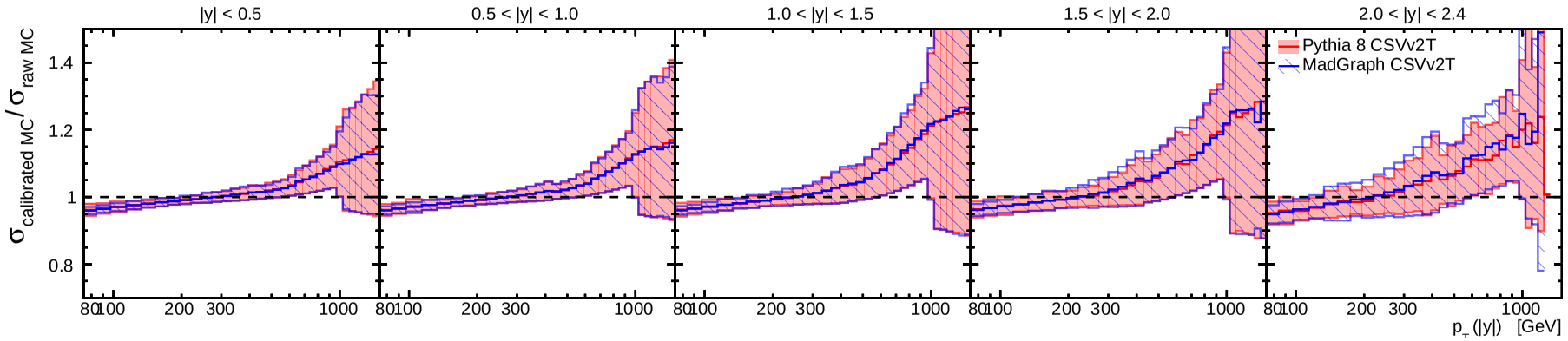
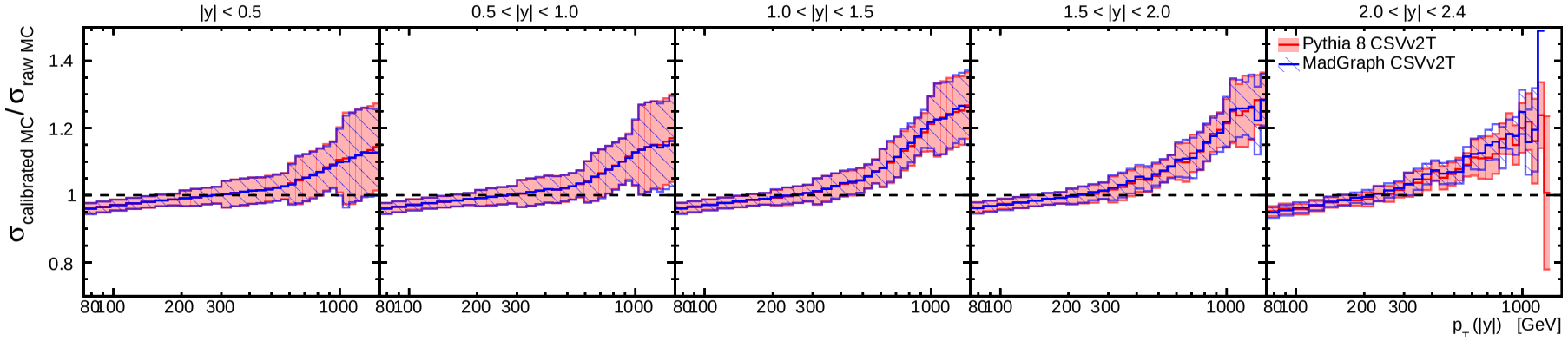


FIGURE 7.15 – Scale factors for the tight selection for the different flavours. No rapidity dependence is included, therefore, the five bins are identical.



(A) Light scale factors



(B) Heavy-flavour scale factors

FIGURE 7.16 – Effect of the b calibration on the spectrum compared for PYTHIA 8 and MADGRAPH. Charm and bottom jets share the same scale factors, as they have very similar properties (charm have doubled uncertainties). The five columns correspond to the rapidity bins.

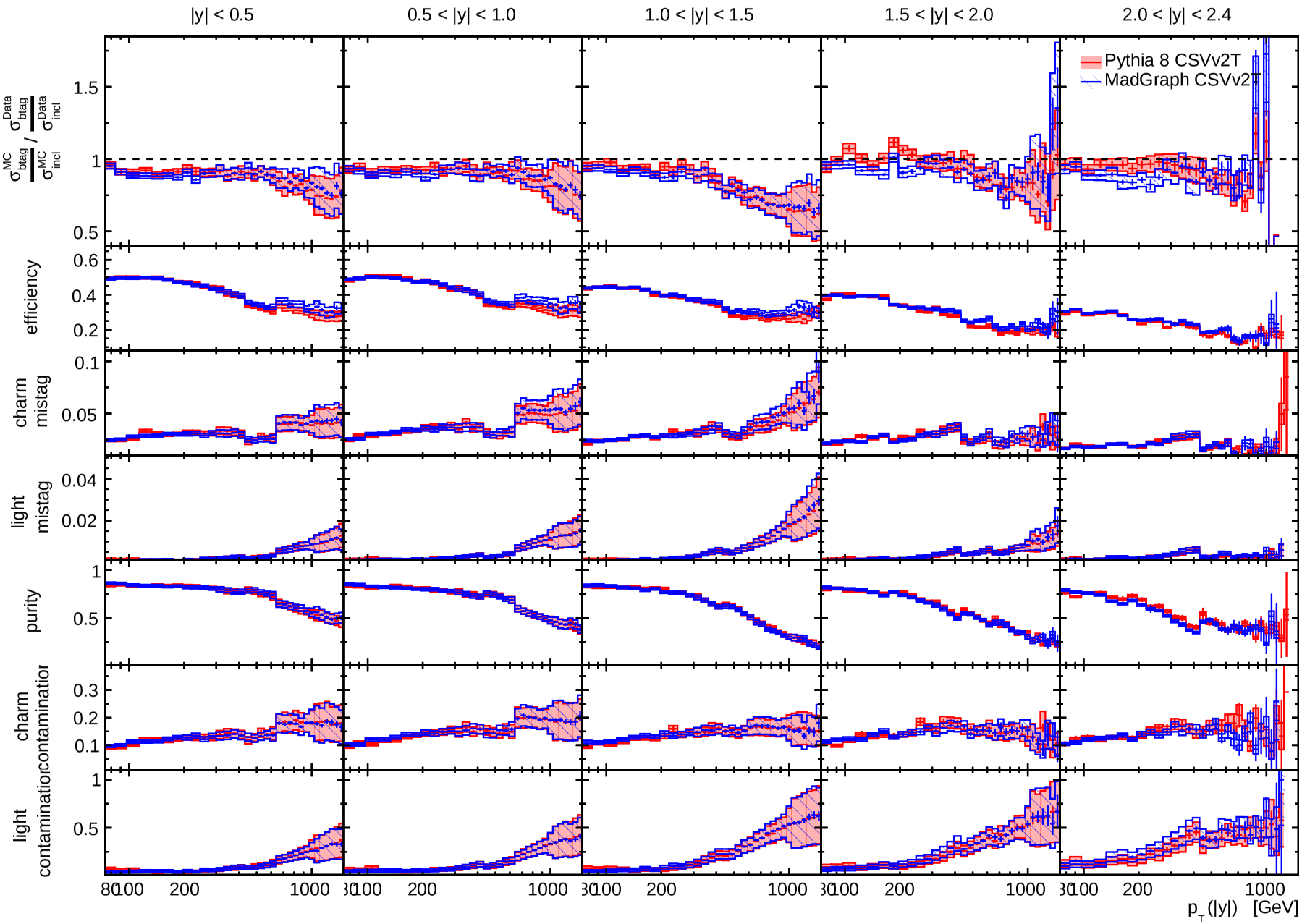


FIGURE 7.17 – Performance of `CSVv2` in simulation after calibration with the tight selection. Calibration is only available for `CUETP8M1`. A significant disagreement from one in the fraction ratio can be seen for both simulations.

7.A Missing Transverse Energy

The *missing transverse momentum* corresponds to the momentum vector imbalance in the perpendicular plane (xy) to the beam axis (z):

$$\mathbf{p}_T^{\text{miss}} = -\sum_{\text{jets}} \mathbf{p}_T^{\text{jet}} \tag{7.26}$$

The Missing Transverse Energy (MET) is the magnitude of the missing transverse momentum:

$$E_T^{\text{miss}} = |\mathbf{p}_T^{\text{miss}}| \tag{7.27}$$

Some physics processes or detector effects may cause it to be significant. In this section, we check these effects in the inclusive jet and inclusive b-tagged jet reconstructed spectra, first at generator level, then at detector level.

7.A.1 Detector studies

General effects have already been studied in the inclusive jet measurements at CMS at 7 and 8 TeV [19, 20]. In the CMS publication at 8 TeV [21], a cut-off on the fraction of MET is applied in order to reduce the contribution from event suffering from significant detector effects with the least effect on physics processes. However, such a cut-off is not applied in the inclusive-jet analysis at 13 TeV with 2015 data [22], nor in the inclusive-b-jet analysis at 7 TeV [23]. The cut-off is very much dependent on the condition of data taking and needs to be checked on a case-by-base basis.

7.A.1.1 MET variables

We first check three variables:

— the MET itself,

— the fraction of MET with respect to the total transverse energy;

— and the azimuthal angle of the MET.

The ratio of MC samples with data is shown in Fig. 7.18. Three series of curves, corresponding to the three rows, are investigated:

1. the three usual MC samples,
2. the inclusive and tagged samples in PYTHIA 8,
3. and the inclusive sample in PYTHIA 8 with different cut-offs on the fraction of MET (0.2, 0.3, 0.4).

In the first row, HERWIG++ shows a slightly different behaviour while PYTHIA 8 and MADGRAPH are very similar, which is most likely related to the respective simulations of the detector (as we said in Sect. 7.2, the simulation of the CMS detector is older for HERWIG++). The dependence on the azimuthal angle of the MET shows a phase, which is likely due to the simulation of the position of the interaction point; since the final measurement does not depend on the azimuthal angle, this phase is not relevant. The second row shows that the tagging behaves similarly for data and MC, therefore the same agreement is seen. Finally, the different cut-offs act the same way for data and MC. The conclusion is that in general, the MET is well simulated in MC, since the agreement with data does not change.

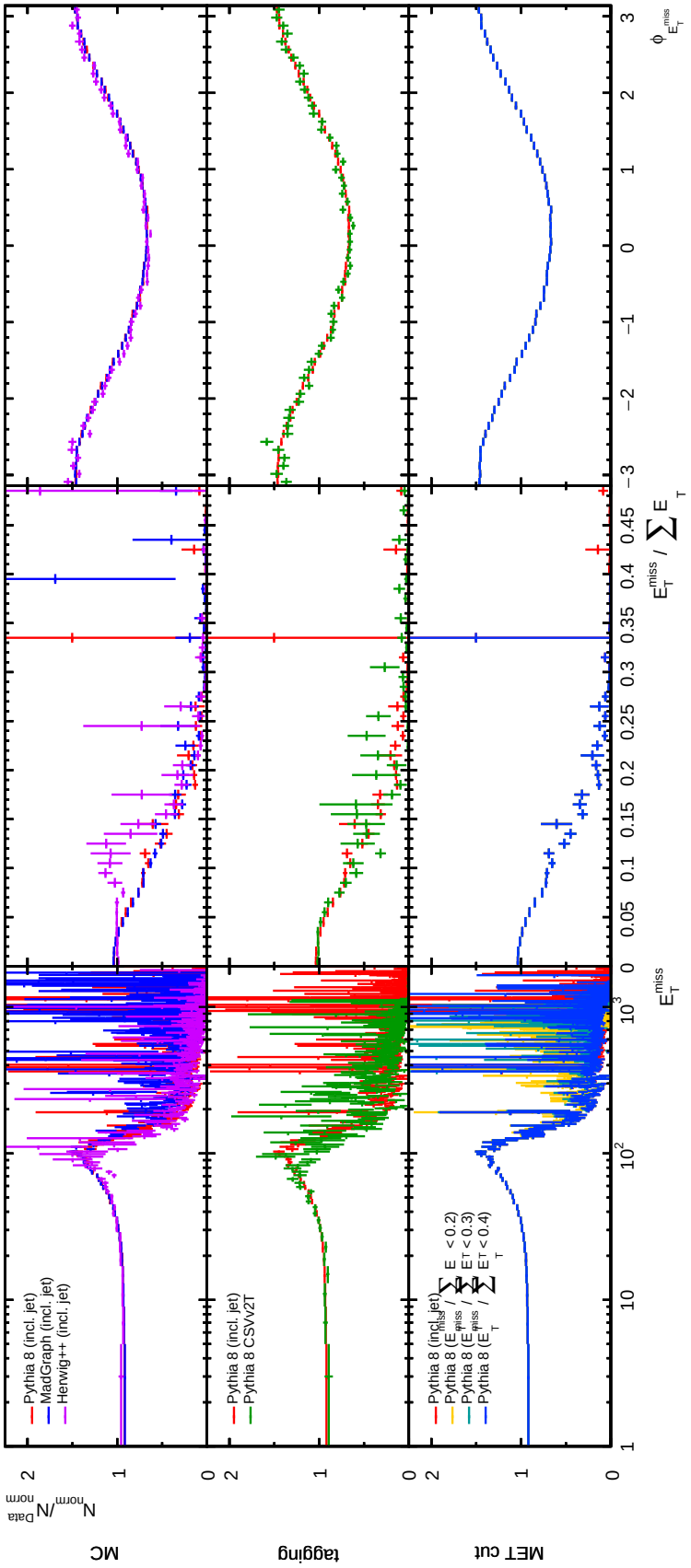


FIGURE 7.18 – Ratios of MC with data for MET, fraction of MET and azimuthal direction of MET are shown (in the columns) for different scenarios (rows).

7.A.1.2 Spectrum of reconstructed transverse momentum

The effect of the cut-off on the fraction of MET for the inclusive jet and inclusive b-tagged jet sample is shown in Figs. 7.19-7.20. In general, an effect starts being visible, though very small, above 1.5 TeV. Since the final measurement will be limited to the region where the calibration of b-tagging discrimination is available, it turns out that the cut-off does not seem relevant for this analysis. However, the effect of the cut-off on different physics processes still needs to be checked.

7.A.2 Generator studies

The effect of the standard cut-off is also checked at generator level in MC studies with PYTHIA 8; it is shown for the first rapidity bin in Fig. 7.21, but the same conclusions may be drawn from other bins. These were done using the Rivet routine of the CMS 2011 b jet analysis [24] and include the contribution of other SM processes, `TTbar`, `W+j` or `Z+j` a.k.a. DY.

In general, there is no significant effect; the difference, for all signals, is of statistical origin. Therefore, applying the cut-off would not affect the signal.

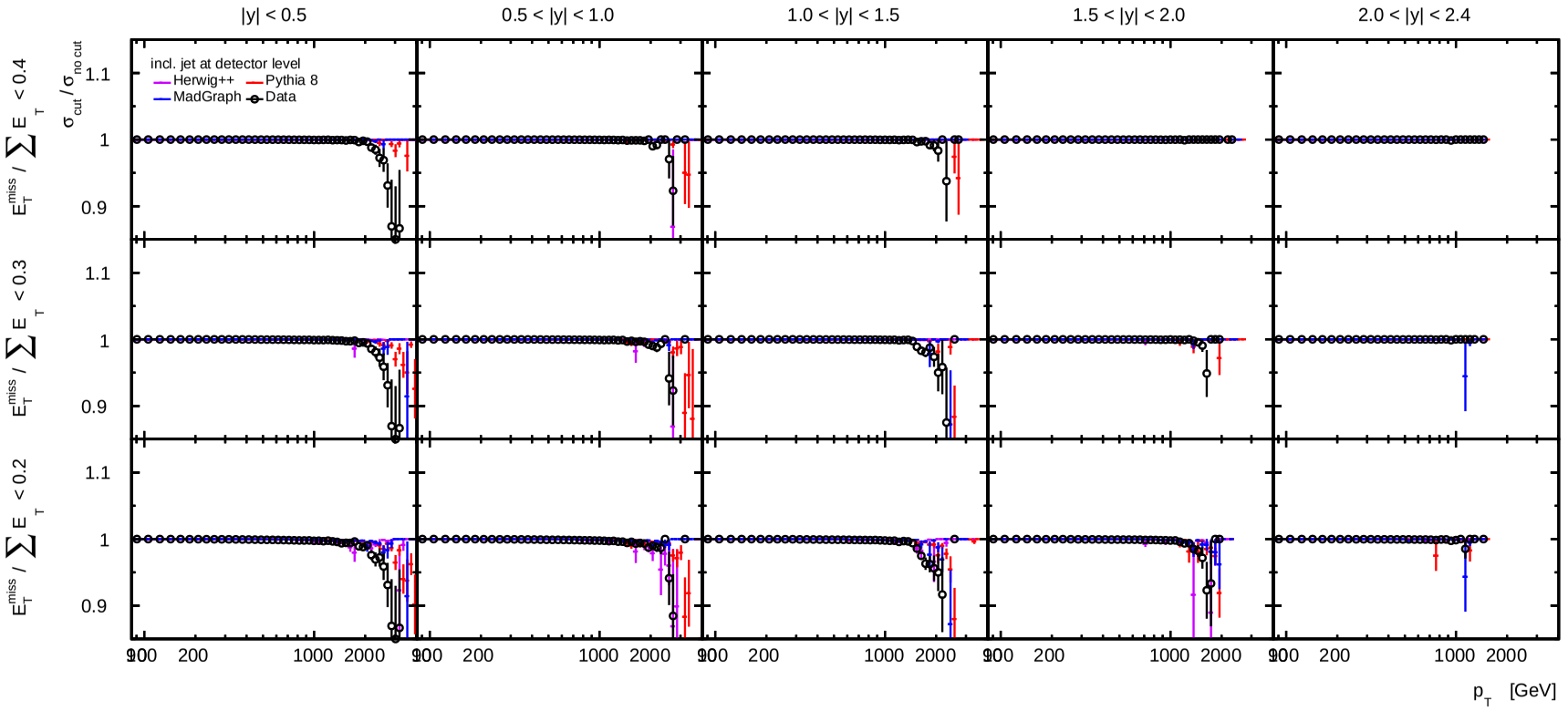
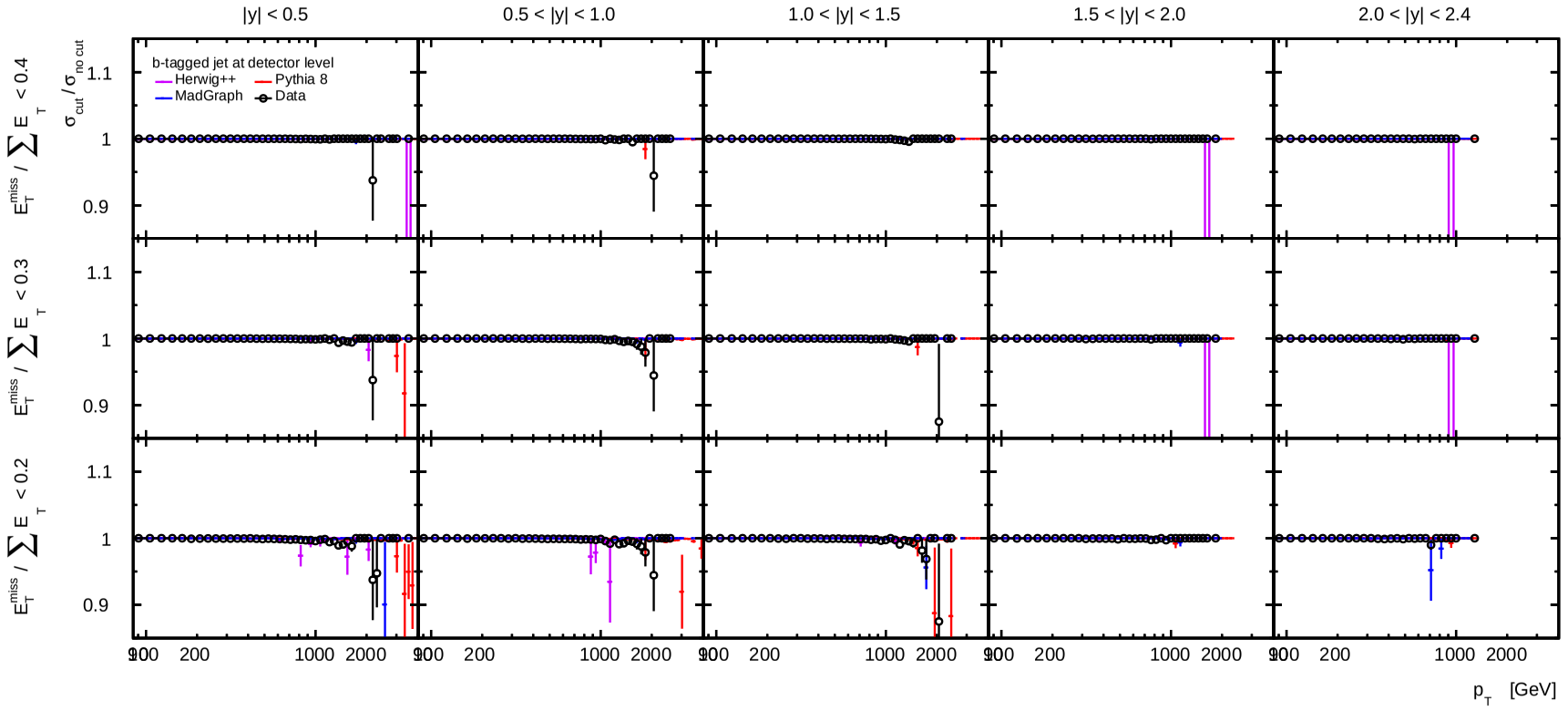
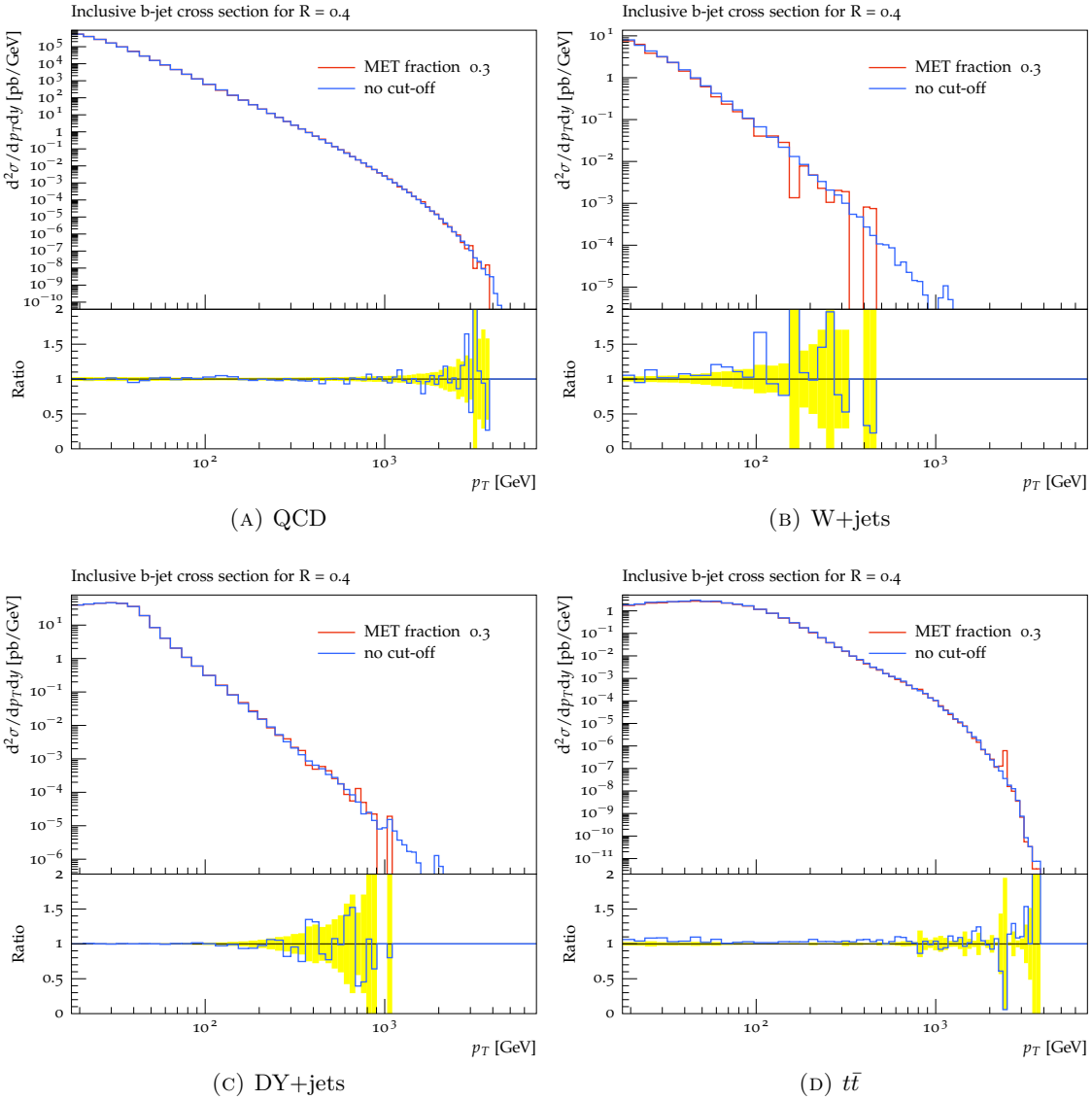


FIGURE 7.19 – The ratios of different samples after and before applying the different values of cut-off on the MET fraction are shown for MC and data. For the inclusive jet selection, an effect may be seen starting from 1.5 TeV.



(A) For the inclusive b-tagged jet, though of smaller magnitude, a similar effect may be seen.

FIGURE 7.20 – The ratios of different samples after and before applying the different values of cut-off on the MET fraction are shown for MC and data. For the inclusive b-tagged jet selection, there is no effect up to 1 TeV.



(A) QCD (B) W+jets

(C) DY+jets (D) $t\bar{t}$

FIGURE 7.21 – Effect of the cut-off on the fraction of MET for the four main signals. It is shown for the four main contribution in the SM in the most central rapidity bin.

7.B Details of trigger efficiency

Tables for turn-on values. In Sect. 7.1.1.1, while presenting the determination of the trigger efficiencies, we presented the emulation and the Tag & Probe methods; they are given in Tabs. 7.3 for each trigger, per era and per rapidity bin.

Representation of the subdivision of the phase space. In Sect. 7.1.1.2, the final choice of the turn-on values was described. In Fig. 7.22, one can visualise the subdivision of the phase space. In complement, the number of events per trigger is given in Tab. 7.4 per era and per rapidity bin.

Run-by-run average pre-scales. One can define the pre-scales as follows:

$$f^{\text{run}}_{\text{trigger}} = \mathcal{L}^{\text{run}}_{\text{total}} / \mathcal{L}^{\text{run}}_{\text{trigger}} \tag{7.28}$$

		emulation method					Tag & Probe method				
trigger	era	$\lvert y\rvert < 0.5$	$0.5 < \lvert y\rvert < 1.0$	$1.0 < \lvert y\rvert < 1.5$	$1.5 < \lvert y\rvert < 2.0$	$2.0 < \lvert y\rvert < 2.4$	$\lvert y\rvert < 0.5$	$0.5 < \lvert y\rvert < 1.0$	$1.0 < \lvert y\rvert < 1.5$	$1.5 < \lvert y\rvert < 2.0$	$2.0 < \lvert y\rvert < 2.4$
40	BCD						66	66	65	59	55
	EFearly						65	65	64	58	54
	FlateG						66	67	64	60	54
	H						65	66	65	60	56
60	BCD	77	77	77	70	66	77	77	74	69	66
	EFearly	77	78	76	69	66	77	77	74	70	61
	FlateG	77	78	77	70	68	76	77	74	69	64
	H	77	77	76	71	66	77	77	77	72	180
80	BCD	113	114	92	91	90	114	114	110	103	101
	EFearly	113	113	91	91	91	113	110	107	102	97
	FlateG	112	113	80	91	88	113	112	110	101	98
	H	110	112	92	91	91	113	109	108	102	100
140	BCD	179	179	173	164	162	179	179	173	151	162
	EFearly	177	177	171	164	162	177	177	173	150	164
	FlateG	177	179	173	165	160	177	177	171	155	164
	H	177	179	169	164	150	179	177	171	169	150
200	BCD	244	247	239	236	228	242	242	239	228	226
	EFearly	242	244	261	223	228	244	244	239	226	231
	FlateG	242	244	242	231	226	244	242	236	228	226
	H	242	242	244	223	236	236	242	236	228	226
260	BCD	311	314	304	281	294	314	311	300	281	294
	EFearly	311	307	291	284	291	314	307	294	281	291
	FlateG	314	314	291	278	284	314	311	297	287	291
	H	311	314	291	284	291	314	311	294	278	291
320	BCD	366	366	346	346	358	382	350	362	346	362
	EFearly	370	350	350	346	350	378	378	350	346	362
	FlateG	374	350	350	346	346	382	382	346	346	346
	H	350	370	350	346	350	378	374	350	346	350
400	BCD	447	447	400	447	447	452	452	447	447	447
	EFearly	447	447	442	442	447	452	452	447	447	447
	FlateG	452	447	447	442	447	452	452	447	447	447
	H	452	447	447	447	447	452	452	447	447	447
450	BCD	492	486	486	486	492	492	492	486	486	492
	EFearly	492	486	486	486	492	492	492	486	486	492
	FlateG	492	492	486	486	486	492	492	486	486	486
	H	492	492	486	486	492	492	492	486	486	492

TABLE 7.3 – Turn-on points for trigger efficiency in bins of rapidity per trigger and era, shown for the two methods.

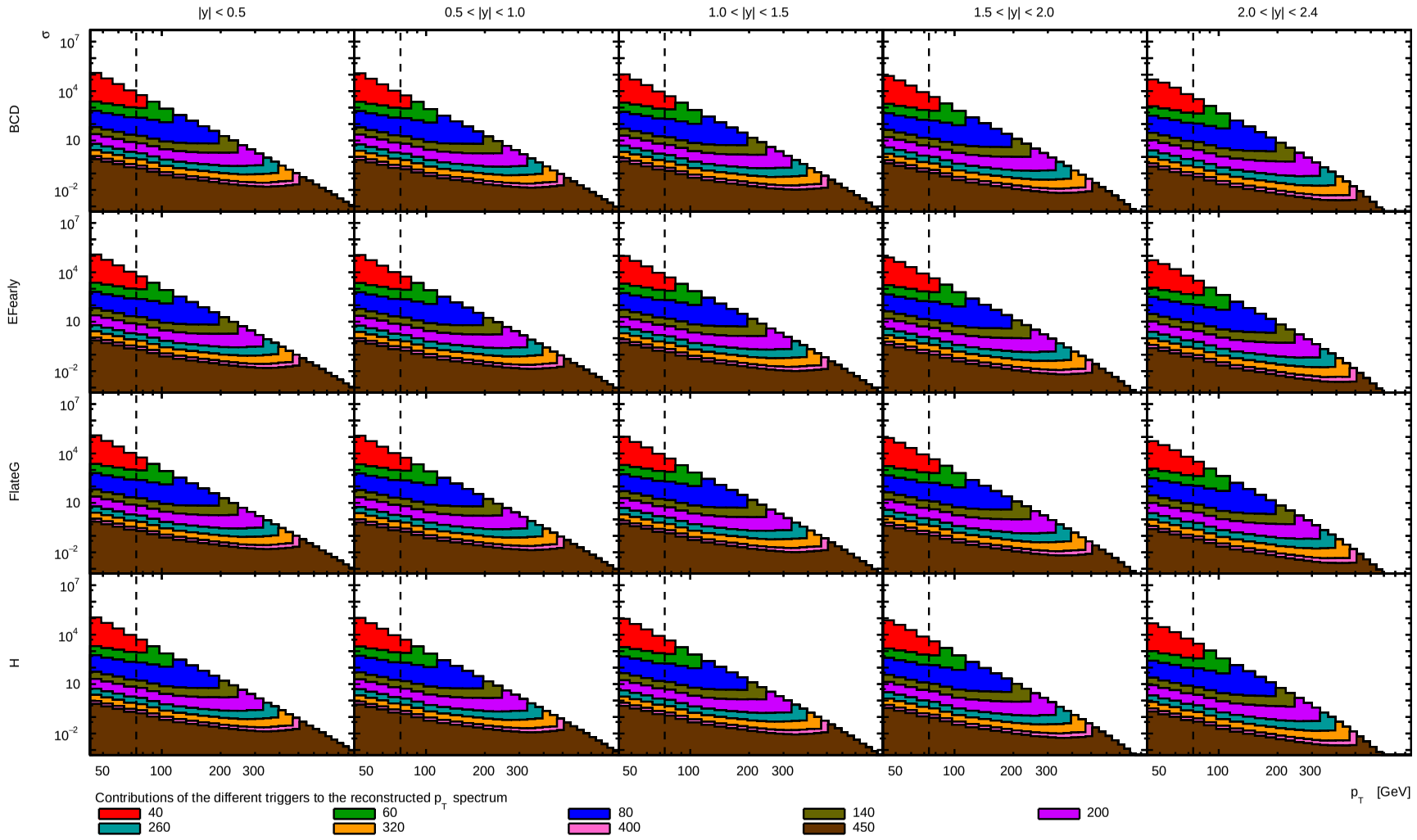


FIGURE 7.22 – The different contributions from the different triggers to the total cross section are stacked in bins of rapidity and per era.

trigger	era	$\lvert y\rvert < 0.5$	$0.5 < \lvert y\rvert < 1.0$	$1.0 < \lvert y\rvert < 1.5$	$1.5 < \lvert y\rvert < 2.0$	$2.0 < \lvert y\rvert < 2.4$
40	BCD	868500	828016	763223	692237	488531
	EFearly	442347	419102	387150	345752	253388
	FlateG	443188	424510	391895	359266	256978
	H	445452	423063	390693	350886	253246
60	BCD	255272	243855	214973	189559	132760
	EFearly	106111	100988	89955	77792	53484
	FlateG	103910	100259	88543	78389	53386
	H	88482	85385	75478	66468	45574
80	BCD	487734	460517	401419	343303	224716
	EFearly	127217	120671	104117	88841	56976
	FlateG	123821	117553	102022	87289	55822
	H	107724	101845	88748	75023	48389
140	BCD	296982	276715	236009	192082	116147
	EFearly	102830	95836	82035	66477	39104
	FlateG	93642	86624	74113	60087	35394
	H	78884	74281	63098	50886	29596
200	BCD	633278	587602	488492	385465	221183
	EFearly	206868	191584	158455	125192	70133
	FlateG	137591	127759	106361	84264	46584
	H	116448	107935	90157	70600	39153
260	BCD	561595	514174	419367	317151	168201
	EFearly	218751	200452	163406	122542	63516
	FlateG	230859	211254	171348	129581	67287
	H	195725	179780	146313	110239	56299
320	BCD	691473	625971	501672	365205	184002
	EFearly	265658	241153	192068	139358	68677
	FlateG	285935	260335	207869	150886	73817
	H	244515	222672	177559	129016	62229
400	BCD	497249	447734	352633	246763	118903
	EFearly	200072	180630	141768	99387	46147
	FlateG	211914	191209	150252	105383	49041
	H	180553	163071	127227	90076	41326
450	BCD	5277234	4677644	3528306	2333147	1057756
	EFearly	2731499	2428308	1829205	1200886	536510
	FlateG	3300980	2929738	2214669	1465589	648822
	H	2964714	2637357	1995284	1314709	577339

TABLE 7.4 – Number of jets per trigger and per era in rapidity bins, after the final phase space subdivision.

Then, Eq. 7.7 can also be written as follows:

$$\sigma^{\text{run}}_{\text{trigger}} = \frac{f^{\text{run}}_{\text{trigger}} N^{\text{run}}_{\text{trigger}}}{\mathcal{L}^{\text{run}}_{\text{total}}} \tag{7.29}$$

The run-by-run averaged trigger pre-scales in Fig. 7.23; as it can be seen, the variations are significant. (Similarly, as it was already mentioned in Chap. 3, Fig. 3.8, the luminosity can become smaller by a factor of two during a single run, and pre-scales can be adapted on-line to compensate.) Higher instantaneous luminosity implies higher pile-up conditions, and therefore higher pre-scales. The figure shows that, along the year of 2016, the LHC has achieved better and better performances in terms of luminosity.

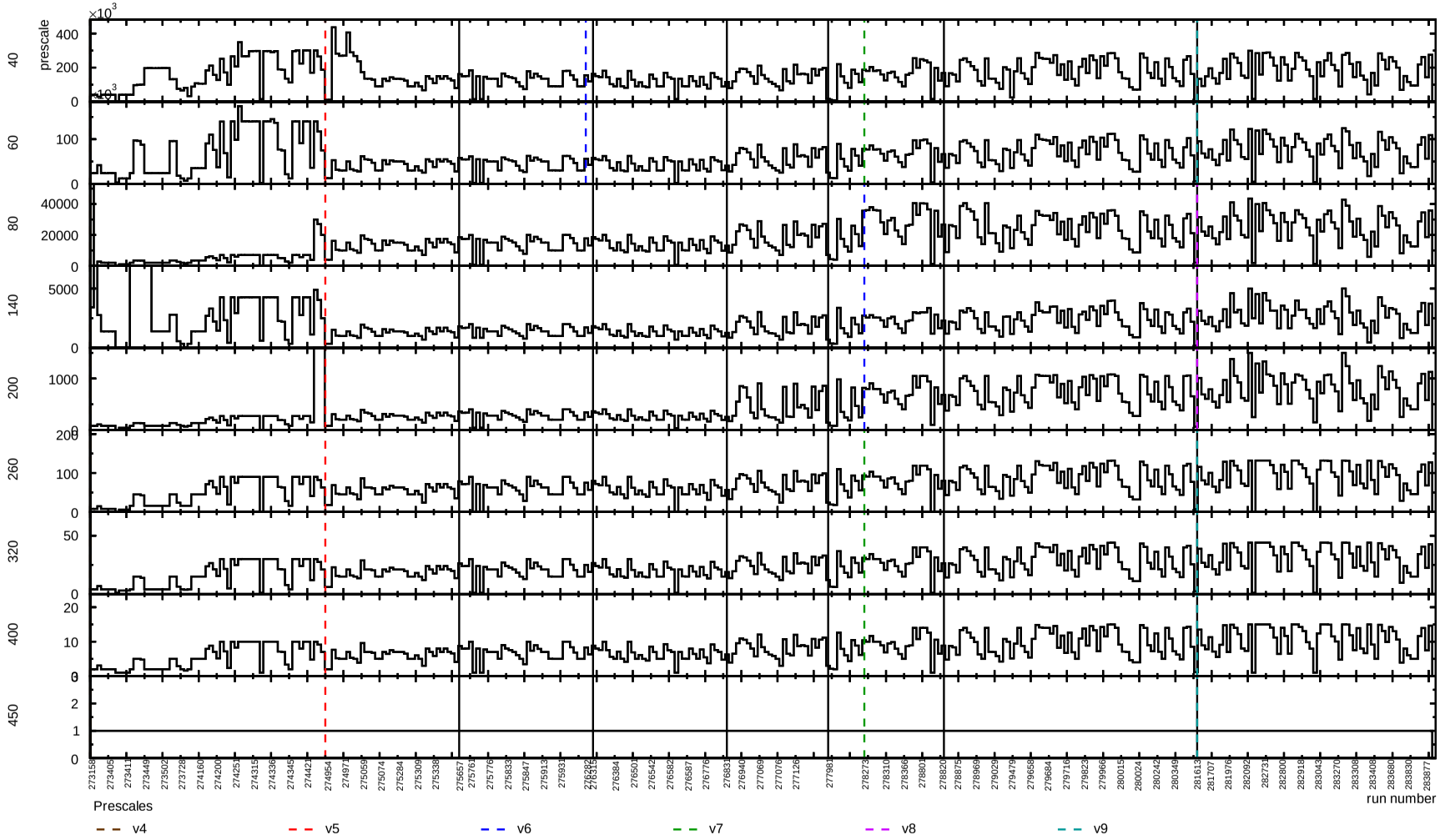


FIGURE 7.23 – Run-by-run averages of the pre-scales for each of the triggers used in this analysis (the threshold is indicated on the very left). 372 runs are shown, but only a few of them can be explicitly written for readability. Black vertical lines separate the different eras and the changes of version are shown by dashed lines.

7.C Additional control plots about jets

Angular response. The angular response of the detector for jet reconstruction is compared in data and in the simulation in Fig. 7.24; it can be seen that the imperfections of the detector are well described.

More on binning. The following binning in considered in this analysis:

p_T 43,49,56,64,74, 84, 96, 114, 133, 153, 174, 196, 220, 245, 272, 300, 330, 362, 395, 430, 468,507, 548, 592, 638, 686, 737, 790, 846, 905, 967,1032, 1101, 1172, 1248, 1327, 1410, 1497, 1588, 1684, 1784, 1890, 2000,2116, 2238, 2366, 2500, 2640, 2787, 2941

$|y|$ 0, 0.5, 1.0, 1.5, 2.0, 2.4

Additional checks are shown in Fig. 7.25. Most importantly, the resolution for b jets is slightly worse, especially in the forward region ($|y| > 1.5$); we will see later that during the procedure of unfolding, bins have to be merged roughly in pairs, which will cover the migrations.

7.D More on tagging

We show in Fig. 7.26 the performance of the tagger before calibration. In this case, one can also compare to HERWIG++: the performance is extremely similar, and the discrepancy at high transverse momentum is even more pronounced.

Additional checks may be performed by comparing taggers and WPs.

7.D.1 Comparison of the working points

The SFs have been shown on Fig. 7.15 for the tight selection. The medium and loose WPs are also investigated, and they can be seen on Figs. 7.27-7.28; the effect of different WPs on the $p_T(y)$ can be seen on Fig. 7.29. Allowing larger statistics, the SFs for the mistag of light jets can be provided with rapidity dependence in addition to transverse momentum dependence in order to attempt to mitigate the disagreement, especially in the $1.0 < |y| < 1.5$ region where it is the strongest. The rapidity dependence is defined in different binning schemes according to the WP:

— for the medium selection, it is done in three bins of width 0.8, *i.e.* $0.0, 0.8, 1.6, 2.4$;

— for the loose selection, it is done in seven bins of different widths, *i.e.* $0.0, 0.3, 0.6, 0.9, 1.2, 1.5, 1.8, 2.4$.

From Fig. 7.31, the fraction ratio is better described; on the other hand, the contamination from light (charm) jets reaches 60% (20%), and the purity is around $10-20\%$ in the whole spectrum. The improvement seen in the fraction can be explained by the presence of more n jets. Therefore, we cannot conclude any improvement from the rapidity dependence of the SFs with looser WPs. The same conclusions may be drawn with the different WPs of the different taggers (not shown here).

7.D.2 Comparison of the taggers

The different taggers can be compared on Fig. 7.32. The uncertainties from the `CSVv2` (`JP`) tagger are the smallest (greatest) one. However, they all show similar tendencies, and confirm that possible biases are not due to their respective performances.

7.E Jet constituents

The content of jets and of $\hat{b}$ jets is described by two categories of variables:

The *energy fractions* (Fig. 7.33):
- charged-hadron energy fraction
- neutral-hadron energy fraction
- charged e.m. energy fraction
- neutral e.m. energy fraction
- muon energy fraction

The *multiplicities* (Fig. 7.34):
- charged-hadron multiplicity
- neutral-hadron multiplicity
- electron multiplicity
- photon multiplicity
- muon multiplicity

(The figures for the energy fractions and for the multiplicities are shown opposite to one another.)

7.E.1 Jet ID

The *jet ID*, already addressed while describing the selection in Sect. 6.3.1, is based on these variables. In this analysis, the *tight ID* is used, to which the corresponding cut-off values for $|y| < 2.4$ are shown in Tab. 7.5.

PF Jet ID	Tight
neutral hadron fraction	< 0.90
neutral e.m. fraction	< 0.90
number of constituents	> 1
charged hadron fraction	> 0
charged multiplicity	> 0
charged e.m. fraction	< 0.99

TABLE 7.5 – Tight jet ID definition in $|y| < 2.4$.

7.E.2 Jet constituents in bins of rapidity

In order to investigate the discrepancy in the fraction ratio (Sect. 7.4), we show some elements of additional investigations on the jet constituents in bins of rapidity.

Figs. 7.33-7.34 show the agreement with data in bins of rapidity, after the tight jet ID selection, for inclusive jet and inclusive $\hat{b}$ jet production. Statistical uncertainties are included, but systematical uncertainties have not been investigated.

Despite the different showering used, it is interesting that PYTHIA 8 and MADGRAPH on the one hand and HERWIG++ and the other hand do not show any large difference on any of these variables. In general, only the variables involving neutral particles show a sensitive difference (Figs. 7.33b, 7.33d, 7.34b and 7.34d). But most importantly, the agreement is not affected by the b tagging.

This having been said, the statistics is usually low, and it is hard to conclude. To be perfectly rigorous in our investigations, the same investigations should be performed in bins of p_T; unfortunately, this is not possible, since the statistics are too low.

References

[1] Butz, Erik. *Mitigation of the Strip Tracker Dynamic Inefficiency (previously known as HIP)*. https://indico.cern.ch/event/560226/contributions/2277448/attachments/1324704/1988050/wgm_vfp_change_ebutz.pdf. Accessed: 2018-01-25.

[2] CMS Collaboration. "Pileup Jet Identification". In: (2013).

[3] CMS Collaboration. "Pileup Removal Algorithms". In: (2014).

[4] David Krohn et al. "Jet Cleansing: Pileup Removal at High Luminosity". In: *Phys. Rev.* D90.6 (2014), p. 065020. DOI: 10.1103/PhysRevD.90.065020. arXiv: 1309.4777 [hep-ph].

[5] The CMS collaboration. "Determination of jet energy calibration and transverse momentum resolution in CMS". In: *Journal of Instrumentation* 6.11 (2011), P11002. URL: http://stacks.iop.org/1748-0221/6/i=11/a=P11002.

[6] CMS Collaboration. "CMS Luminosity Measurements for the 2016 Data Taking Period". In: (2017).

[7] Salfeld, Jakob. *Utilities for Accessing Pileup Information for Data*. https://twiki.cern.ch/twiki/bin/view/CMS/PileupJSONFileforData. CMS Private area, Accessed: 2018-01-25.

[8] Matteo Cacciari and Gavin P. Salam. "Pileup subtraction using jet areas". In: *Phys. Lett.* B659 (2008), pp. 119–126. DOI: 10.1016/j.physletb.2007.09.077. arXiv: 0707.1378 [hep-ph].

[9] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. "The Catchment Area of Jets". In: *JHEP* 04 (2008), p. 005. DOI: 10.1088/1126-6708/2008/04/005. arXiv: 0802.1188 [hep-ph].

[10] CMS Collaboration. "Jet algorithms performance in 13 TeV data". In: (2017).

[11] Vardan Khachatryan et al. "Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV". In: *JINST* 12.02 (2017), P02014. DOI: 10.1088/1748-0221/12/02/P02014. arXiv: 1607.03663 [hep-ex].

[12] Viola Sordini. *Jet Energy Resolution*. https://twiki.cern.ch/twiki/bin/viewauth/CMS/JetMET. CMS Private area, Accessed: 2018-01-26.

[13] Serguei Chatrchyan et al. "Identification of b-quark jets with the CMS experiment". In: *JINST* 8 (2013), P04013. DOI: 10.1088/1748-0221/8/04/P04013. arXiv: 1211.4462 [hep-ex].

[14] Luca Scodellaro. "b tagging in ATLAS and CMS". In: *5th Large Hadron Collider Physics Conference (LHCP 2017) Shanghai, China, May 15-20, 2017*. 2017. arXiv: 1709.01290 [hep-ex]. URL: https://inspirehep.net/record/1621595/files/arXiv:1709.01290.pdf.

[15] Albert M Sirunyan et al. "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV". In: (2017). arXiv: 1712.07158 [physics.ins-det].

[16] D. Buskulic et al. "A precise measurement of $\Gamma\ Z \to bb/\Gamma\ Z \to$ hadrons". In: *Physics Letters B* 313.3 (1993), pp. 535–548. ISSN: 0370-2693. DOI: https://doi.org/10.1016/0370-2693(93)90028-G. URL: http://www.sciencedirect.com/science/article/pii/037026939390028G.

[17] G. Borisov and C. Mariotti. "Fine tuning of track impact parameter resolution of the DELPHI detector". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 372.1 (1996), pp. 181–187. ISSN: 0168-9002. DOI: https://doi.org/10.1016/0168-9002(95)01287-7. URL: http://www.sciencedirect.com/science/article/pii/0168900295012877.

[18] Andra Rizzi and Dinko Ferencek. *B Tagging and Vertexing*. https://twiki.cern.ch/twiki/bin/view/CMS/BTagSFMethods. CMS Private area, Accessed: 2018-01-27.

[19] The CMS collaboration. "Missing transverse energy performance of the CMS detector". In: *Journal of Instrumentation* 6.09 (2011), P09001. URL: http://stacks.iop.org/1748-0221/6/i=09/a=P09001.

[20] The CMS collaboration. "Performance of the CMS missing transverse momentum reconstruction in pp data at $\sqrt{s} = 8\ TeV$". In: *Journal of Instrumentation* 10.02 (2015), P02006. URL: http://stacks.iop.org/1748-0221/10/i=02/a=P02006.

[21] Vardan Khachatryan et al. "Measurement and QCD analysis of double-differential inclusive jet cross sections in pp collisions at $\sqrt{s} = 8$ TeV and cross section ratios to 2.76 and 7 TeV". In: *JHEP* 03 (2017), p. 156. DOI: 10.1007/JHEP03(2017)156. arXiv: 1609.05331 [hep-ex].

[22] Vardan Khachatryan et al. "Measurement of the double-differential inclusive jet cross section in proton – proton collisions at $\sqrt{s} = 13\,\text{TeV}$". In: *Eur. Phys. J.* C76.8 (2016), p. 451. DOI: 10.1140/epjc/s10052-016-4286-3. arXiv: 1605.04436 [hep-ex].

[23] Serguei Chatrchyan et al. "Inclusive b-jet production in pp collisions at $\sqrt{s} = 7$ TeV". In: *JHEP* 04 (2012), p. 084. DOI: 10.1007/JHEP04(2012)084. arXiv: 1202.4617 [hep-ex].

[24] Andy Buckley et al. "Rivet user manual". In: *Comput. Phys. Commun.* 184 (2013), pp. 2803–2819. DOI: 10.1016/j.cpc.2013.05.021. arXiv: 1003.0694 [hep-ph].

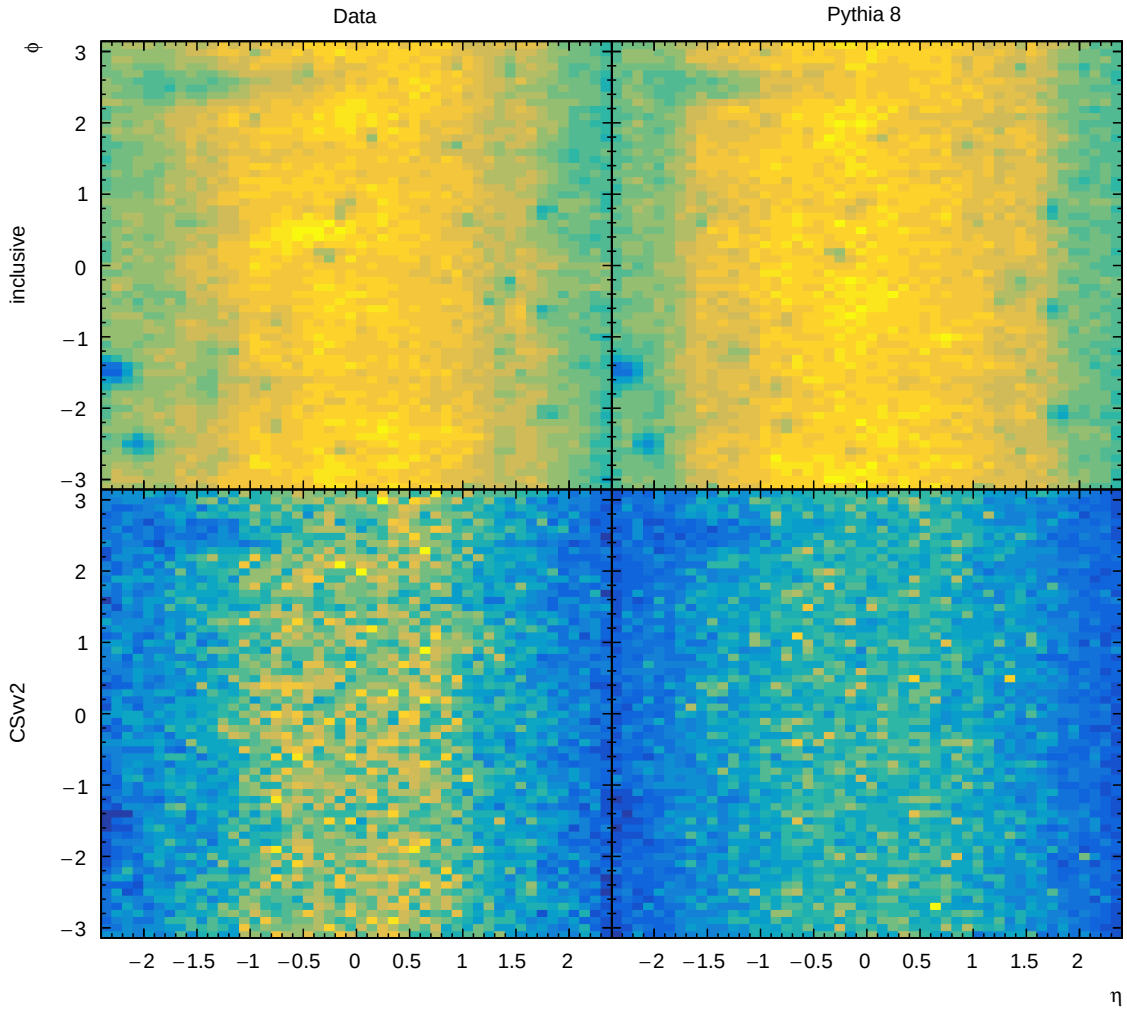
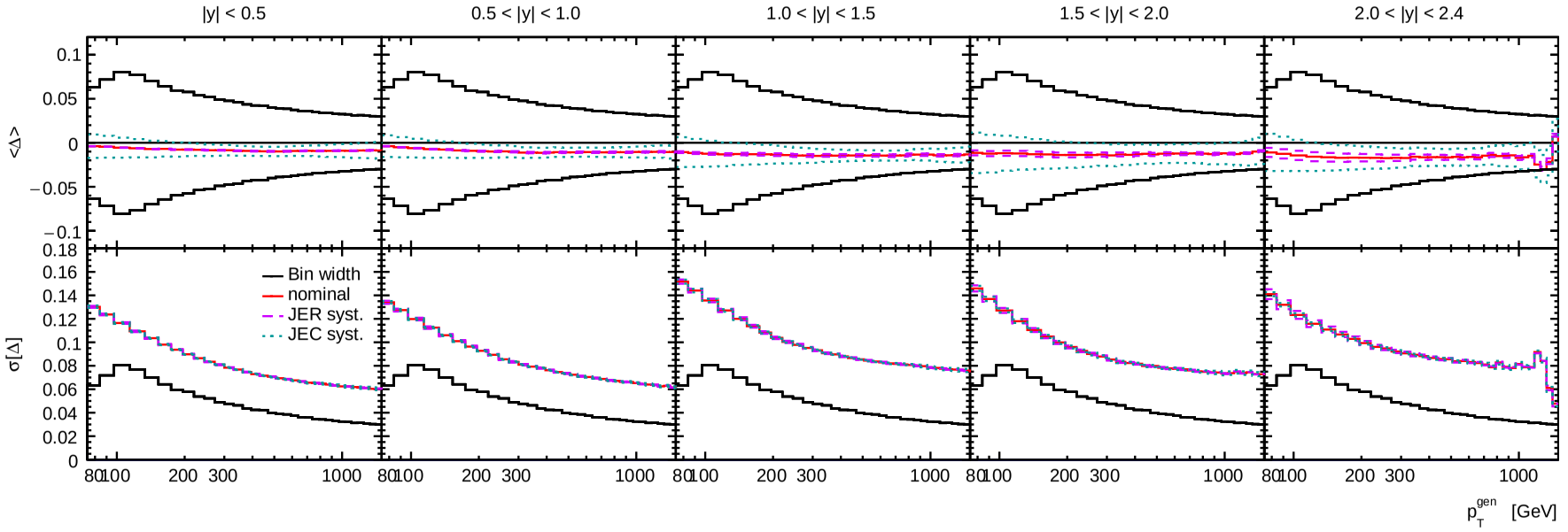
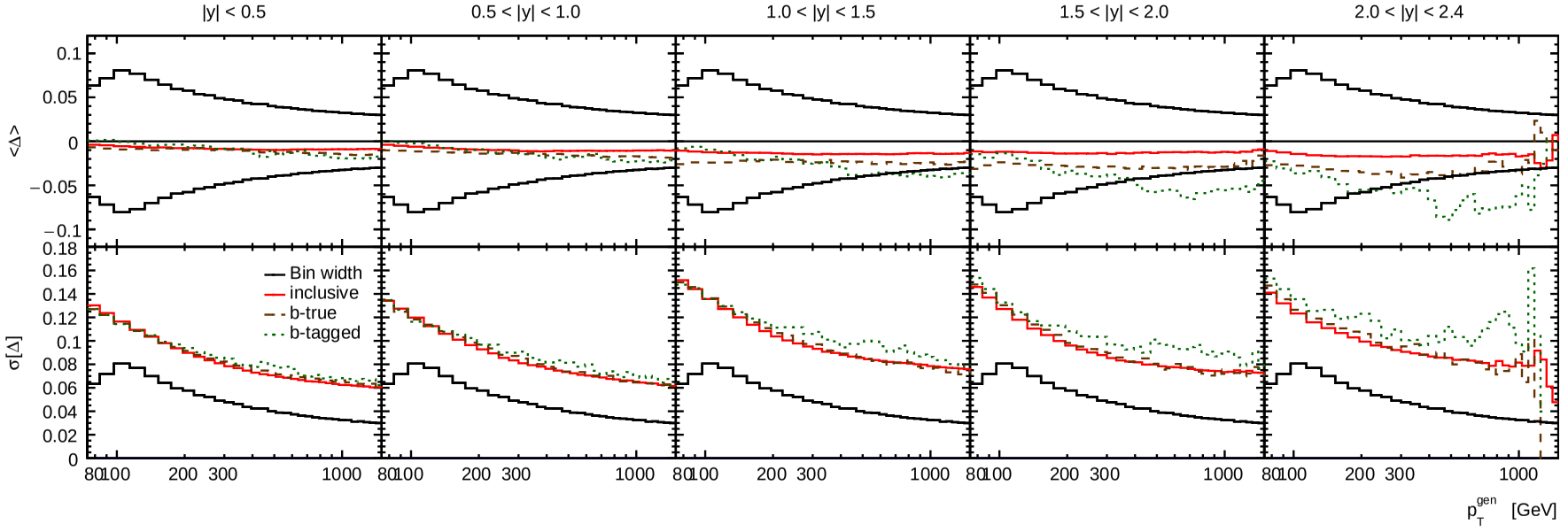


FIGURE 7.24 – Jet $\eta - \phi$ spectrum in data (left) and simulation (right), for inclusive jet (top) and inclusive b jet (bottom). The bins corresponds to the count of jets (not to a cross section) with arbitrary normalisation.



(A) Uncertainties from JEC and JER.



(B) Resolution for b (tagged) jets.

FIGURE 7.25 – Resolution profiles for systematic shifts and for flavoured samples.

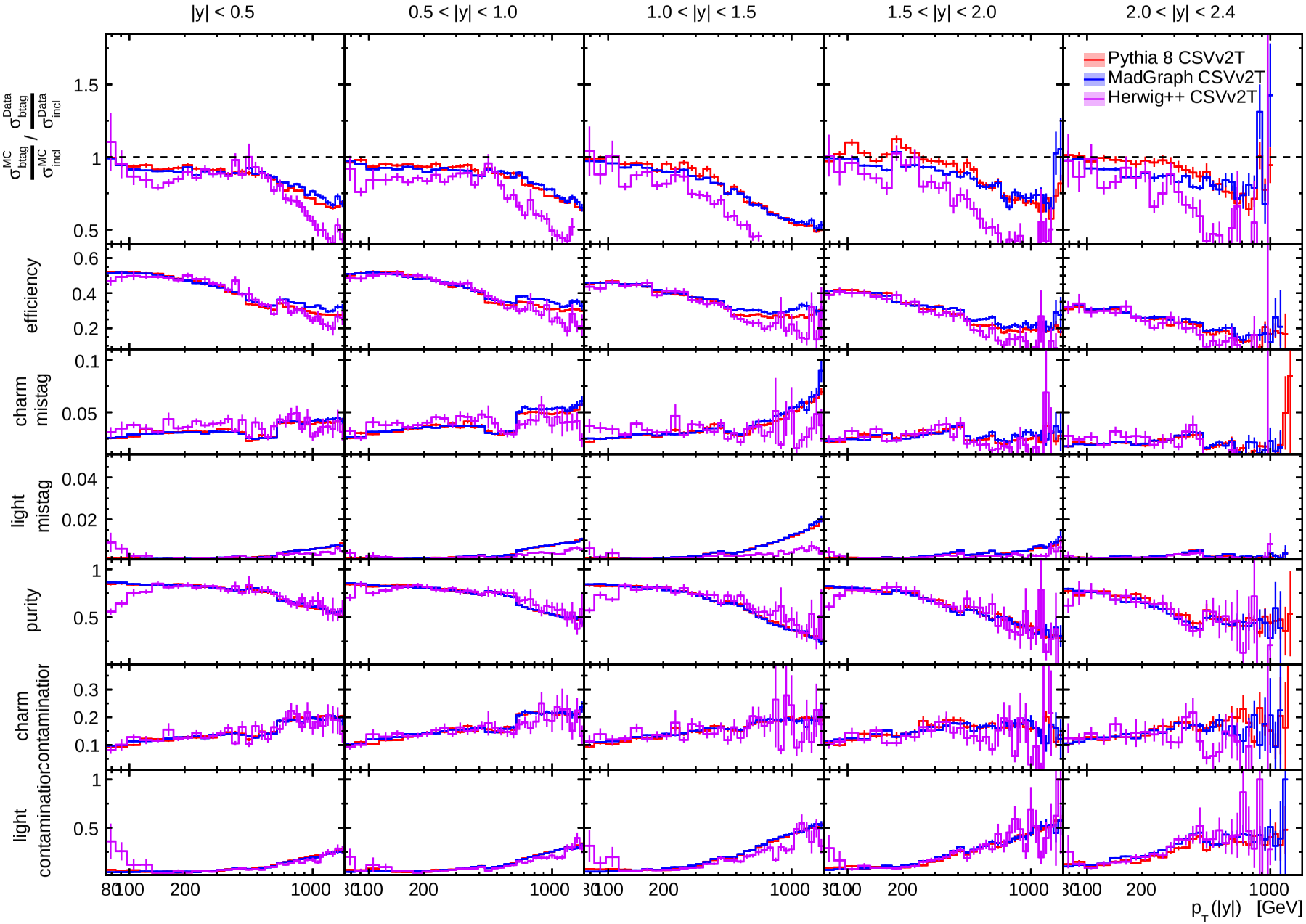


FIGURE 7.26 – Performance of `CSVv2` in simulation before calibration with the tight selection. PYTHIA 8 and MADGRAPH share the same tune `CUETP8M1` while HERWIG++ has its own tune `CUETHppS1`; calibration is only available for PYTHIA 8 and MADGRAPH. A significant disagreement from one in the fraction ratio can be seen for all simulations.

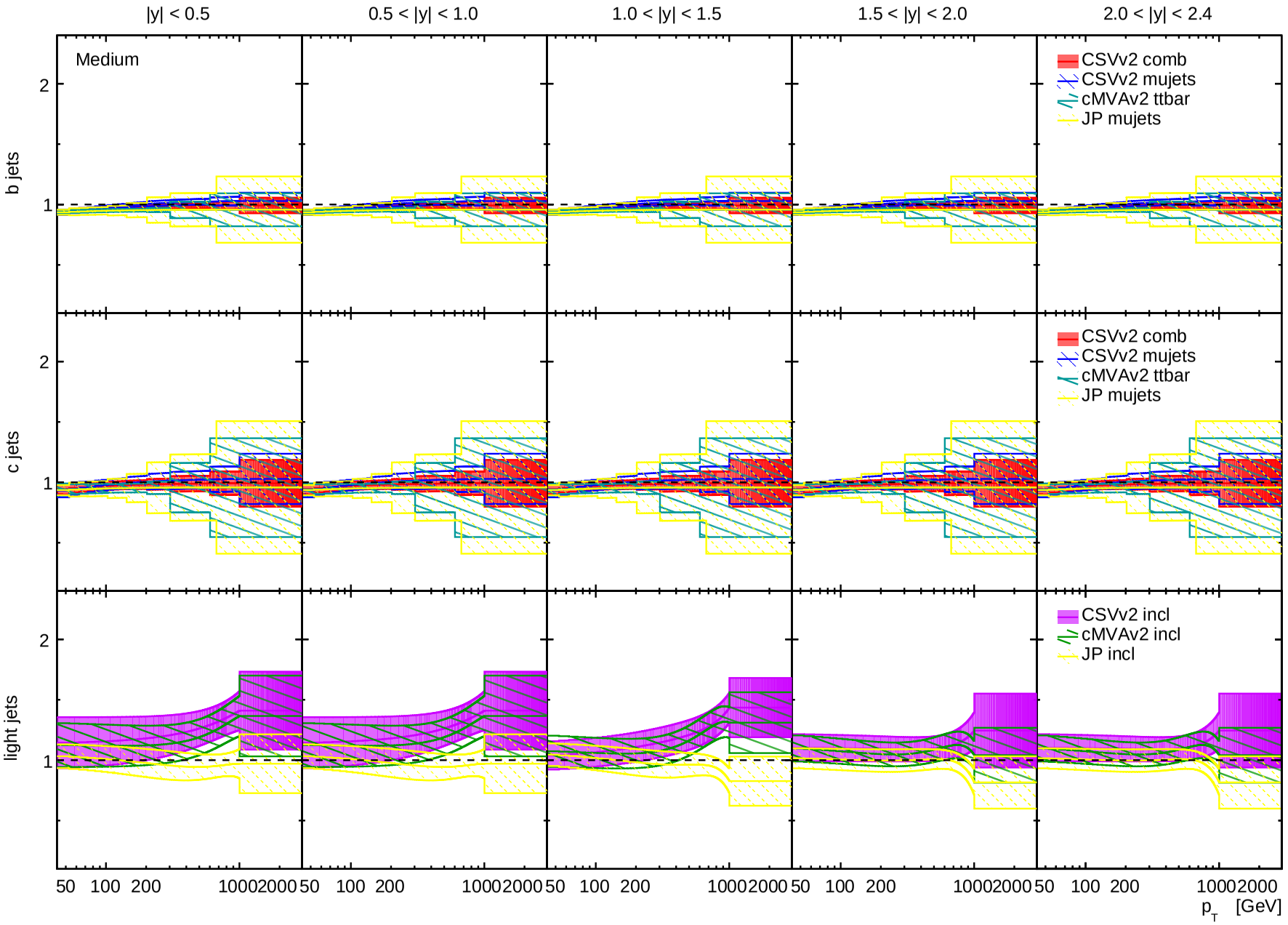


FIGURE 7.27 – SFs for the medium selection. Mistag scale factors for light jets include a rapidity dependence in 3 bins.

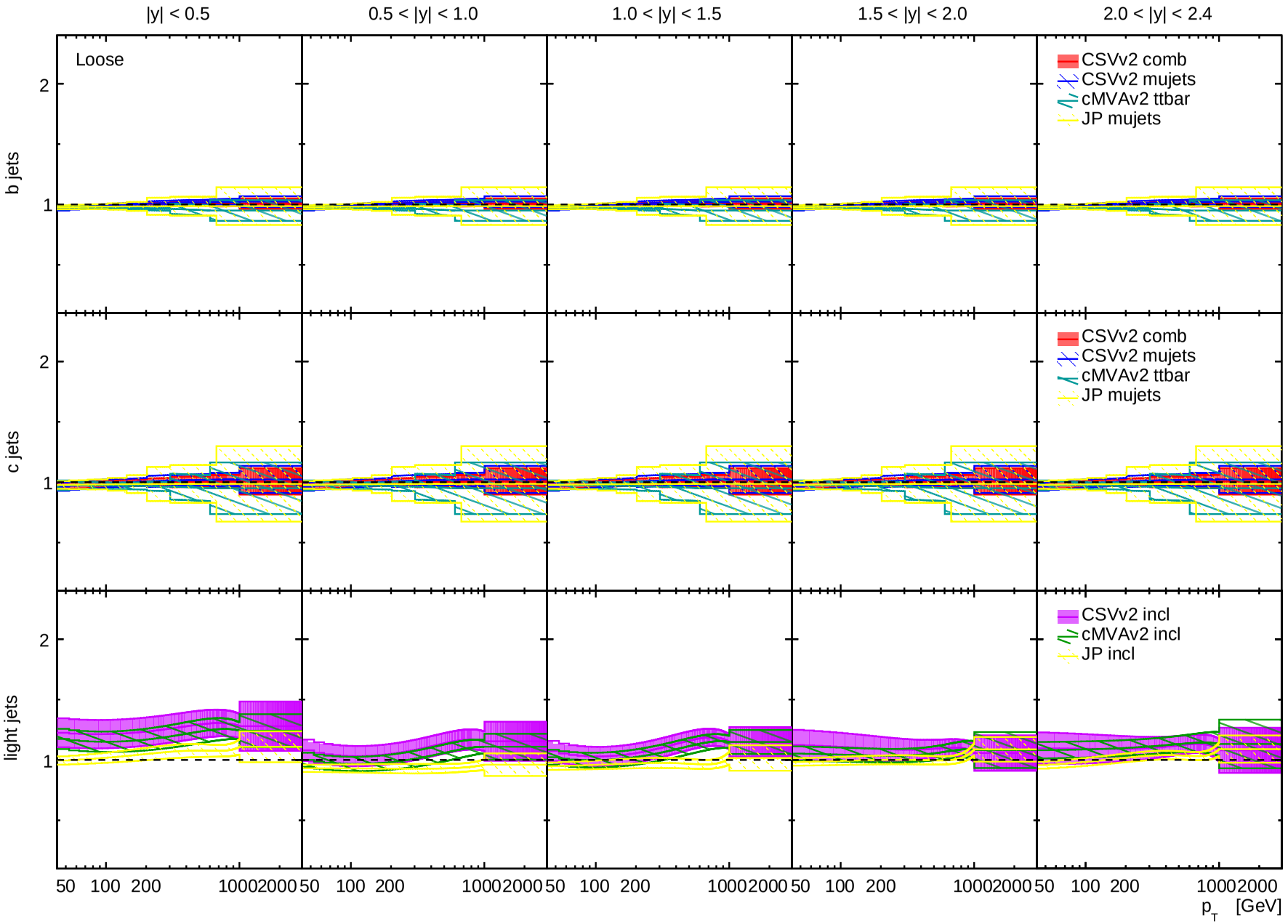
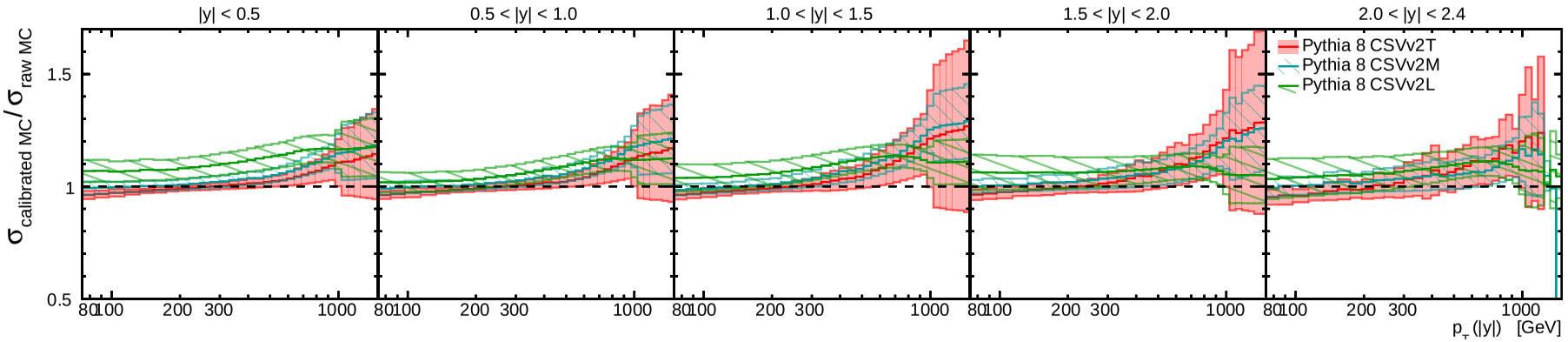
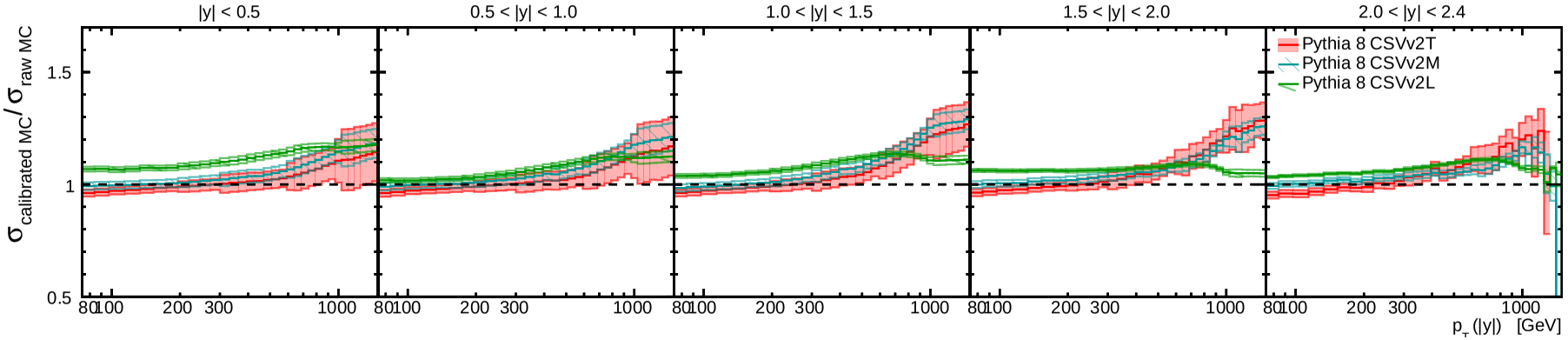


FIGURE 7.28 – SFs for the loose selection for the different flavours. Mistag scale factors for light jets include a rapidity dependence in 3 bins.

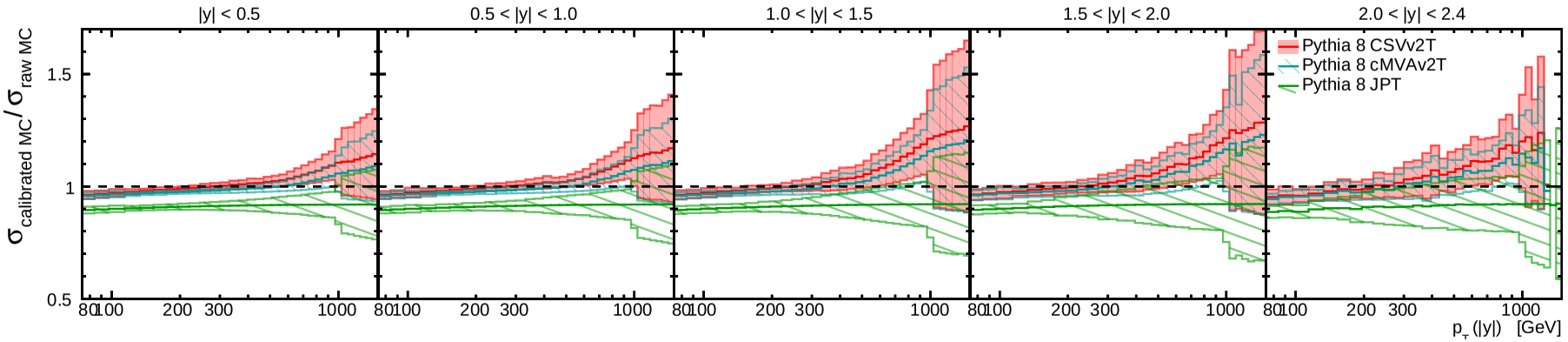


(A) Light scale factors

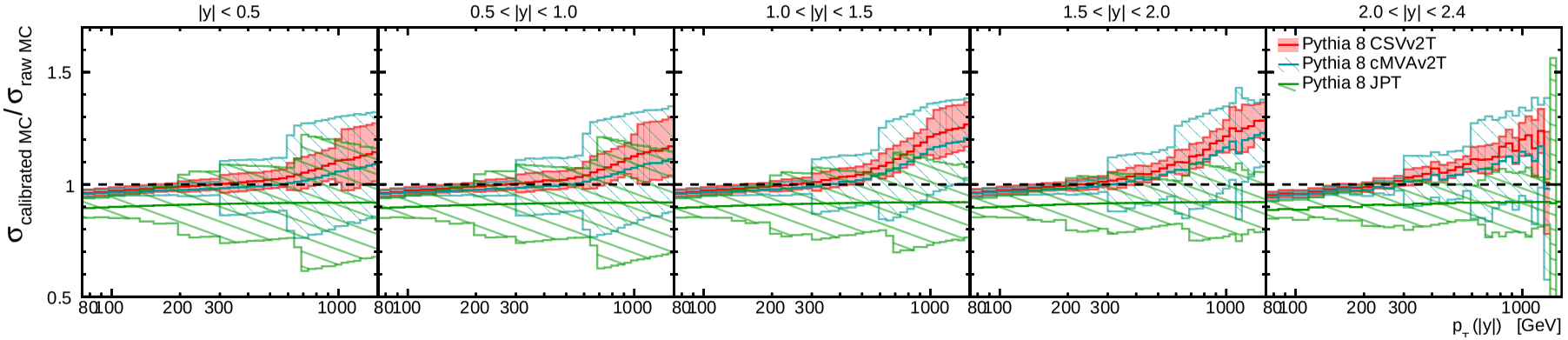


(B) Heavy-flavour scale factors

FIGURE 7.29 – Effect of the b calibration on the spectrum with different WPs. Different WPs are compared for `CSVv2`. The five columns correspond to the five rapidity bins. The coloured, shaded bands correspond to the uncertainties of the SFs.



(A) Light scale factors



(B) Heavy-flavour scale factors

FIGURE 7.30 – Effect of the b calibration on the spectrum with different WPs. Different taggers are compared for the tight WP. The five columns correspond to the five rapidity bins. The coloured, shaded bands correspond to the uncertainties of the SFs.

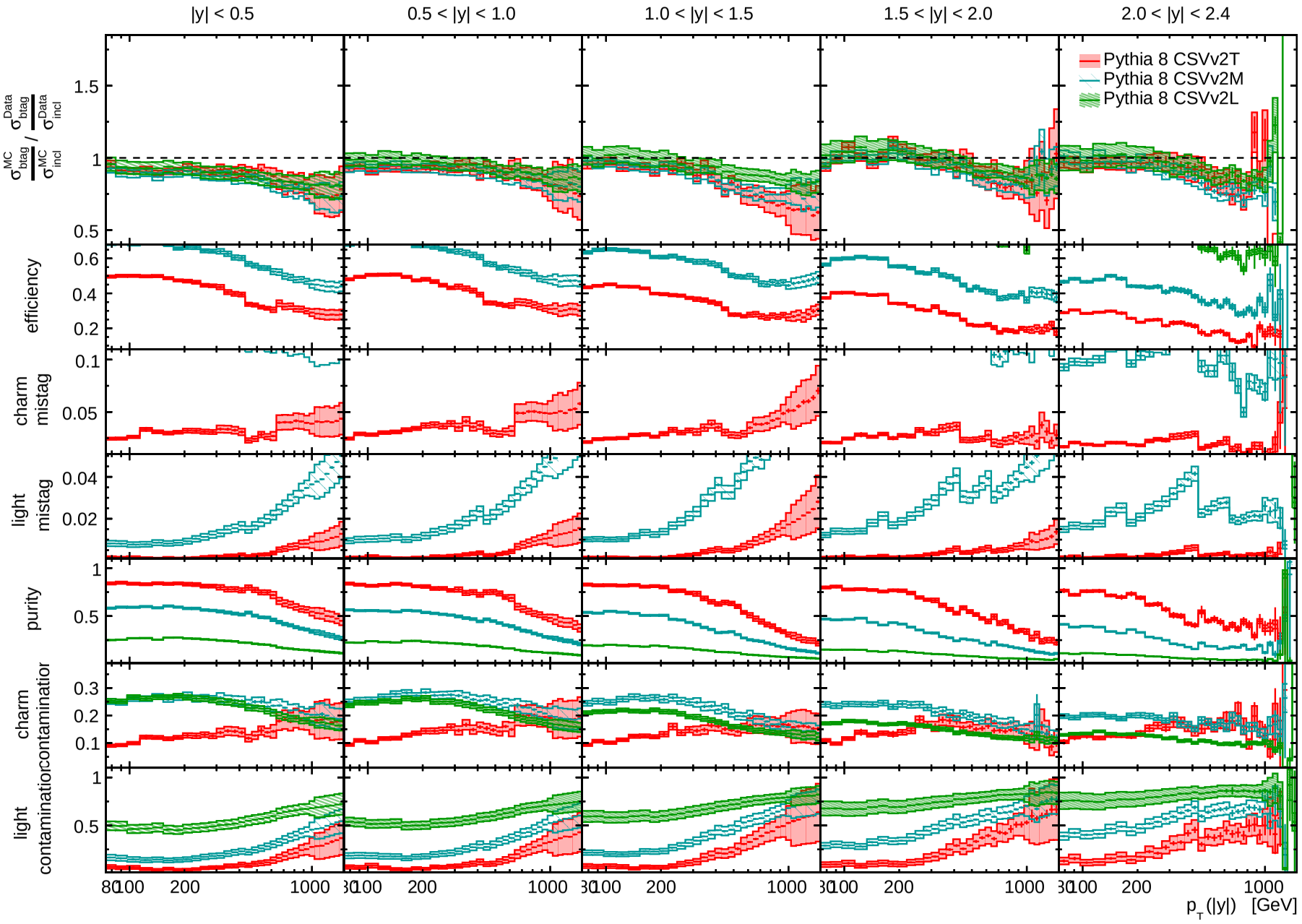


FIGURE 7.31 – Comparison of the performance of `CSVv2` of PYTHIA 8 with different WPs.

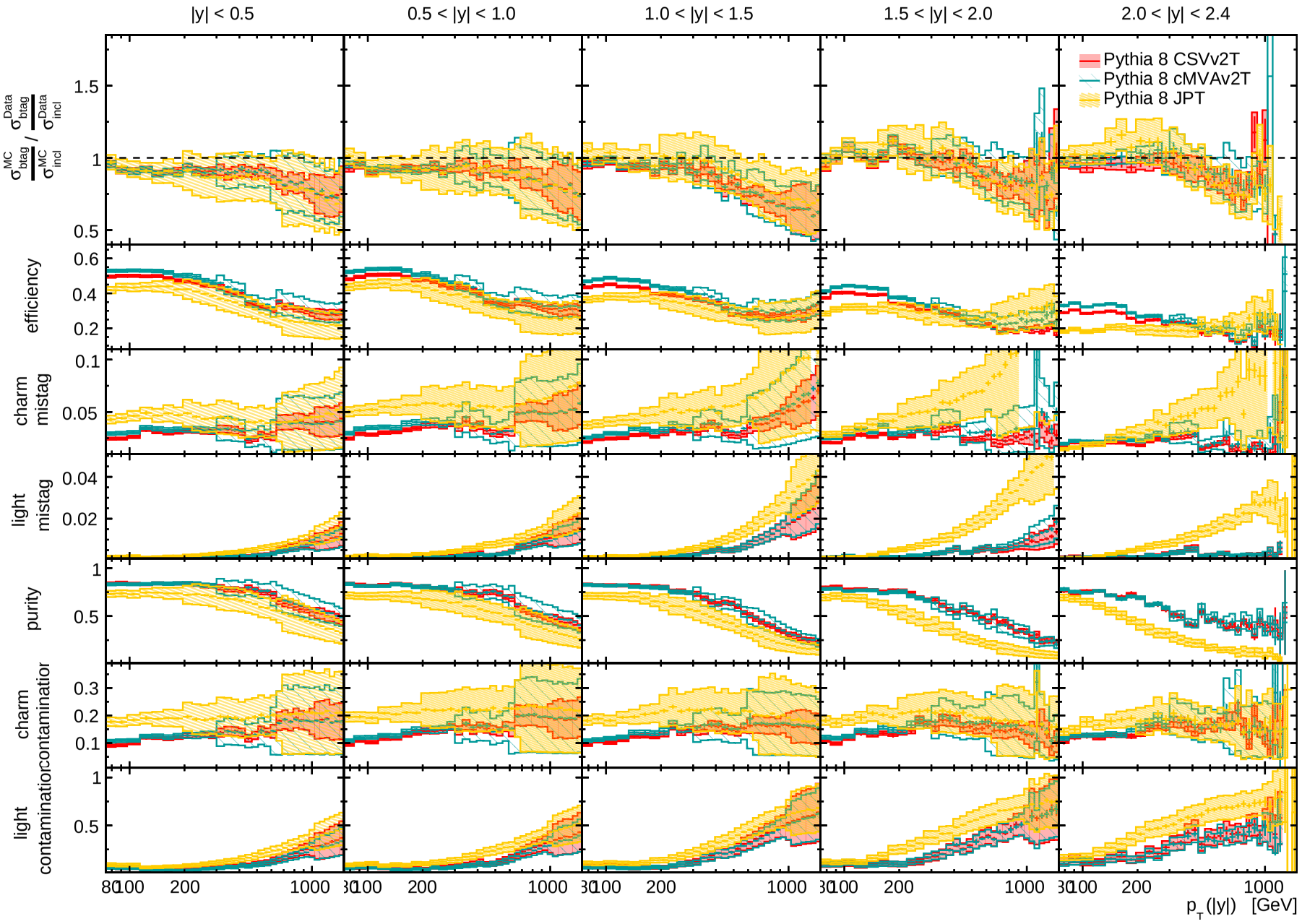
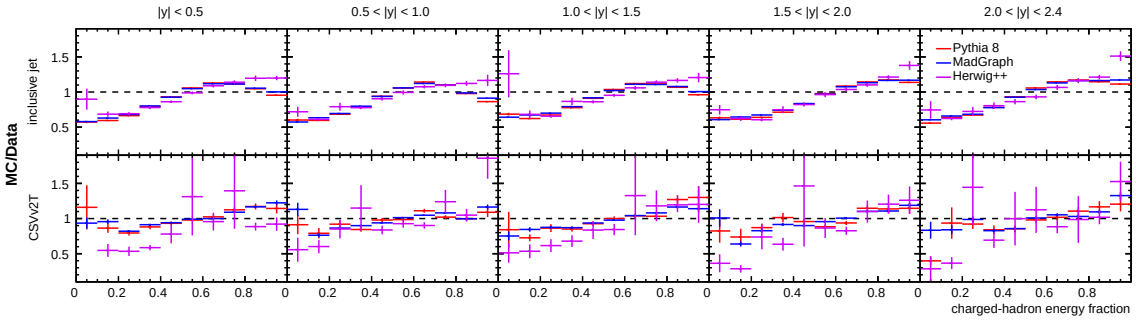
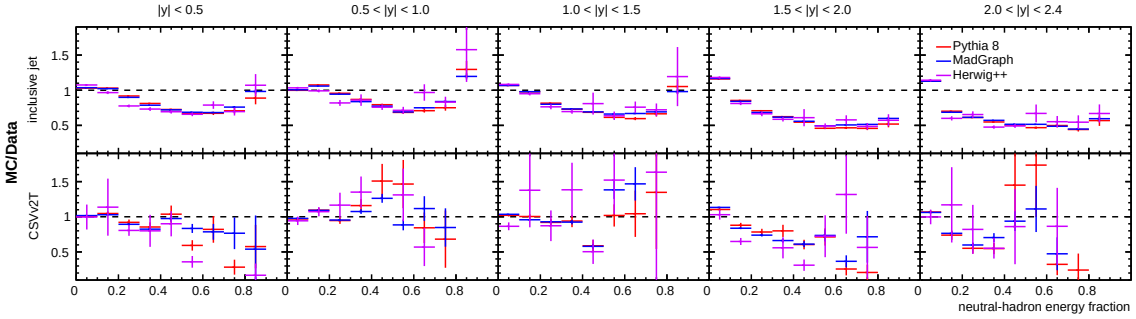


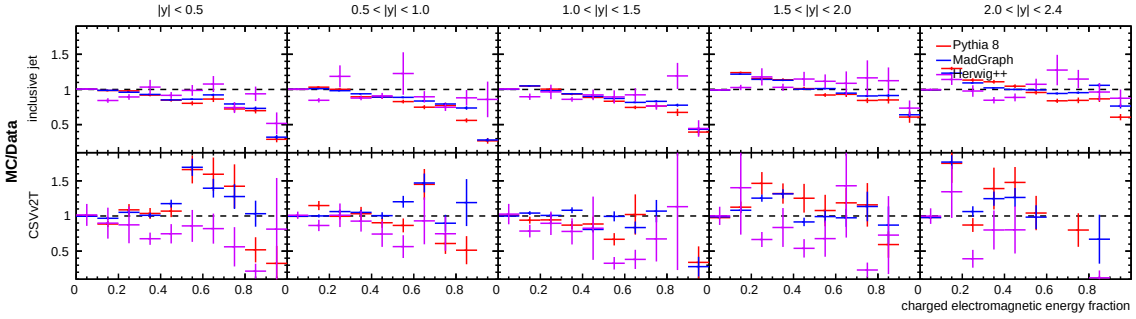
FIGURE 7.32 – Comparison of the performance of `CSVv2` of PYTHIA 8 with different taggers.



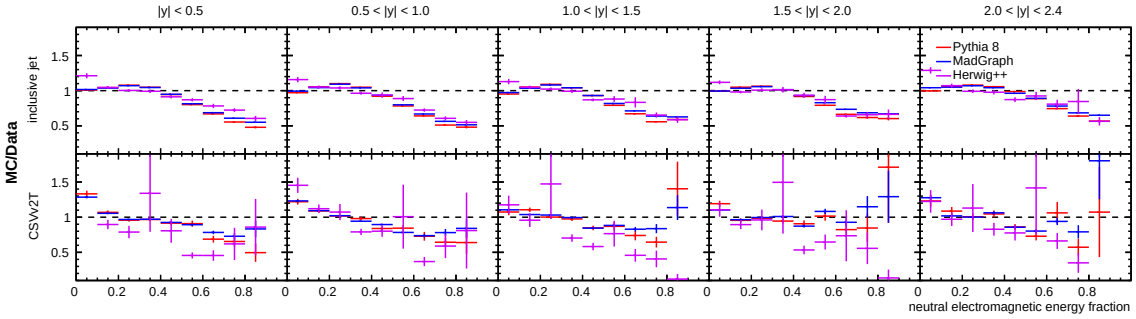
(A) Charged-hadron energy fraction.



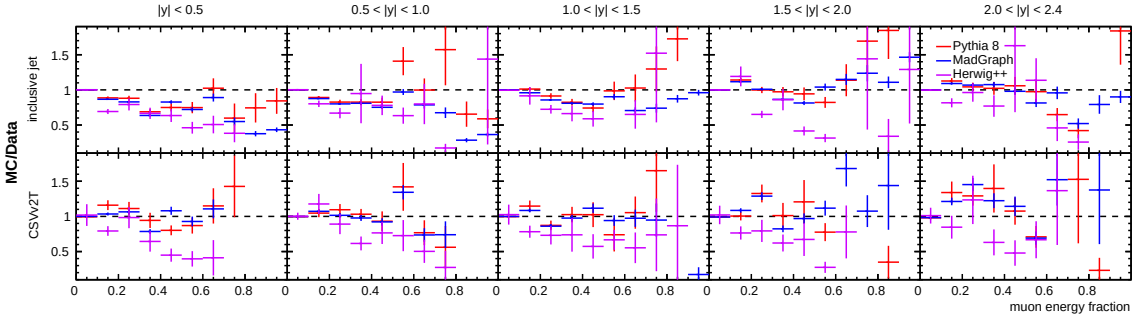
(B) Neutral-hadron energy fraction.



(C) Charged electromagnetic energy fraction.

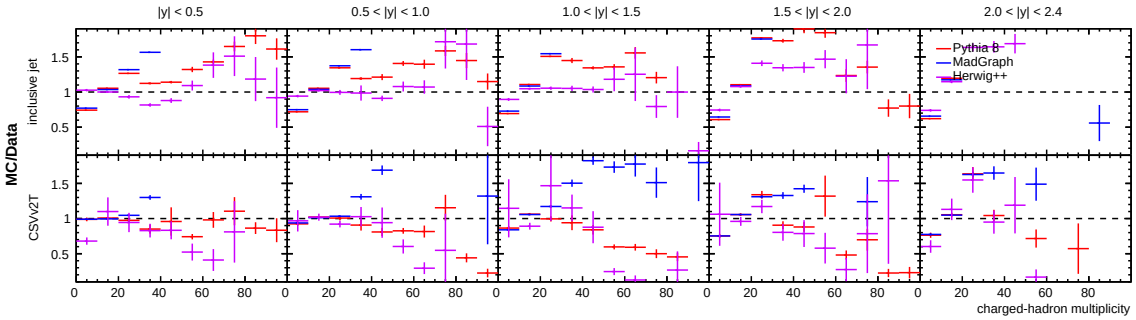


(D) Neutral electromagnetic energy fraction.

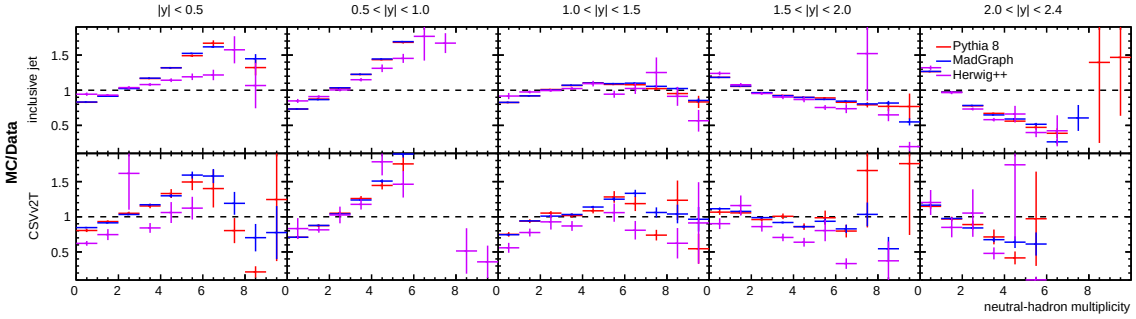


(E) Muon energy fraction.

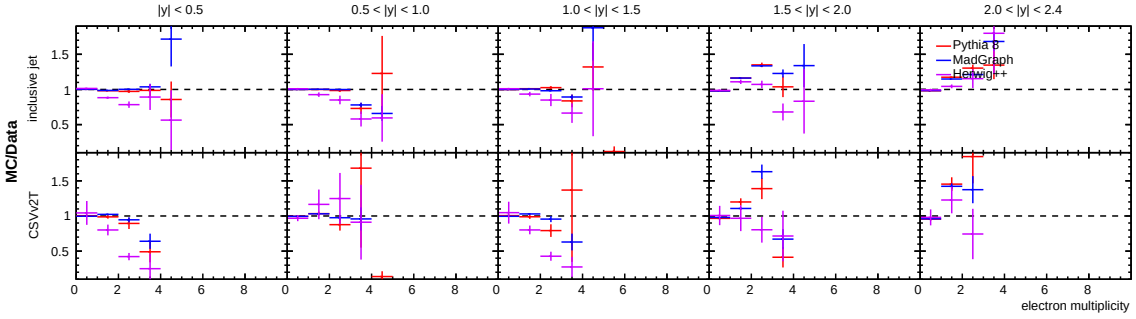
FIGURE 7.33 – Comparison of data and simulation of the fractions of the jet constituents.



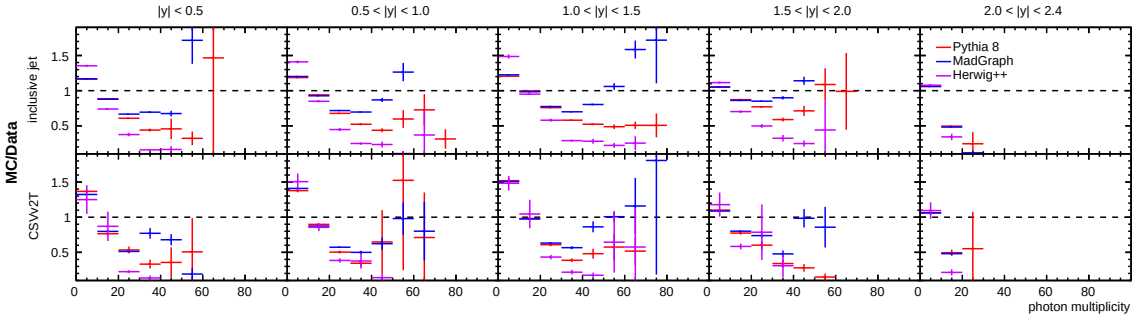
(A) Charged-hadron multiplicity.



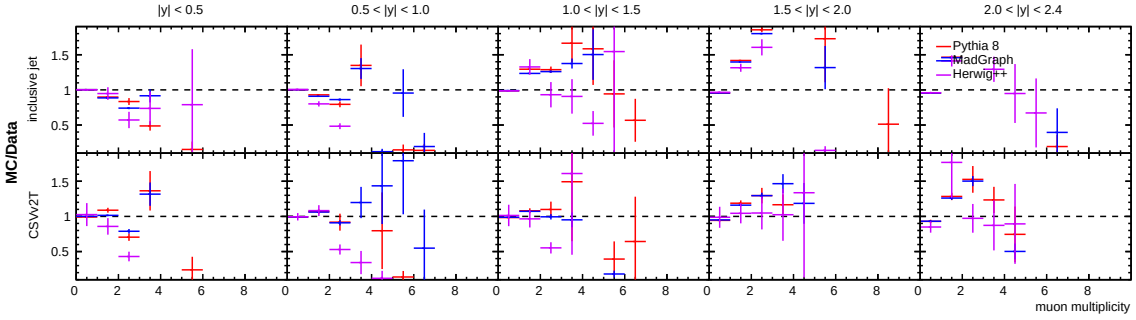
(B) Neutral-hadron multiplicity.



(C) Electron multiplicity.



(D) Photon multiplicity.



(E) Muon multiplicity.

FIGURE 7.34 – Comparison of data and simulation of the multiplicities of the jet constituents.

8

Analysis at particle level

> *"The method of Least Squares is seen to be our best course when we have thrown overboard a certain portion of our data—a sort of sacrifice which has often to be made by those who sail the stormy seas of Probability."*
>
> — Francis Ysidro Edgeworth

In the present chapter, the extraction of the particle-level cross section from the measurement is described.

The chapter is organised as follows. First, the disagreement in the b-tagged fraction observed in the previous chapter is investigated; a correction to the simulation is applied to fix the disagreement. Then, the b-jet and n-jet cross sections are simultaneously extracted together with advanced techniques of unfolding; the treatment of the systematic uncertainties in the unfolding is also discussed.

8.1 Purity

The CMS taggers were described in Sect. 7.4; the performance was given. The efficiency and mistag rates, as well as the purity and the contaminations, are two equivalent ways to describe the effect of tagging from different point of views: the former (latter) describes the content of the tagged (true) sample in terms of true (tagged) sample:

— from true to tagged:

$$\begin{bmatrix} \sigma_{\hat{n}} \\ \sigma_{\hat{b}} \end{bmatrix} = \begin{bmatrix} 1-m & 1-\epsilon \\ m & \epsilon \end{bmatrix} \times \begin{bmatrix} \sigma_n \\ \sigma_b \end{bmatrix} \tag{8.1}$$

where ϵ (m) is the *efficiency* (*mistag*);

— or *vice-versa*:

$$\begin{bmatrix} \sigma_n \\ \sigma_b \end{bmatrix} = \begin{bmatrix} P_{n\hat{n}} & 1-P_{b\hat{b}} \\ 1-P_{n\hat{n}} & P_{b\hat{b}} \end{bmatrix} \times \begin{bmatrix} \sigma_{\hat{n}} \\ \sigma_{\hat{b}} \end{bmatrix} \tag{8.2}$$

where $P_{b\hat{b}}$ ($P_{\hat{b}b}$) is the *purity* of the tagged (non-tagged) sample.

The purity and contaminations are related to the efficiency and mistag rates by matrix inversion. Therefore, the b calibration described in the previous chapter corrects the purity of the samples at detector level.

The extraction of the true cross section can in principle be performed by matrix inversion; technically, this operation from tagged to true cross section is included in the unfolding.

In Fig. 7.17, a strong disagreement the fraction of b-tagged jet in the inclusive jet sample was observed between simulation and data, despite the calibration. In this section, we discuss origin of the disagreement and show how a correction on the purity can fix it.

8.1.1 Origin of the disagreement

At high p_T, in addition to the disagreement in the fraction ratio, one can observe that the efficiency and the purity (mistag rate and contamination) are significantly low (high).

In principle, after the calibration of the b-tagging, the efficiency and mistag rates are supposed to be correct; however, at high p_T, the SFs are not derived in the inclusive-jet sample, where the fraction of b jets is small, but from sub-samples where the statistics of b and n jets are of similar order. Consequently, in the present analysis, a tiny under- or overestimation of the performance can lead to important differences between data and simulation in the tagged sample. For instance, numerically, Eq. 8.1 reads:

$$\begin{bmatrix} \sigma_{\hat{n}} \\ \sigma_{\hat{b}} \end{bmatrix} = \begin{bmatrix} 0.99 & 0.5 \\ 0.01 & 0.5 \end{bmatrix} \times \begin{bmatrix} \sigma_n \\ \sigma_b \end{bmatrix} \tag{8.3}$$

Since $\sigma_n \approx 50\sigma_b$, a variation of the order of the percent is enough to induce significant effects in the tagged cross section; if this is not correct at the level of the detector, it infers at particle level.

Moreover, the treatment of rapidity dependence in the training and in the calibration of `CSVv2` is not satisfactory: only two bins to train the tagger are considered, and no rapidity dependence is included in the calibration. The high p_T region is also treated with a very rough binning (see Fig. 7.2), which is not so fine as the one used in the current analysis.

Therefore, an additional correction has to be derived.

8.1.2 Method of the determination of the purity

In Fig. 8.1, the two first sub-figures correspond to the procedure of calibration described in the previous chapter, while the third corresponds to the additional correction that is the object of this section.

The principle consists in performing a *template fit* of the purity separately in the $\hat{b}$ and $\hat{n}$ samples, using a variable sensitive to b jets but loosely correlated with the `CSVv2` tagger. A good choice for such a variable is for instance the `JP` tagger (see correlations in Sect. 7.4 in Fig. 7.13).

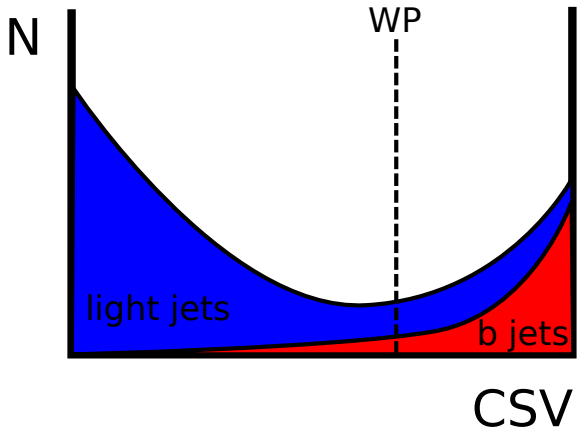
In a template fit, rather than an analytical function, one uses *templates* t_is (for instance obtained from the simulation) to define a function:

$$f(x) = \sum_i p_i t_i(x) \tag{8.4}$$

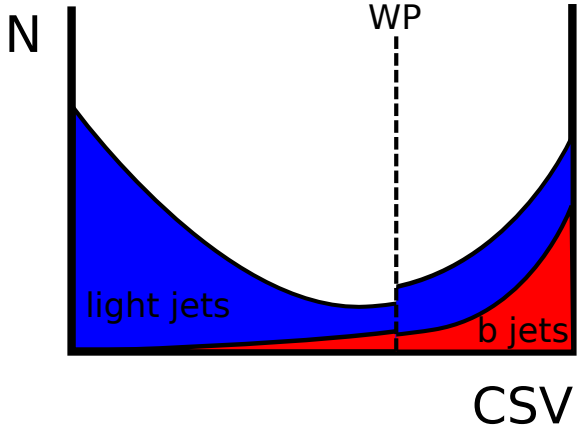
where the p_is are the fit parameters. The technique of template fits is standard when the analytical shape of a function is unknown. Here, a template fit of the discriminant is performed in the data $N_{\text{data}}^{\text{total}}$ with templates corresponding to the discriminant for the different flavours in simulation $N_{\text{MC}}^{\text{flavour}}$:

$$N_{\text{data}}^{\text{total}}(\text{JP}) = \sum_{\text{flavours}} p_{\text{flavour}} N_{\text{MC}}^{\text{flavour}}(\text{JP}) \tag{8.5}$$

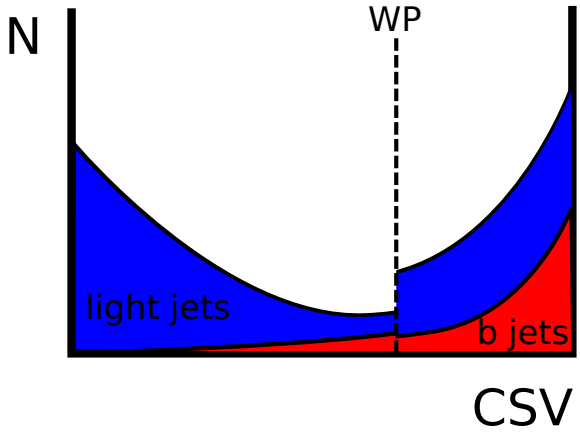
where p_{flavour} are the fit parameters. Fig. 8.2 illustrates the templates of `JP` for the different flavours in the `CSVv2`-tagged and non-`CSVv2`-tagged sub-samples. After the fit, the contributions to the `CSVv2` distribution are renormalised.



(A) Before correction.

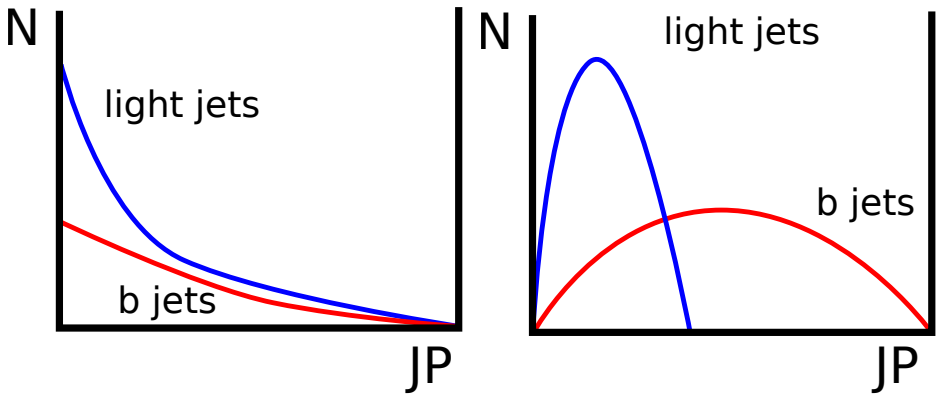


(B) After calibration.



(C) After additional correction.

FIGURE 8.1 – Sketch of the calibration. The three sub-figures represent the `CSVv2` tagger at different stages of the calibration. The blue (red) stands for the light (b) jets; c jets are not represented.



(A) Non-CSVv2-tagged region.

(B) CSVv2-tagged region.

FIGURE 8.2 – Sketch of the JP templates in the two regions. Charm is neglected in this picture.

A priori, one could simply perform a template fit of the CSVv2 tagger itself; however, using another variable results in a better fit since the power of discrimination of JP is used in addition to the one of CSVv2. Also, in the CSVv2-tagged region, the JP templates have very distinct shapes (as illustrated in Fig. 8.2) whereas the different flavours have similar shapes for CSVv2; the fit has more chances to converge with distinct shapes.

8.1.3 Results

The fits are performed in bins of transverse momentum and rapidity. The step-by-step procedure is described in details in App. 8.A.

In practice, in order to make the fit converge, we have to face two difficulties:

1. In the whole discussion till now, the special behaviour of c jets has been neglected. However, in practice, the template for the c component has a shape halfway between light and b templates. Different alternatives can be thought of: either the three templates are considered independently, or the c template is constrained together with one of the other templates (either the light or the b template). The most stable configuration happens when the b and c templates are constrained together. In order to cover this arbitrary constrain, a conservative uncertainty on the procedure is derived by varying the normalisation of the c template with a factor of 1.0 ± 0.5:

$$N_{\text{data}}^{\text{total}} = p_{\text{udsg}} N_{\text{MC}}^{\text{udsg}} + p_{\text{b+c}} \left((1.0 \pm 0.5)\, N_{\text{MC}}^{\text{c}} + N_{\text{MC}}^{\text{b}} \right) \tag{8.6}$$

2. Moreover, in the non-CSVv2-tagged region, the statistics from the light and c components is much larger than the b component's; for this reason, the template fit in this region did not converge or give satisfactory results. As an alternative, we apply the same normalisation factor to b and c jets as obtained in the CSVv2-tagged region, and simply rescale the light component [1].

The simulation is corrected to match the purity in data. Fig. 8.3 shows the performance of the CSVv2 tagger in PYTHIA 8 before and after the applying the correction. The performance of the CSVv2 tagger is improved by the procedure: one can see that the fraction ratio agrees at one and that the mistag rate is higher at high p_T after the correction than before.

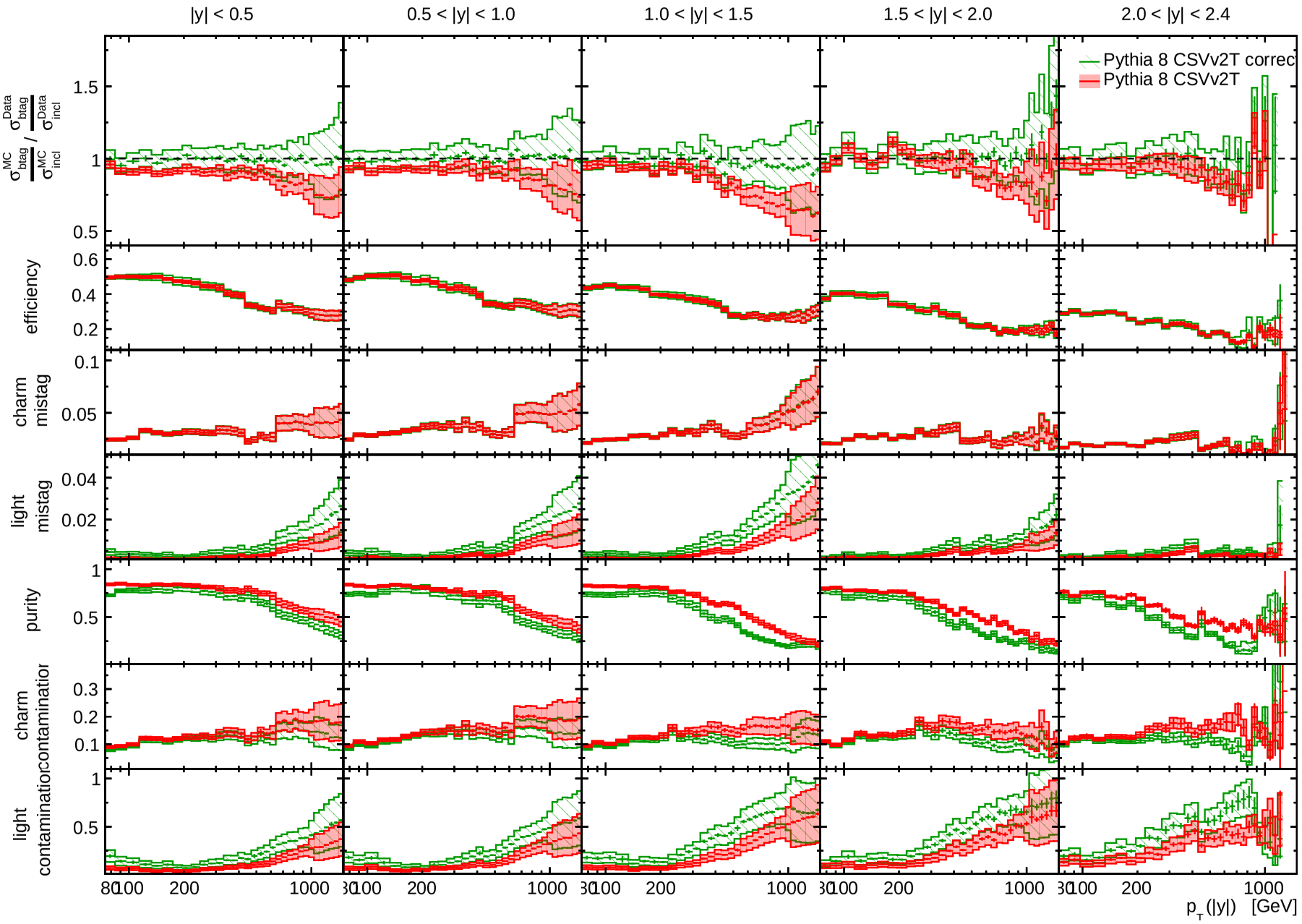


FIGURE 8.3 – Comparison of the performance of b-tagging before (red) and after (green) the additional correction. The columns correspond to the rapidity bins; the rows correspond to the ratio of the fractions in data and simulation (first row), to the efficiency and mistag rate (three next rows) and to the purity and contaminations (three last rows).

8.2 Unfolding

The aim of unfolding is to correct for different effects of the detector on the measurement of a differential distribution [2]. These effects are the limited resolution, acceptance, inhomogeneities and anisotropies of the detector. They can result in a smearing due to migrations or in reconstruction inefficiencies.

The problem of unfolding may be formulated in the following mathematical terms:

$$\mathbf{A}\mathbf{x} + \mathbf{b} = \mathbf{y} \tag{8.7}$$

where

- $\mathbf{x}$ is the *truth* vector at particle-level, which represent the distribution that we want to determine in this section;
- $\mathbf{y}$ is the *measurement* vector at detector-level, which is known from Chap. 7;
- $\mathbf{b}$ is the *background* vector at detector-level, discussed in Chap. 6;
- and $\mathbf{A}$ is the Probability Matrix (PM), which we are going to construct with the MC samples, where an element A_{ij} corresponds to the probability that a quantity with true value (or generated values for MC) in bin j will be measured (or reconstructed) in bin i.

Given $\mathbf{y}$, $\mathbf{b}$ and $\mathbf{A}$, the exact solution $\mathbf{x} = \mathbf{A}^{-1}(\mathbf{y} - \mathbf{b})$ is not guaranteed to work due to numerical instabilities, both in the measurement and in the PM. Therefore the matrix inversion is replaced with the so-called *unfolding*; here, we mainly use the unfolding formulated as a least square minimisation with Tikhonov regularisation, implemented with the `TUnfold` package (version 17.6) [3, 4].

In this section, the construction of the response of the detector and the technique of unfolding are first described. The unfolding is repeated in different conditions:

- we perform a Closure Test (CT) on pseudo-data with the simulations;
- we perform the unfolding with two different MC samples and with two different taggers;
- we cross-check the result with another unfolding technique.

The treatment of the uncertainties is explained along the different steps. Additional tests like the *backfolding* and Bottom Line Test (BLT) are presented in App. 8.B.

8.2.1 Response of the detector

In this subsection, we describe how the response of the detector is implemented in the Response Matrix (RM), and what the RM describes in general and in the particular case of the current analysis.

8.2.1.1 Response matrix and probability matrix

A RM consists in a matrix that contains the information both at particle level and detector level. A RM can be constructed for instance from a MC sample containing event records at both levels, *e.g.* PYTHIA 8 or MADGRAPH.

In the case of the unfolding of, say, the p_T spectrum, the RM is a 2-dimensional matrix, with one coordinate corresponding to each level; it is obtained from the simulation by filling a 2-dimensional histogram with pairs $(p_T^{\text{rec}}, p_T^{\text{gen}})$ corresponding to the values of the

transverse momentum of a jet before and after the simulation of the detector. The correspondence between particle level and detector level is however not obvious, since a particle-level jet may be reconstructed as two separated detector-level jets or *vice-versa*. One needs to define a *matching* between the jets of the two levels to obtain the pairs $(p_T^{\text{rec}}, p_T^{\text{gen}})$; the matching is described in one of the next paragraphs (Sect. 8.2.1.2).

As already mentioned in Sect. 6.3.4 while discussing the strategy of the analysis, we consider in the present analysis three variables simultaneously in the unfolding:

1. the transverse momentum,
2. the absolute rapidity,
3. and the flavour.

The unfolding will be 3-dimensional, and RM 6-dimensional, but the principles are the same: jets are matched between the two levels, and fill the RM with a couple of values for each variable.

The RM from PYTHIA 8 is shown in Fig. 8.4: the four large sectors correspond to the flavour components, similarly to Eq. 8.1; inside of each of the sectors are the 5×5 rapidity cells; finally, each rapidity cell consists of a simple 2-dimensional RM for the transverse momentum. Inside of a cell, the entries are concentrated around the diagonal, since jets are most likely reconstructed at the same value as at particle level; moreover, the dispersion on either side of the diagonal is slightly asymmetric, since it is more likely to miss some elements of a jet at the reconstruction than including extra activity. These two facts were already observed in Sect. 7.3; in Fig. 7.11, the mean value of the resolution profile is smaller than the bin width (which here translates in being concentrated around the diagonal) but is not at zero (which reflects the asymmetry of the dispersion). That the RM is mostly diagonal is important to perform the unfolding, as will be discussed later (Sect. 8.2.2.1).

The binning schemes of the generated and reconstructed axes differ only for the p_T component, where bins are merged by two (or more) with respect to the binning described in Sect. 7.3; merging bins is part of the regularisation, as will be defined later in this section (Sect. 8.2.2.2).

So far, we have described the construction of the RM; but in Eq. 8.7, the unfolding is performed with the PM. The PM $\mathbf{A}$ is obtained by normalising every column of the RM $\widehat{\mathbf{A}}$ to the total number of the generated events in the corresponding bin $\mathbf{x}^{MC}$ at particle level:

$$A_{ij} = \frac{1}{x_j^{MC}} \sum_i \hat{A}_{ij} \tag{8.8}$$

As some jets are not reconstructed, the sum of elements in every column of the PM is smaller than 1; this value corresponds to the efficiency of reconstruction for the given bin at particle level. We discuss in the next sections the migrations among bins and the inefficiencies of reconstruction.

8.2.1.2 Matching

To build up the RM from the MC samples, we adopt the following procedure to match jets from detector level to particle level:

1. Reconstructed jets are considered one by one from the highest to the lowest transverse momentum.

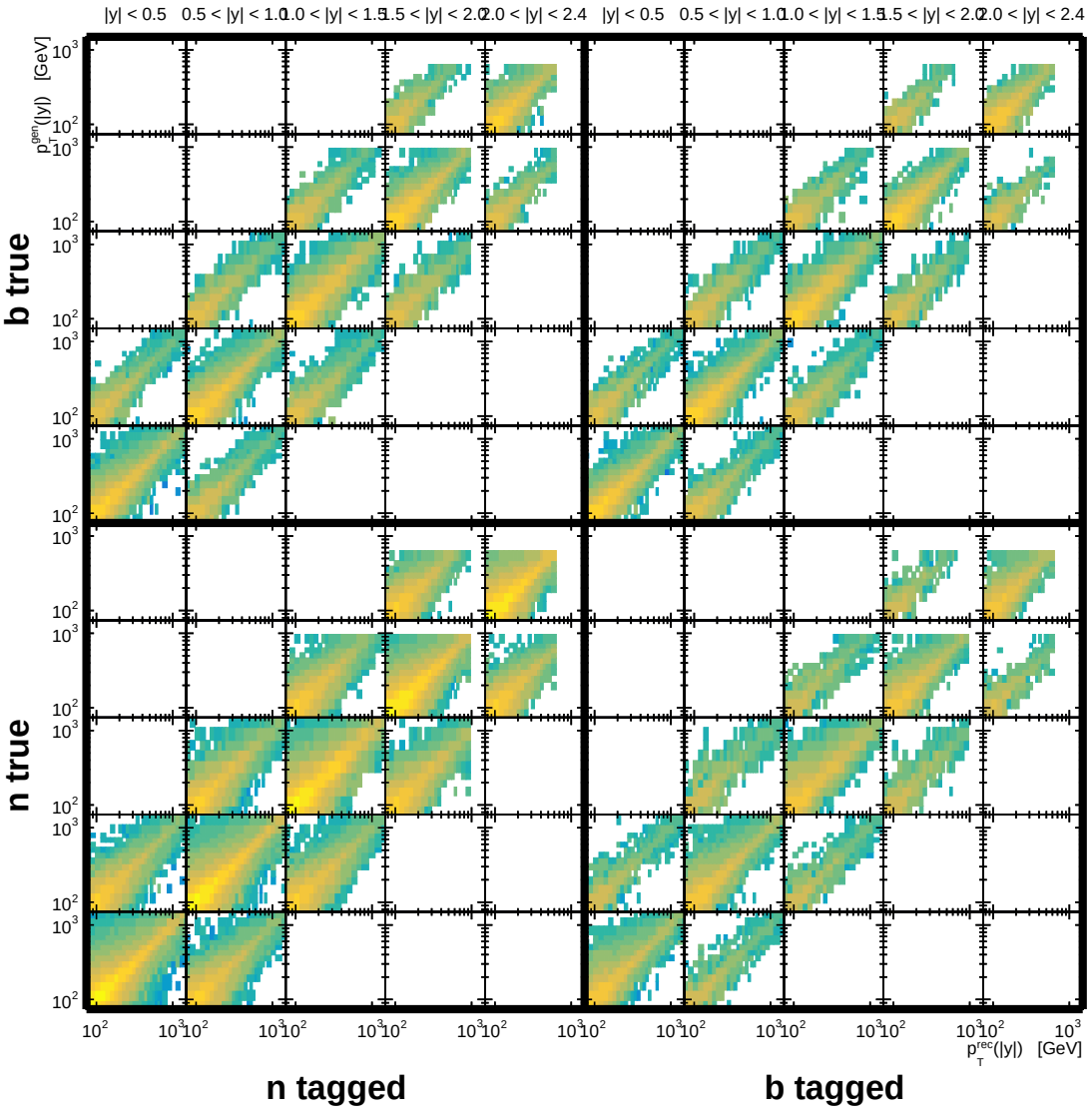


FIGURE 8.4 – Six-dimensional RM extracted from the PYTHIA 8 sample: migrations are shown for p_T (smallest cells), $|y|$ (intermediate-size cells) and the flavour (large sectors). The vertical (horizontal) axis corresponds to the generated (reconstructed). The cells contain event counts, up to $\sim 10^2$ (yellow) and down to $\sim 10^{-13}$ (blue).

2. Generated jets are considered around the reconstructed jet in a cone of radius $\Delta R = \sqrt{\phi^2 + y^2} = 0.2$. It was checked that the description of the migrations is not sensitive to variations of the radius $\Delta R \in [0.15, 0.40]$ (not shown here).

3. Inside of the cone, the generated jet of highest transverse momentum is considered for the matching, unless it was already matched previously. Another strategy is also possible, by considering the closest generated jet instead of the one of highest transverse momentum, but this has a negligible impact.

The matching is not always defined with a cone, and is sometimes only based on p_T ordering. However, this would be sensitive to additional jet activity.

By definition of the matching condition, there cannot be migrations among non-neighbouring rapidity bins; this explains the empty bins of middle size in the top left and bottom right corners of the flavour sectors of the RM in Fig. 8.4.

8.2.1.3 Miss and fake jets

The generated (reconstructed), non-matched jets are called *miss* (*fake*) jets:

— The miss jets are jets that were not matched, most often because they were reconstructed outside of the measured phase space. The miss jets are treated as yet another possible bin in the RM (although not represented in Fig. 8.4).

— The fake jets are jets that were reconstructed in addition; several reasons can hold for this: it can come from additional activity (pile-up or MPI); it can be due to a bad clustering of the jet; it can also be related to difficulties of reconstructing jets in certain regions of the detector (*e.g.* in presence of a gap or of a joint). The fake jets are considered as a background; in fact, this is the only background in the analysis, and it only affects the bins at the edge of the phase space, *i.e.* the two or three first p_T bins and the last rapidity bin.

8.2.1.4 Description of the migrations

Fig. 8.4 helps visualising the migrations globally. In complement, the composition of the bins of p_T and y in terms of migrations is shown in Fig. 8.5 for $\hat{n}$ and $\hat{b}$ jets. Except the green, which stands for jet generated and reconstructed in the same $(p_T, |y|, \text{flavour})$ bin, each colour corresponds to a type of migration.

We can describe the migrations in terms of *purity* (not in the restricted sense of the flavour):

— the non-b-tagged-jet cross section at detector level shows mainly migrations in p_T, and has a purity above 60%;

— however, the b-tagged-jet cross section suffers also from migrations due to the tagging, and the purity goes down to around 20% at $p_T = 1\,\text{TeV}$.

The gray area corresponds to the fake jets, *i.e.* the background, and shows that the first p_T bin has typically a background of 30%, rapidly decreasing to 10% in the second bin and getting negligible in bins at higher values. An uncertainty is extracted by scaling up and down the subtracted background with 10%; this is a rough but very conservative estimation, since the matching is very robust against variations of ΔR.

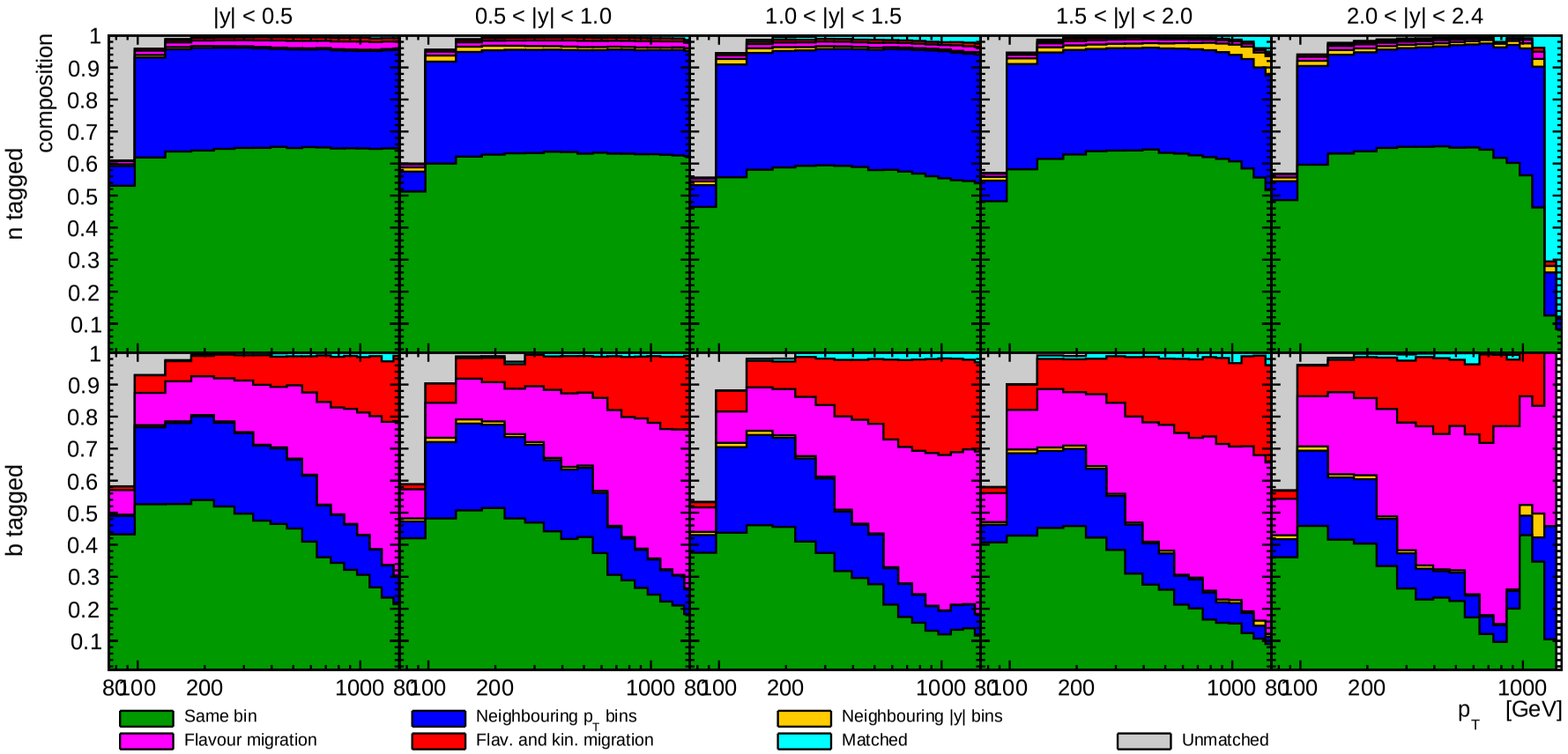


FIGURE 8.5 – Migrations of the spectrum from generator level to detector level. The upper (lower) row shows the $\hat{n}$ ($\hat{b}$) jets; the rows (columns) correspond to flavour (rapidity bin). Each colour (except green) corresponds to a different type of migration: blue (yellow, purple) corresponds to variation of p_T ($|y|$, flavour) only from neighbouring bins; red (cyan) corresponds to simultaneous p_T and $|y|$ migrations in neighbouring (non-neighbouring) bins. Finally, fake jets are shown in gray.

8.2.1.5 Uncertainties

In Chap. 7, the following systematic uncertainties have been discussed:

— pile-up reweighting (Sect. 7.2)

— pile-up removal (Sect. 7.2)

— JER smearing (Sect. 7.3)

— JES correction (Sect. 7.3)

— b tagging calibration (Sect. 7.4)

— b purity (Sect. 8.1)

In addition, we mentioned the background subtraction in Sect. 8.2.1.3.

The RM is constructed separately for each variation up and down. The systematics uncertainties are therefore inferred at particle level by repeating the unfolding for each variation of the RM.

In addition, the RM itself has statistical uncertainties that contribute to the unfolded spectrum as an additional uncertainty.

8.2.1.6 Model reweighting

A last procedure is performed to improve the construction of the RM. Since the models can have distinct distributions from the data, the unfolding can infer some significant bias in the data. The model reweighting aims at reducing the model dependence of the unfolding.

In the present case, the model reweighting is performed together with the correction of the flavour purity: the fit of the purity is performed without normalisation of the fitted discriminant; then the fit parameters are directly applied to the template, which results in corrected simulations describing the data. Since the flavour is fitted in bins of p_T and y, the spectrum in the simulation is corrected to the one in the data also in terms of kinematics.

8.2.2 Procedure

We have now determined the different components entering Eq. 8.7:

— the measurement at detector level $\mathbf{y}$ from Chap. 7,

— the background $\mathbf{b}$

— and the PM $\mathbf{A}$ from Sect. 8.2.1.

Therefore we can now describe the procedure of unfolding.

In principle, the unfolding consists in the inversion of $\mathbf{A}$ (or more generally pseudo-inversion, as the RM is not necessarily square). In practice however, the inversion of the matrix is often numerically unstable and may not lead to reasonable results, and may even deliver negative values.

In this subsection, the origin of the instabilities are explained, and the strategy to prevent instabilities from degrading the final result, the *regularisation procedure*, is described.

8.2.2.1 Origin of the instabilities

The origin of fluctuations can be understood in terms of Singular Value Decomposition (SVD) [5]:

$$\mathbf{A} = \mathbf{USV}^{\intercal} \tag{8.9}$$

where $\mathbf{U}$ and $\mathbf{V}$ are orthogonal matrices transforming $\mathbf{A}$ into $\mathbf{S}$, a square matrix with the eigenvalues s_i of $\mathbf{S}$ on the diagonal. The inverse of $\mathbf{S}$ depends on the inverse of the determinant of the matrix $|\mathbf{S}| = \Pi s_i$; if the eigenvalues differ in order of magnitude, the inversion of the PM is dominated by small eigenvalues. The *regularisation* consists of limiting the contributions from the small eigenvalues.

A relevant quantity quantifying whether the matrix inversion is possible (or in other words, how much regularisation is necessary) is the *condition* of the matrix, which is the ratio of the highest eigenvalue with the lowest eigenvalue: if they are of different orders of magnitude, the *condition* of the matrix should be very different from 1. In this analysis, the *condition* of the PM from PYTHIA 8 is 14.5909, therefore no large impact of the regularisation is expected.

8.2.2.2 Tikhonov regularisation

The regularisation chosen for the present analysis is the *Tikhonov regularisation* [5].

The problem of unfolding is reformulated as a problem of least-square minimisation:

$$\chi^2 = (\mathbf{y} - \mathbf{Ax})^T \mathbf{V}_y^{-1} (\mathbf{y} - \mathbf{Ax}) \tag{8.10}$$

where $\mathbf{V}_y$ is the covariance matrix from the measurement (shown in App. 8.B). As the inclusive b jet measurement is a multi-count observable, off-diagonal elements are different from zero.

Minimising χ^2 naturally leads to a good estimate of the invert of the PM $\mathbf{A}$ within the available statistics.

Two types of regularisation are applied:

— The number of free parameters to fit is reduced; in other words, a finer binning is used for the measurement than for the final result at particle level. It should first be noted that the perfect solution, $\mathbf{x} = \mathbf{A}^{-1}\mathbf{y}$, using a square RM, leads to $\chi^2 = 0$. On the other hand, the opposite case of a single-row matrix does not necessarily lead to $\chi^2 = 1$, which means that, ideally, the number of degrees of freedom should be tuned to perform the unfolding. However, no attempt has been done to find an optimal merging scheme; the recommendation in the `TUnfold` package is to perform the measurement with twice more bins than at particle level has been followed [6].

— A second term is added in the χ^2 in Eq. 8.10, the *regularisation term*:

$$\chi^2 = \underbrace{(\mathbf{y} - \mathbf{Ax})^T \mathbf{V}_y^{-1} (\mathbf{y} - \mathbf{Ax})}_{\chi^2_{\mathbf{A}}} + \tau^2 \underbrace{||\mathbf{Lx}||^2}_{\chi^2_{\mathbf{L}}} \tag{8.11}$$

where $\mathbf{L}$ is the *regularisation matrix* and τ the *regularisation parameter*. In general, this additional term aims at constraining close bins to one another. The form of the regularisation matrix is specific to each unfolding problem and its choice will be detailed in the next section. The regularisation parameter balances the contribution of the two terms to the final result; the choice of the regularisation has to be performed according to objective criteria and will be detailed later in the second next section.

In the present case, only migrations related to the transverse momentum may need to be regularised:

— The rapidity spectrum is rather flat, and the resolution is finer than the rapidity binning; no regularisation needs to be applied in that case.

— The flavour determination, as discussed in the previous section, is only performed in two bins; no regularisation can even be applied.

— In contrast, the transverse-momentum spectrum is steeply falling, and the resolution is of the order of the bin sizes; therefore, significant migrations cause the smearing of the spectrum.

8.2.2.3 Regularisation matrix

Different prescriptions exist to build the regularisation matrix (or **L** matrix) [3]: here, we choose the prescription consisting in minimising the second derivative of the cross section with respect to p_T, *i.e.* in relating any three adjacent bins in p_T so as to prevent fluctuations from getting significant; this is called the *curvature regularisation*. A default regularisation mode is available in `TUnfold`; however, in the present case, we will see that it is not convenient and that we need to modify it.

In mathematical terms, the curvature regularisation means that, given any three consecutive bins j_1, j_2, j_3, we minimise the following quantity $(x_{j_3} - x_{j_2}) - (x_{j_2} - x_{j_1}) = x_{j_3} - 2x_{j_2} + x_{j_1}$ where the x_is are the true values (the unknowns); this quantity can be seen as the numerical second derivative [1]. The **L** matrix is then:

$$\mathbf{L} = \begin{bmatrix} 0 & \dots & 0 & 0 & 1 & -2 & 1 \\ 0 & \dots & 0 & 1 & -2 & 1 & 0 \\ 0 & \dots & 1 & -2 & 1 & 0 & 0 \\ & & \vdots & \vdots & \vdots & \vdots & \vdots \\ & & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{8.12}$$

Note that as one constraint binds three consecutive p_T bins, there are two less constraints (rows) than there are p_T bins (columns).

Because of the steeply falling character of the cross section with respect to p_T, Eq. 8.12 needs to be modified, *i.e.* in order for the high-p_T bins to be numerically significant in the minimisation, an additional factor may be added to build up the matrix so that all terms in χ^2_L are of similar order of magnitude. In terms of formula, this means that we want to introduce weights ms in the second derivative $m^j_a(x_{j_3} - x_{j_2}) - m^j_b(x_{j_2} - x_{j_1})$; the **L** matrix in Eq. 8.12 becomes:

$$\mathbf{L} = \begin{bmatrix} 0 & \dots & 0 & 0 & m^j_a & -m^j_a - m^j_b & m^j_b \\ 0 & \dots & 0 & m^k_a & -m^k_a - m^k_b & m^k_b & 0 \\ 0 & \dots & m^l_a & -m^l_a - m^l_b & m^l_b & 0 & 0 \\ & & \vdots & \vdots & \vdots & \vdots & \vdots \\ & & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{8.13}$$

1. If x_j is the event count in bin j, the first derivative is seen as $x'_j \approx \frac{x_j - x_{j-1}}{j-(j-1)} = (x_j - x_{j-1})$ (one can choose to define it on the left with $(j+1, j)$ or on the right with $(j, j-1)$); the second derivative is consequently seen as $x''_j \approx \frac{x'_j - x'_{j-1}}{j-(j-1)} = x_{j+1} - 2x_j + x_{j-1}$ (taking the first derivative once on the left and once on the right).

For the current analysis, we simply use the MC prediction at particle level $\mathbf{x}^{\mathrm{MC}}$, which means that $m_a^j = \frac{1}{x_{j_3}^{\mathrm{MC}}}$ and $m_b^j = \frac{1}{x_{j_1}^{\mathrm{MC}}}$. The matrix then looks as follows:

$$\mathbf{L} = \begin{bmatrix} 0 & \dots & 0 & 0 & \frac{1}{x_{j_3}^{\mathrm{MC}}} & -\frac{1}{x_{j_3}^{\mathrm{MC}}} - \frac{1}{x_{j_1}^{\mathrm{MC}}} & \frac{1}{x_{j_1}^{\mathrm{MC}}} \\ 0 & \dots & 0 & \frac{1}{x_{k_3}^{\mathrm{MC}}} & \frac{1}{x_{k_3}^{\mathrm{MC}}} - \frac{1}{x_{k_1}^{\mathrm{MC}}} & \frac{1}{x_{k_1}^{\mathrm{MC}}} & 0 \\ 0 & \dots & \frac{1}{x_{l_3}^{\mathrm{MC}}} & \frac{1}{x_{l_3}^{\mathrm{MC}}} - \frac{1}{x_{l_1}^{\mathrm{MC}}} & \frac{1}{x_{l_1}^{\mathrm{MC}}} & 0 & 0 \\ & & \vdots & \vdots & \vdots & \vdots & \vdots \\ & & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{8.14}$$

8.2.2.4 Regularisation parameter

The `TUnfold` package offers several methods to determine the regularisation parameter τ [3]:

— the L-curve scan,

— the minimisation of the global correlation coefficients.

Here, we use the L-curve scan method: the so-called L-curve is defined by the pairs (L_x, L_y), depending on the parameter τ:

$$L_x = \log_{10} \chi^2_{\mathbf{A}} \tag{8.15}$$

$$L_y = \log_{10} \chi^2_{\mathbf{L}} \tag{8.16}$$

where $\chi^2_{\mathbf{A}}$ ($\chi^2_{\mathbf{L}}$) corresponds to the first (second) term of the solution in Eq. 8.11. The L-curve has a L-shape (hence its name), and the value of τ is then chosen at the point of highest curvature ("in the L"). This choice gives a good compromise in the minimisation of the two terms, where χ^2_A does not increase (which means that the agreement of $\mathbf{Ax}$ and $\mathbf{y}$ is not degraded) while χ^2_L is minimised (which means that the fluctuations are limited).

The L-curve is shown in Sect. 8.2.2.6, while performing the unfolding in data.

An uncertainty on the regularisation is then obtained by varying the regularisation parameter τ (but the regularisation matrix $\mathbf{L}$ is kept identical).

8.2.2.5 Closure test

The Closure Test (CT) is performed to validate the procedure of unfolding and the MC samples. It consists in unfolding pseudo-data, *i.e.* a simulation where the truth is already known. Then, the unfolded and true spectra can be compared.

The CT can be seen in Fig. 8.6; in this figure, three levels of CTs are presented:

— First, we unfold PYTHIA 8 with itself (in red in the figure); in that case, the unfolded spectrum and the generated curve are identical and the agreement is therefore at one.

— Then, we split the PYTHIA 8 into two statistically orthogonal sub-samples; one is used as pseudo-data and the other as MC to construct the RM (yellow). This way, statistical correlations are avoided, but the pseudo-data is *per se* perfectly modelled. The agreement is not exactly at one but around is for all values of p_T.

— Finally, we unfold MADGRAPH with PYTHIA 8 (blue), where, despite the model reweighting, remaining differences are still included. In this case, one observe more fluctuations of the order of a few percents, and small deviations at high p_T.

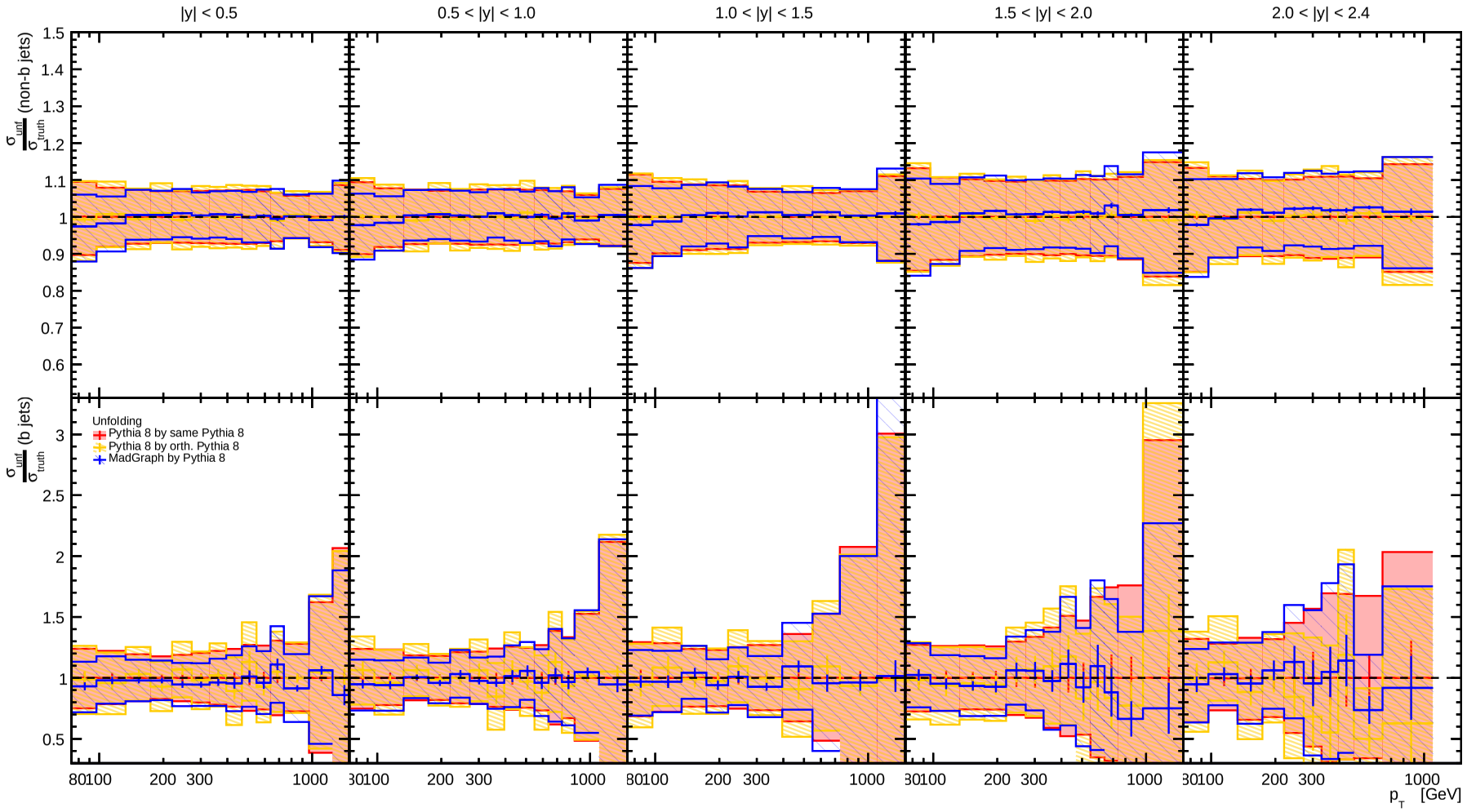


FIGURE 8.6 – The three CTs are shown. The rows (columns) correspond to the flavour (rapidity). The ratio is taken between the unfolded curve and the generated curve. The bands correspond to the statistical and systematic uncertainties.

8.2.2.6 Results in data

We consider the unfolding with the three scenarios in parallel, in order to cross-check the choice of the MC sample and of the tagger. The three scenarios correspond to the following:

— PYTHIA 8 using `CSVv2` (in red in Figs.8.7-8.10)

— PYTHIA 8 using `cMVAv2` (in orange)

— MADGRAPH using `CSVv2` (in blue)

We perform the unfolding with `TUnfold`, including all uncertainties described in Sect. 8.2.1.5 and the variation of the regularisation parameter.

L-curve scan. The L-curve scan is performed by minimising Eq. 8.11 200 times consecutively for different values of the parameter τ in the range $[1 \times 10^{-4}, 3 \times 3 \cdot 10^{-1}]$. The L-curve scan is presented in Fig. 8.7 in either scenario: in the all cases, the regularisation is weak, and small values of τ are taken (therefore the L-shape of the curve may be difficultly to recognise). In addition, the two contributions to the total χ^2 are shown as a function of τ: one can then see that τ is taken before the χ^2_A increases.

Cross section. The ratio of the double differential cross sections for b jets and n jets of data with the three simulations is shown in Fig. 8.8 (8.9) before (after) unfolding. The two figures are shown opposite and are similarly organised, with the flavour (rapidity) in the rows (columns). Before unfolding, since the simulations have been reweighted to describe the data (see Sect. 8.2.1.6), we see that the simulation agrees indeed well with the data, except for low and large p_T values for the $\hat{n}$ bins in MADGRAPH. After unfolding, in Fig. 8.9, one can observe the same disagreement; it should be noted that each simulation is considered separately: the ratio after unfolding with a given MC is taken with the same MC.

Comparison of the scenarios. The comparison of the unfolded curves in the three different scenarios is shown in Fig. 8.10, where PYTHIA 8 is taken as reference. The figure is organised in the same way as the figure for the cross sections, with the rows (columns) corresponding to the flavour (rapidity). The disagreement of MADGRAPH with data in the $\hat{n}$ cross section at detector level does not play a rôle in the b jet cross section; however, PYTHIA 8 with `cMVAv2` exhibits differences at high p_T up to 20% in the third rapidity bin, though still within the uncertainties.

Uncertainties. Finally, the composition of the total uncertainty is given Fig. 8.11 in the case of PYTHIA 8 using `CSVv2`; the uncertainties in each (p_T, y, flavour) bin are rescaled as follows:

$$1 = \frac{\delta^2_{\text{JEC}}}{\delta^2_{\text{TOT}}} + \frac{\delta^2_{\text{JER}}}{\delta^2_{\text{TOT}}} + \dots \tag{8.17}$$

All uncertainties have been considered except the luminosity uncertainty (which will be added only in Chap. 9). The figure is organised in the same way as the figure for the cross sections. For n jets (b jets), the dominant uncertainty is the JEC (b calibration) in white (dark and light grey).

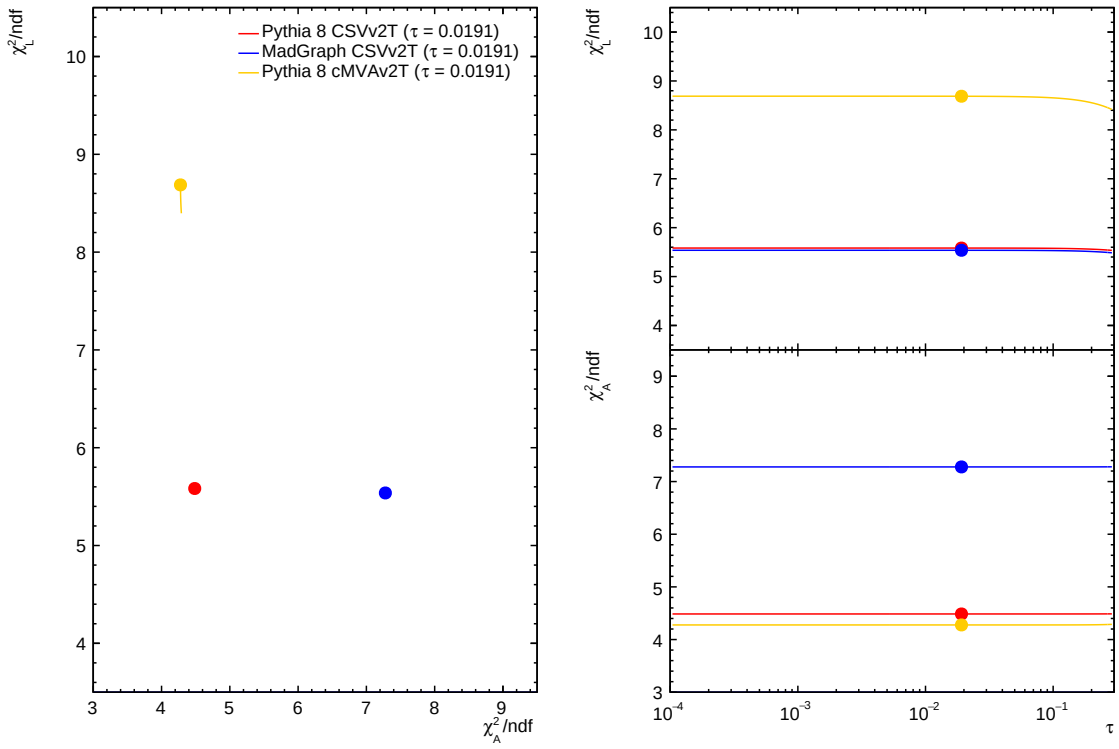


FIGURE 8.7 – The L-curves are represented on the left hand side for the different scenarios. The χ^2_A and χ^2_L are plotted as a function of τ on the right hand side. The points represent the choice of τ.

8.2.3 D'Agostini unfolding

In principle, the unfolded results should be independent from the unfolding algorithm. In this subsection, we present an alternative method: the iterative method from D'AGOSTINI [7, 8] [2] as implemented in the `RooUnfold` package [11].

8.2.3.1 Principles

The unfolding is formulated in terms of probabilities:

$$\hat{n}(C_i) = n(E)\mathbb{P}\left[C_i|E\right] \tag{8.18}$$

where $\hat{n}$ is the unknown distribution at particle level (C stands for *causes*) and n is the measured distribution (E stands for *effects*). The method is based on *Bayes theorem on conditional probabilities*:

$$\mathbb{P}\left[C_i|E_j\right] = \frac{\mathbb{P}\left[E_j|C_i\right]\mathbb{P}\left[C_i\right]}{\sum_{l=1}^{n_C}\mathbb{P}\left[E_j|C_l\right]\mathbb{P}\left[C_l\right]} \tag{8.19}$$

If $\mathbb{P}\left[C_i\right]$ is replaced by a MC prior (in the current case, the $p_T(y)$ spectrum at particle level in PYTHIA 8 or MADGRAPH can be used), one can get an estimation of the true distribution. This approach of unfolding is called *Bayesian unfolding*. However, it is very biased

2. This reference is the standard reference given in HEP. However, the same technique has already been published in other fields of science [9, 10]. In astronomy and optics for instance, it is known as *Lucy-Richardson deconvolution*.

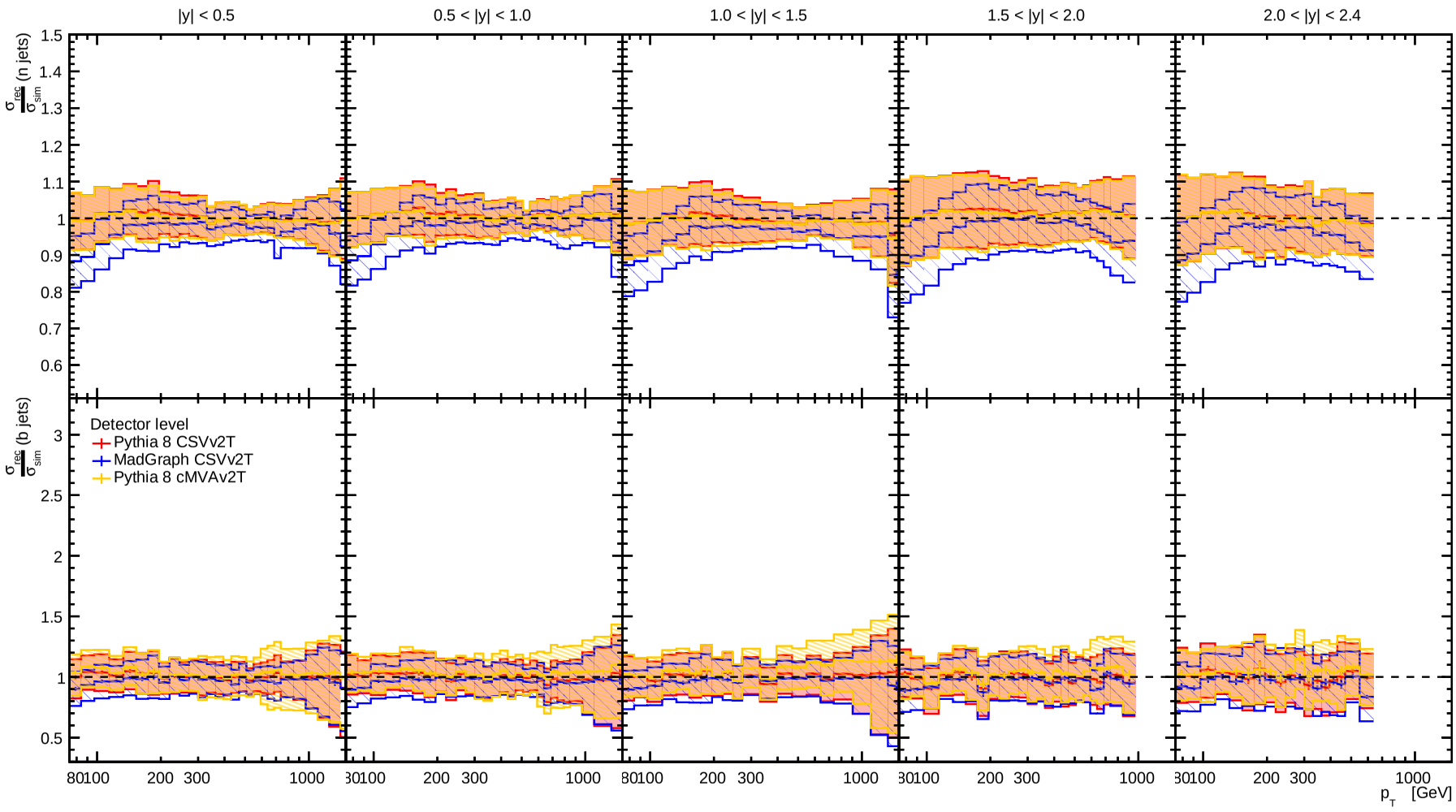


FIGURE 8.8 – Ratio of double differential cross section in transverse momentum and rapidity of the data with the simulations at detector level, before unfolding. The rows (columns) correspond to the flavour (rapidity). The bands corresponds to the total uncertainty.

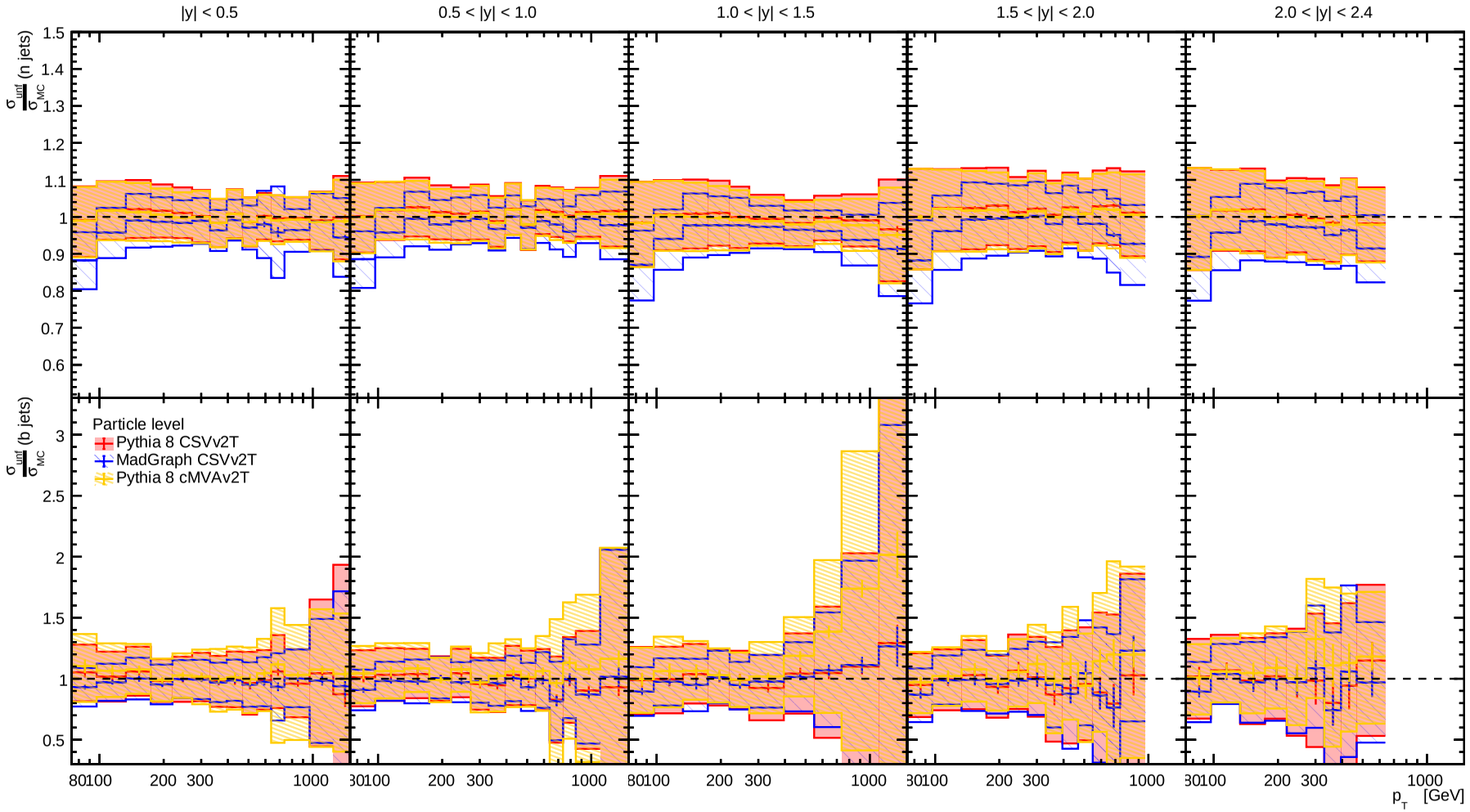


FIGURE 8.9 – Ratio of double differential cross section in transverse momentum and rapidity of the data with the simulation at particle level, before unfolding (the unfolding is performed with the same simulation, respectively). The rows (columns) correspond to the flavour (rapidity). The bands corresponds to the total uncertainty.

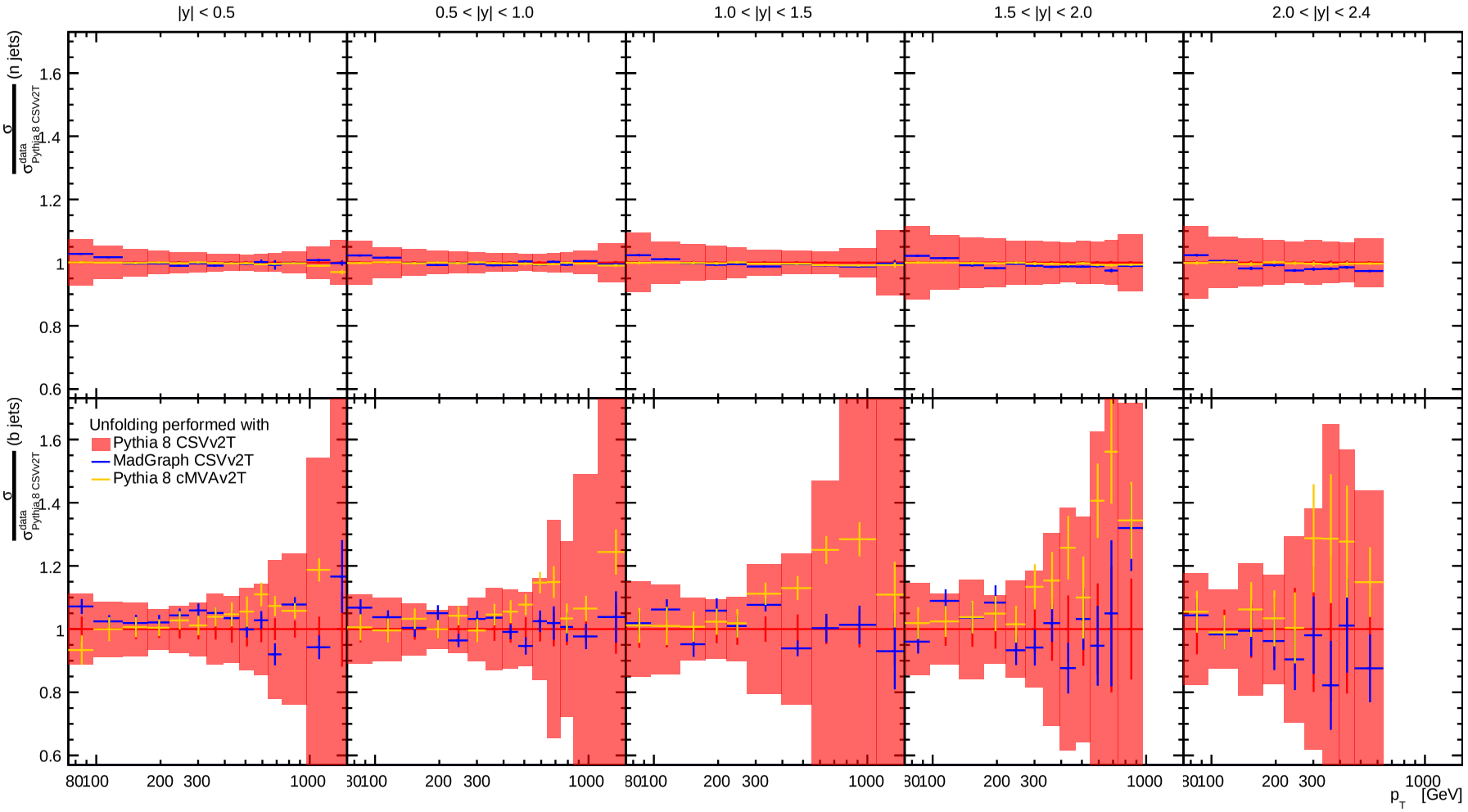


FIGURE 8.10 – Comparison of the unfolding in the different scenarios. The rows (columns) correspond to the flavour (rapidity). All cross sections are divided by the cross section obtained after unfolding with PYTHIA 8 (`CSVv2`). Systematic and statistical (onle statistical) uncertainties are shown for PYTHIA 8 using `CSVv2` (for the two other scenarios).

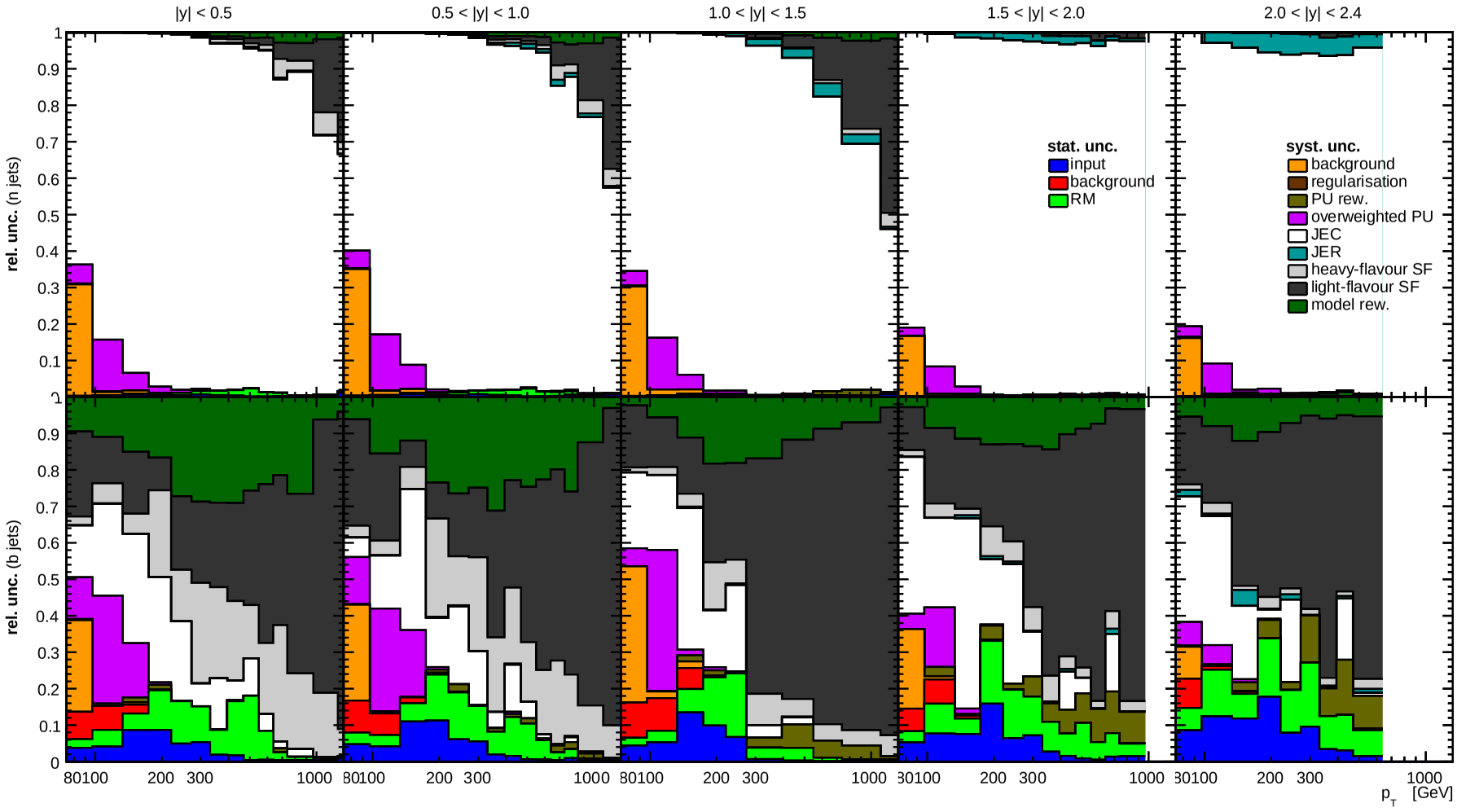


FIGURE 8.11 – Composition of the uncertainties after unfolding. The rows (columns) correspond to the flavour (rapidity). The uncertainties are relative, rescaled bin by bin. Both statistical and systematic uncertainties are shown.

to the MC prior; therefore, one can iterate the procedure by computing $\mathbb{P}[C_i]$ from the result of the previous iteration; with this improvement, the algorithm is called *D'Agostini's unfolding*. At each iteration, the bias to the MC prior is reduced.

Convergence. It converges to a maximum-likelihood estimator with Poisson-like errors. There cannot be any negative values, contrarily to the result from Tikhonov regularisation (which is closer to the matrix inversion). One major difficulty of the iterative approach is to determine when the convergence has been obtained, and when one should stop iterating. The number of iterations has to be determined case by case. In general, "[...] the convergence rate can be very slow and the number of iterations is expected to grow with the number of bins squared." [4]

Regularisation. The regularisation is performed by the choice of a *good* MC prior, *i.e.* a MC that models well the data. Therefore, the more iterations, the less biased to the MC prior, but the less regularised.

8.2.3.2 Results

We show here that the result with D'Agostini unfolding converges and is compatible with the result with Tikhonov unfolding.

The D'Agostini unfolding is performed with different numbers of iterations: 2, 4, 8, 16 and 32. Then the different curves (various yellow levels) are then compared to the Tikhonov unfolding (blue) and the MC prior (red) in Fig. 8.12. The D'Agostini unfolding is stable after a few iterations. For small numbers of iterations, one can see that it is slightly biased to PYTHIA 8. The result after a large number of iterations is compatible within the statistical uncertainties with the Tikhonov unfolding.

Conclusions

The determination of the purity and the procedure of unfolding to particle level with determination of the uncertainties have been presented. Different scenarios involving different samples, different taggers and different unfolding algorithm lead to compatible results. Additional checks are presented in the appendix regarding the determination of the purity and the procedure of unfolding.

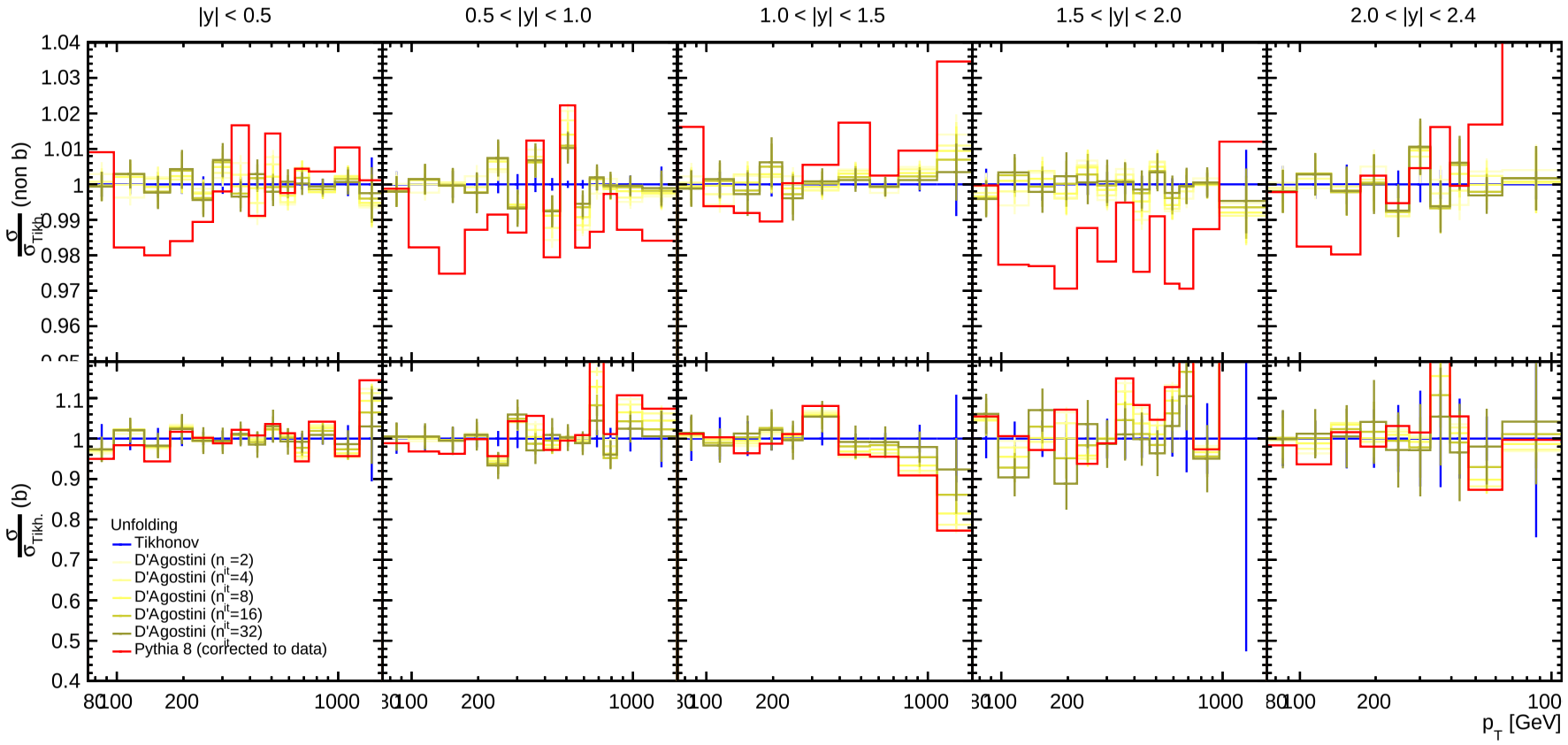


FIGURE 8.12 – The results of the D'Agostini unfolding with different numbers of iterations (yellow levels) are compared to the result with Tikhonov regularisation (blue, at one) and with the MC prior (red). Only the statistical uncertainty is represented. The rows (columns) correspond to the flavour (rapidity).

8.A Details about fit of purity

In this appendix, we describe step by step the determination of the purity. First, we show the templates. Then we investigate the different ways to constrain charm in the CSVv2-tagged region; we compare PYTHIA 8 and MADGRAPH. After, in order to justify our approach in the non-CSVv2-tagged region, we show that the fit is not stable.

8.A.1 Templates

In the discussion conducted in Sect. 8.1, the templates were sketched in Fig. 8.2; the templates in the CSVv2-tagged (non-CSVv2-tagged) region are shown in Fig. 8.13 (Fig. 8.14) in bins of p_T and y. For readability, the statistical errors are not shown; however, they become larger and larger for increasing JP values.

Tagged region. In the CSVv2-tagged region, the different templates are peaked at different values for the different flavours; however, the peak are less and less distinct while going to higher p_T and higher y; one also observes that the c templates lie halfway between the light and b templates.

Non-tagged region. On the other hand, in the non-CSVv2-tagged region, if the templates look different at low p_T, they very similar at high p_T. Moreover, the light component has a roughly 50 times larger statistics.

8.A.2 Determination of the purity in the CSVv2-tagged region

In Figs. 8.15-8.16, the fit is investigated for different ways to constrain the c component with the PYTHIA 8 and MADGRAPH samples:

— as an independent component, *i.e.* just as in Eq. 8.5 (corresponding to the blue circles);

— together with the light component, *i.e.* $N_{\text{data}}^{\text{total}} = p_{\text{udsg+c}}(N_{\text{MC}}^{\text{udsg}} + N_{\text{MC}}^{\text{c}}) + p_{\text{b}} N_{\text{MC}}^{\text{b}}$ (corresponding to the purple squares)

— or together with the bottom component, *i.e.* $N_{\text{data}}^{\text{total}} = p_{\text{udsg}} N_{\text{MC}}^{\text{udsg}} + p_{\text{b+c}}(N_{\text{MC}}^{\text{c}} + N_{\text{MC}}^{\text{b}})$ (corresponding to the orange stars).

In all cases, the configuration where charm is constrained with bottom seems optimal, since at low p_T, the correction is expected to be minimal (the disagreement in the fraction ratio mainly happens at high p_T). The findings were also confirmed using cMVAv2 (not shown here); the result was however of lower quality because the correlation of cMVAv2 with JP is greater.

The ratios of the JP discriminant in bins of p_T and y is shown before (after) the fit in Fig. 8.17 (Fig. 8.18).

8.A.3 Determination of the purity in the non-CSVv2-tagged region

In the non-CSVv2-tagged region, the difference of statistics of the contributions from lights and from HF components. The attempt of fit with a similar approach as in the CSVv2-tagged region (in the previous subsection) is shown in Fig. 8.19; only the case of b and c constrained together converged (in almost all bins, except at low p_T in the central region, the different attempts of fit systematically returned NaN), therefore it is the only one that can be shown here. This failure justifies the solution mentioned in Sect. 8.1, where the

renormalisation factor of the $b+c$ component is propagated from the CSVv2-tagged to the non-CSVv2-tagged region, and the light component is only rescaled to match the data; the result of this procedure, with the different charm constraints, is shown in Fig. 8.20.

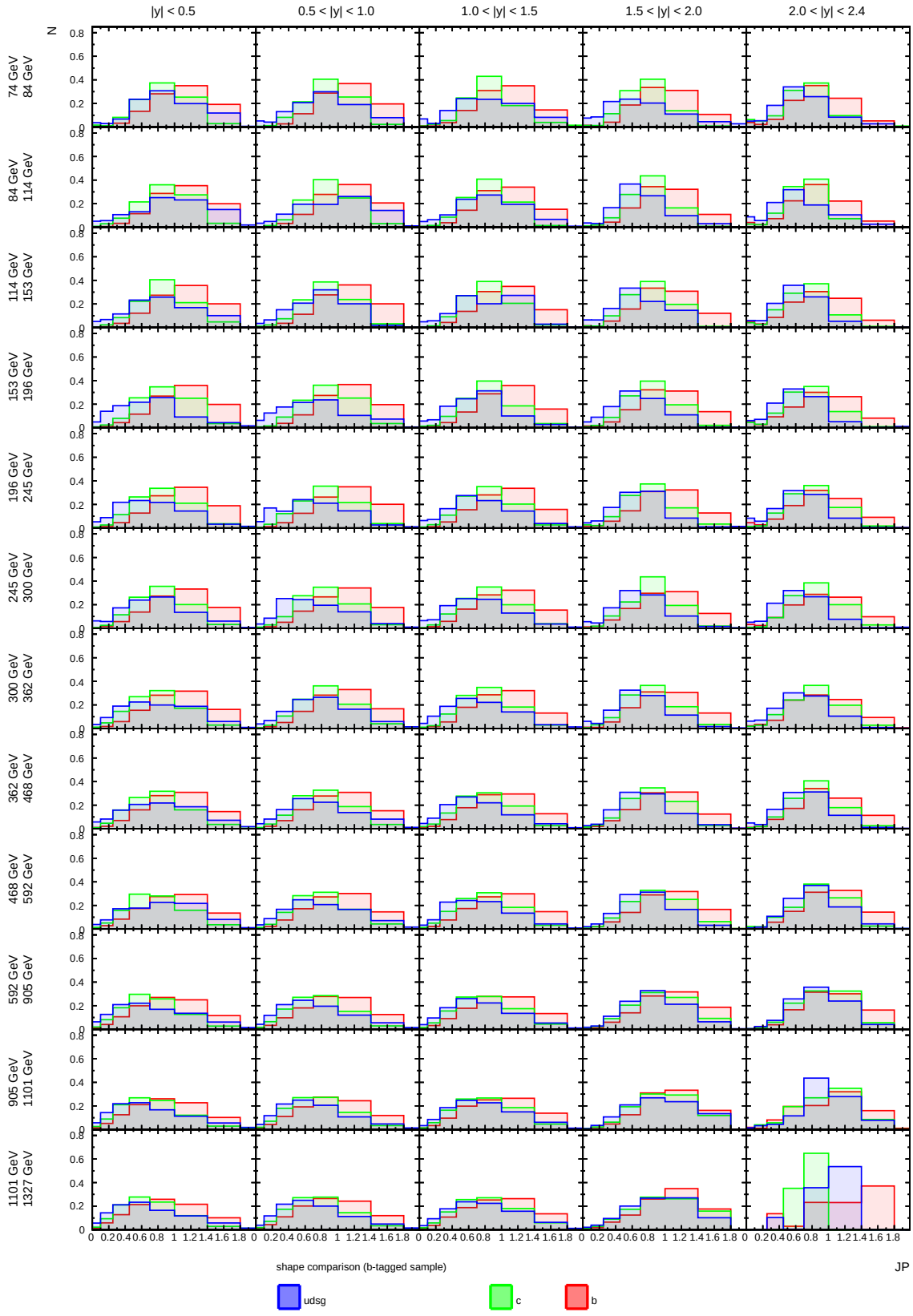


FIGURE 8.13 – The shape of the JP discriminant is shown for the different flavours in the CSVv2-tagged sample. Each grid corresponds to a (p_T, y) bin; the colours represent the different flavours.

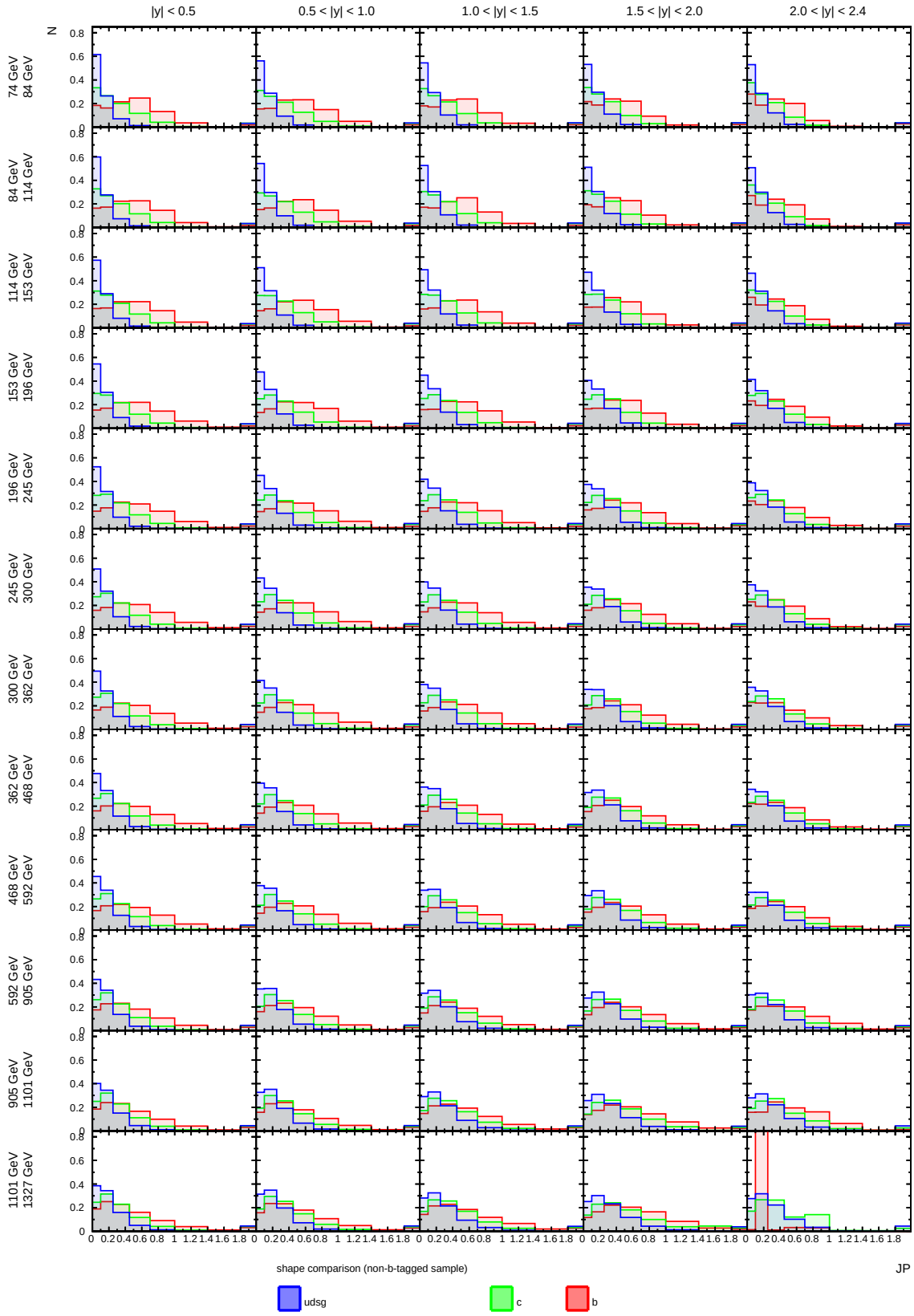


FIGURE 8.14 – The shape of the JP discriminant is shown for the different flavours in the non-CSVv2-tagged sample. Each grid corresponds to a (p_T, y) bin; the colours represent the different flavours.

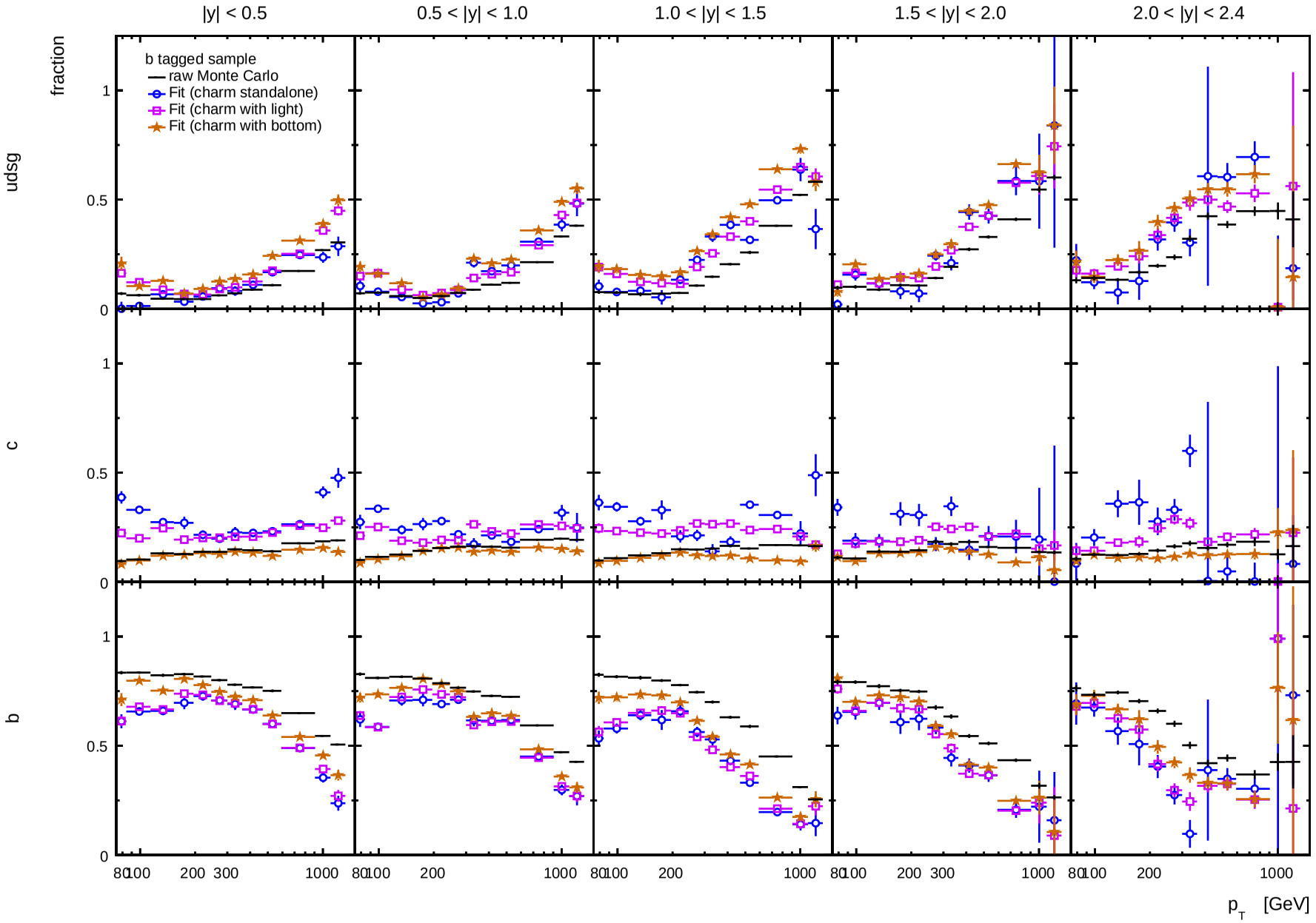


FIGURE 8.15 – Purity in simulation before and after fit in the CSVv2-tagged region with PYTHIA 8. The columns (rows) correspond to the rapidity bins (flavours). Different configurations to constrain charm are considered.

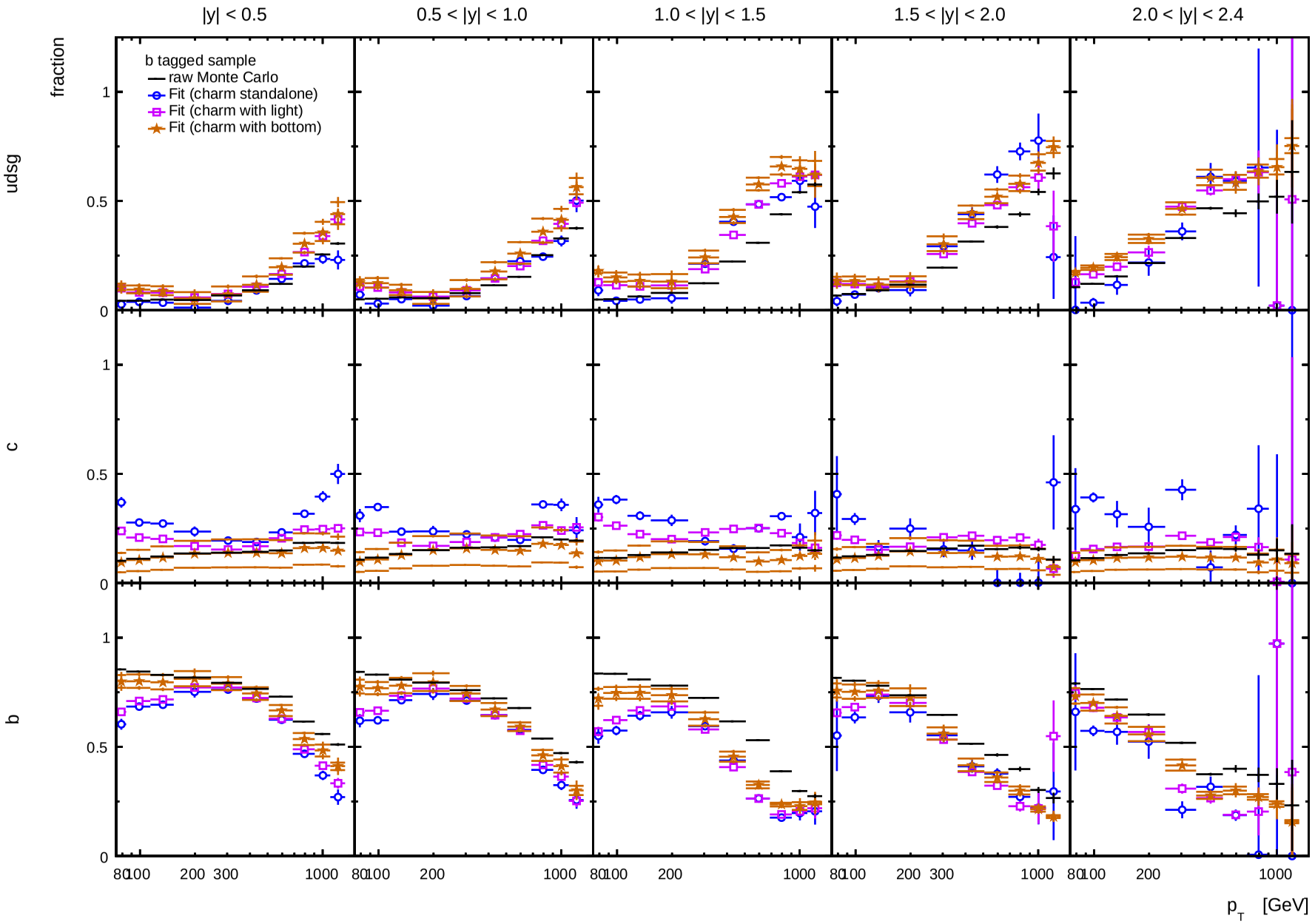


FIGURE 8.16 – Purity in simulation before and after fit in the CSVv2-tagged region with MADGRAPH. The columns (rows) correspond to the rapidity bins (flavours). Different configurations to constrain charm are considered.

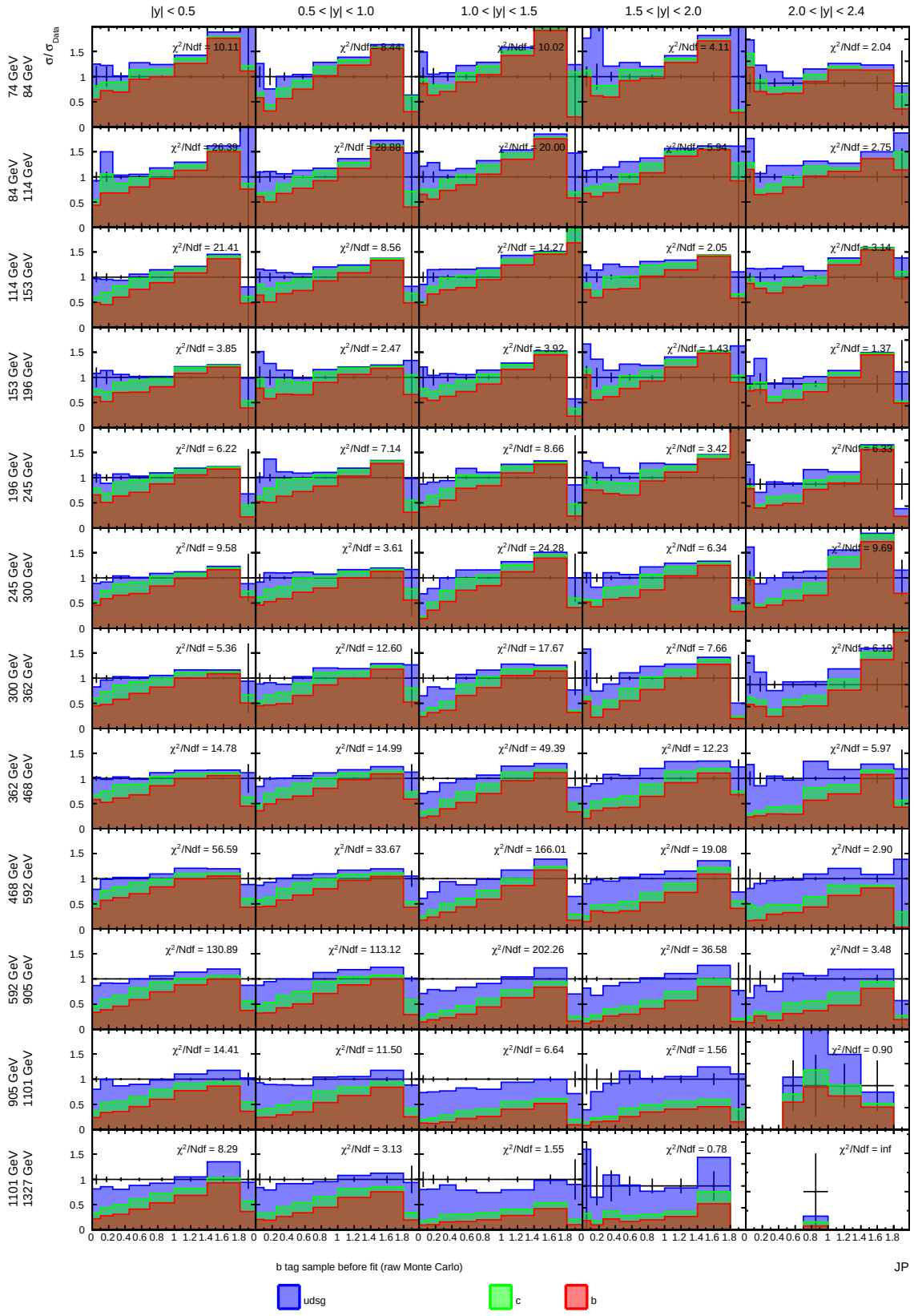


FIGURE 8.17 – Ratio of the JP discriminant in the CSVv2-tagged region, of simulation to data in bins of (p_T, y) before the fit; the colours stands for the flavour. The χ^2 per n.d.f. is given.

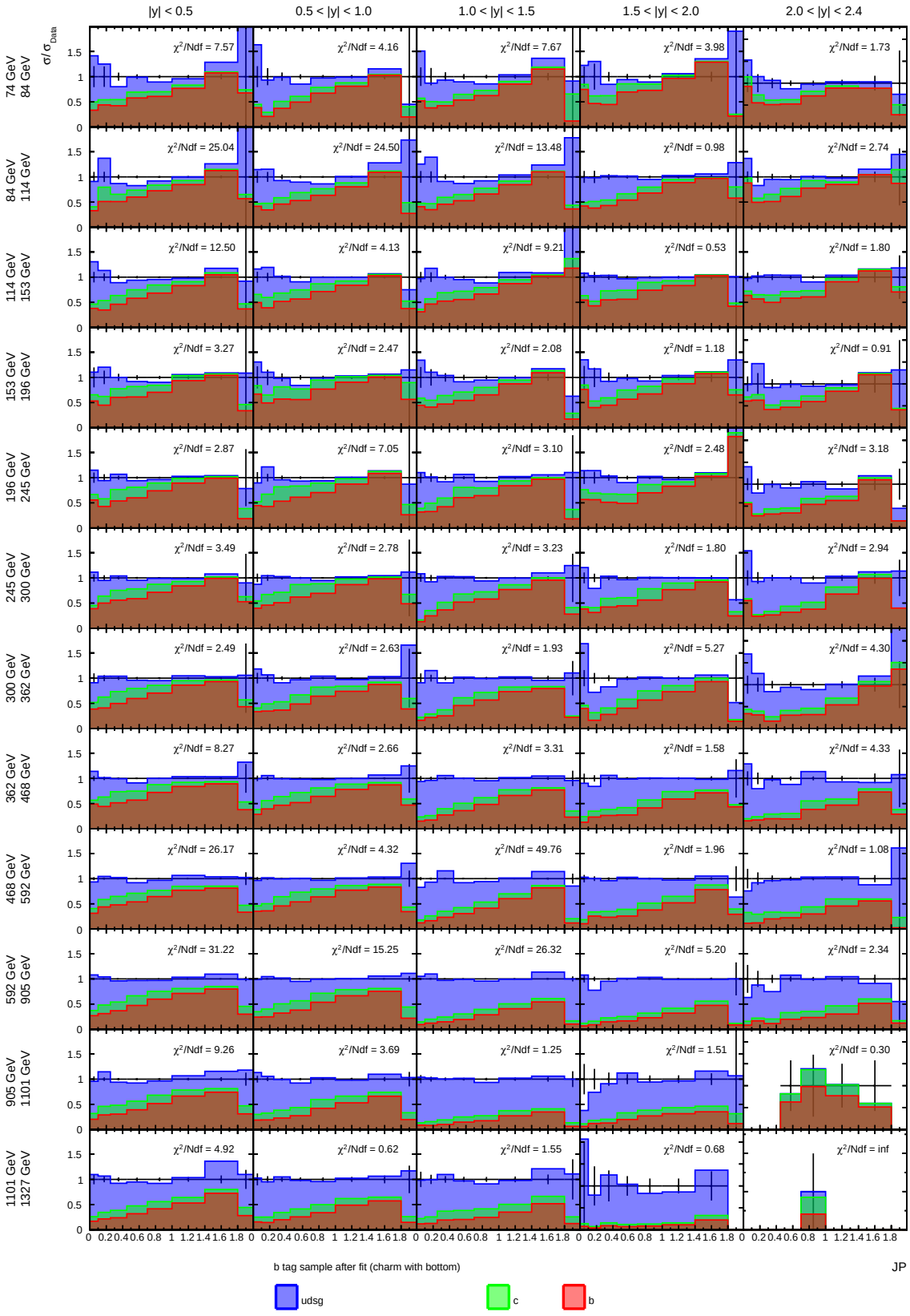


FIGURE 8.18 – Ratio of the JP discriminant in the CSVv2-tagged region, of simulation to data in bins of (p_T, y) after the fit with b and c constrained together; the colours stands for the flavour. The χ^2 per n.d.f. is given.

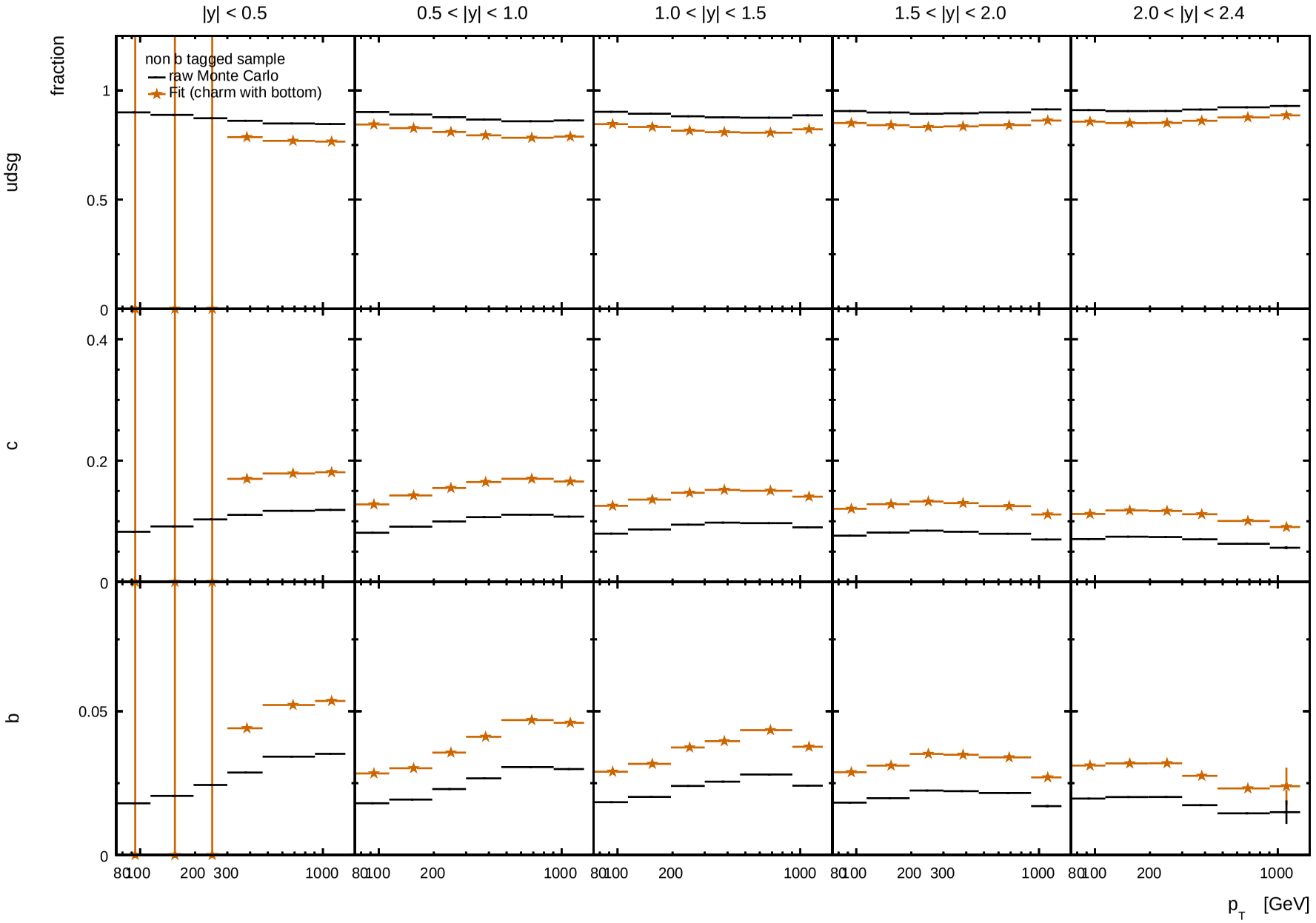


FIGURE 8.19 – Purity in simulation before and after fit in the CSVv2-tagged region with PYTHIA 8. The columns (rows) correspond to the rapidity bins (flavours). Only the case where charm and bottom are constrained together is considered.

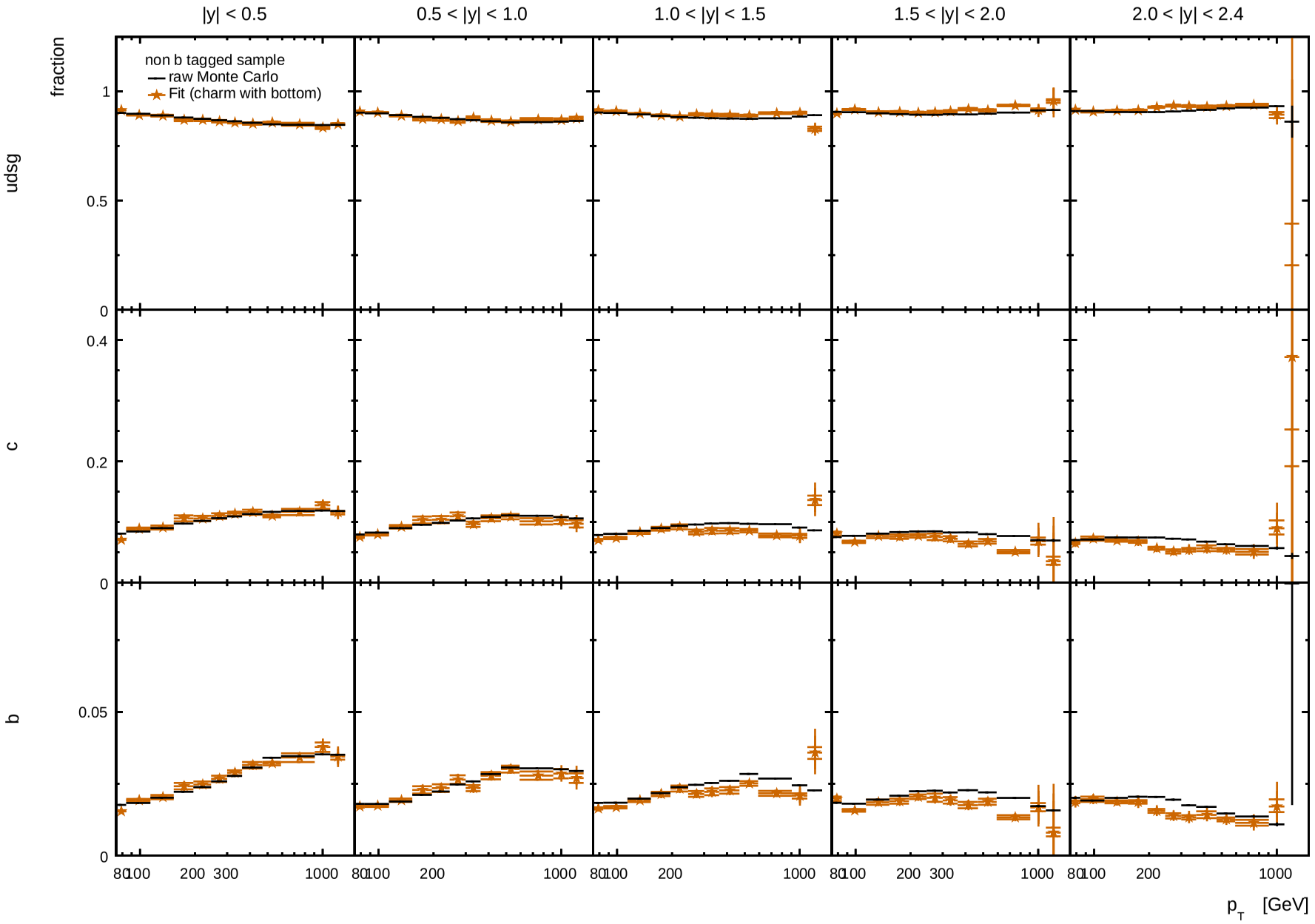


FIGURE 8.20 – Purity in simulation before and after extrapolation of the renormalisation of the b and c component and rescaling of the light component in the non-CSVv2-tagged region with PYTHIA 8. The columns (rows) correspond to the rapidity bins (flavours). Different configurations to constrain charm are considered.

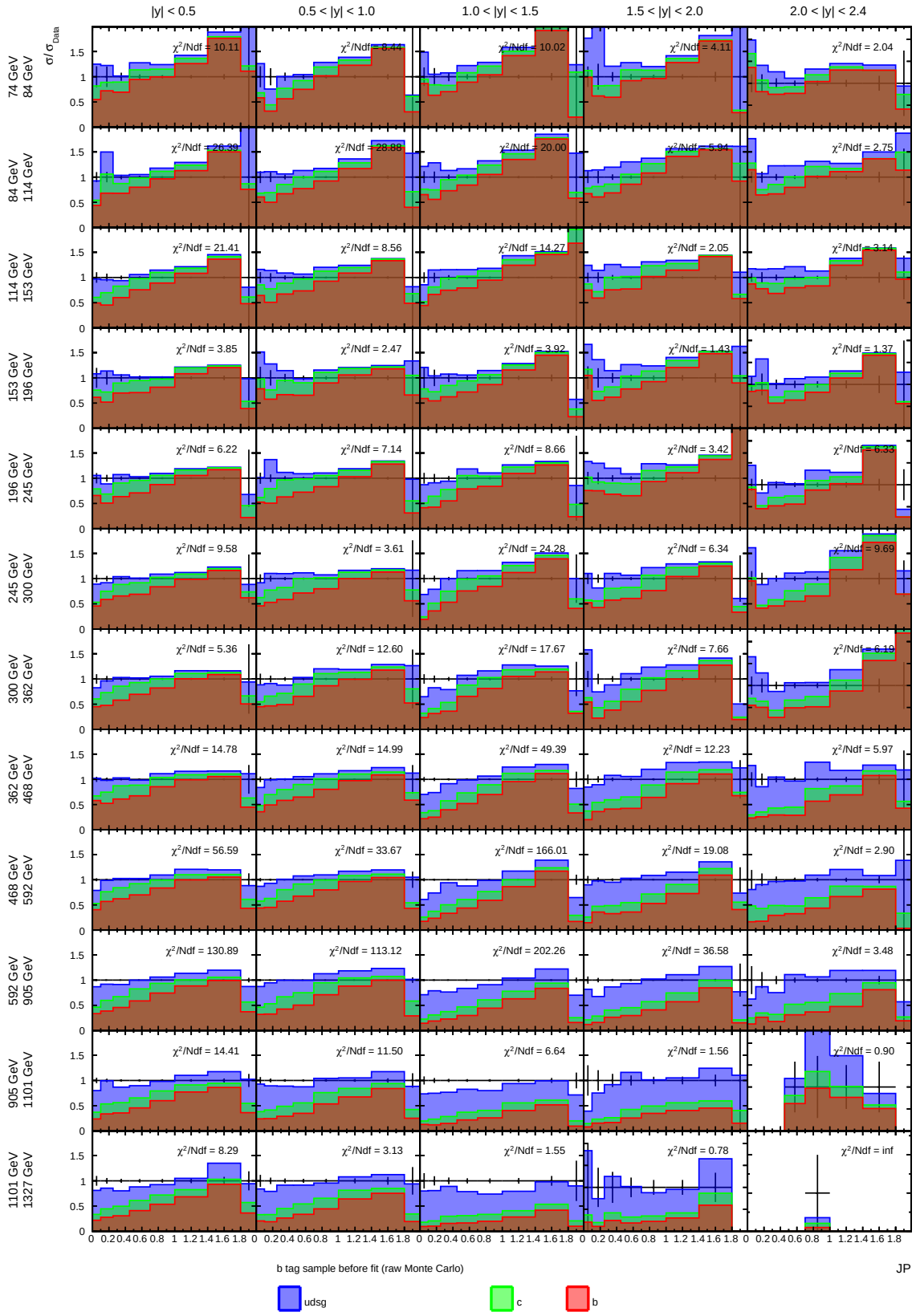


FIGURE 8.21 – Ratio of the JP discriminant of simulation to data in the non-CSVv2-tagged region in bins of (p_T, y) before the fit; the colours stands for the flavour. The χ^2 per n.d.f. is given.

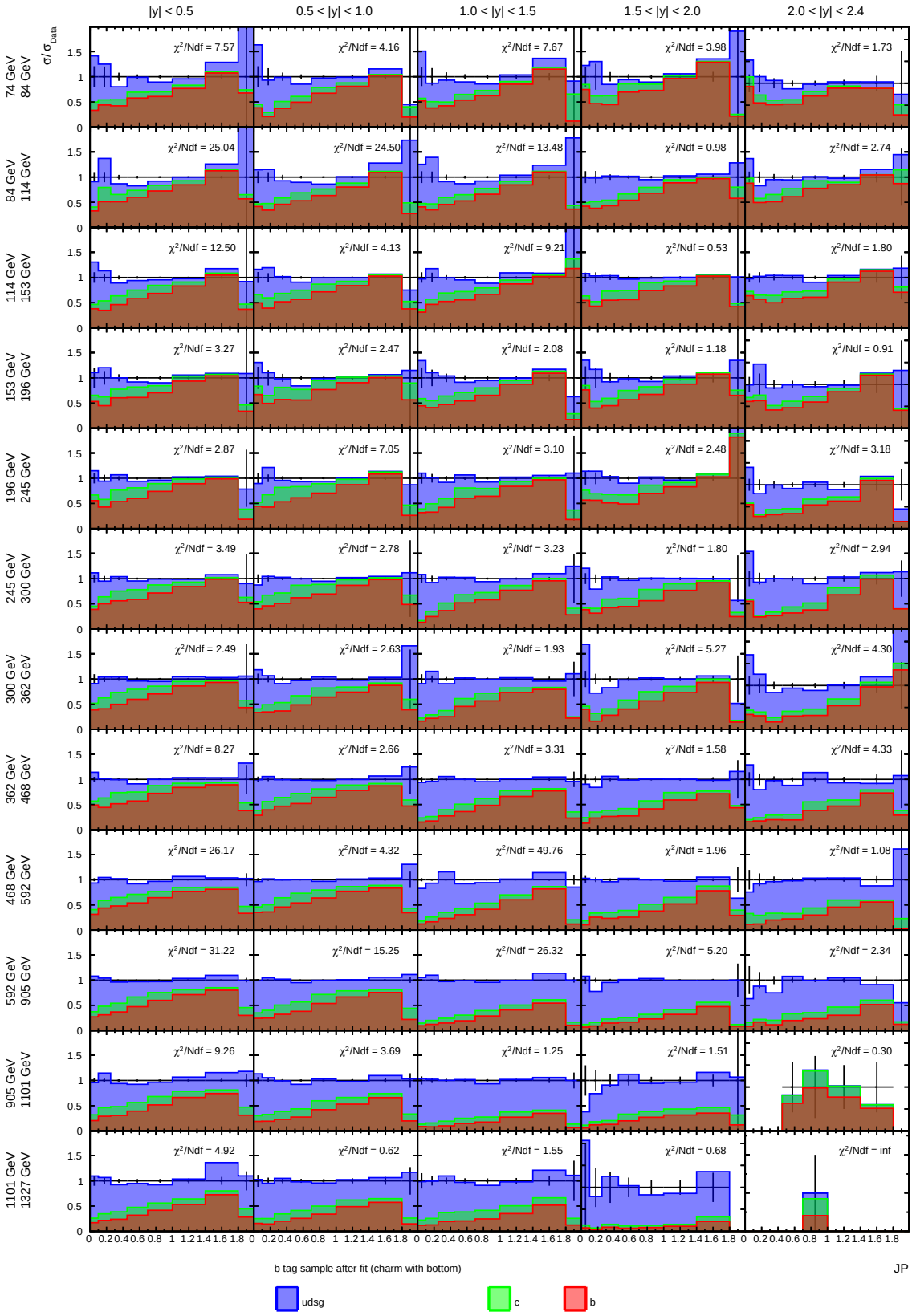


FIGURE 8.22 – Ratio of the JP discriminant of simulation to data in the non-CSVv2-tagged region in bins of (p_T, y) after the fit with b and c constrained together; the colours stands for the flavour. The χ^2 per n.d.f. is given.

8.B Details about unfolding procedure

We give additional details in the procedure of unfolding.

First we show the regularisation matrix obtained from PYTHIA 8. Then we discuss the treatment of the statistical uncertainties. Finally, we present additional checks.

8.B.1 Control plots for Tikhonov regularisation

The $\mathbf{L}$ matrix obtained with the PYTHIA 8 sample after model reweighting may be seen on Fig. 8.23: since only the p_T is regularised, only the diagonal of the rapidity cells are filled; the constraint index corresponds to the row index in Eq. 8.14, *i.e.* to a constraint on three consecutive bins at particle level. The $\mathbf{L}$ matrix is therefore not a square matrix.

The effect of the regularisation on the unfolded spectra can be checked with the product $\mathbf{Lx}$ (second term in Eq. 8.11). The product is shown in the three different scenarios in Fig. 8.24. One sees explicitly which bins need the more regularisation: at high p_T, especially in the third rapidity bin.

8.B.2 Treatment of statistical uncertainties

The statistical uncertainty from the MC (via the RM) and from the data (from the measurement) is considered; the former is included as an additional uncertainty in the unfolding procedure, the latter is part of the unfolded result.

The covariance matrix in data at particle level is given by the following:

$$\mathbf{V_x} = \mathbf{B}^\intercal {\mathbf{V_y}}^{-1} \mathbf{B}^\intercal \tag{8.20}$$

where $\mathbf{B}$, which operates the transformation, is defined as follows:

$$\mathbf{B} = \mathbf{E}\mathbf{A}^\intercal {\mathbf{V_y}}^{-1} \tag{8.21}$$

with

$$\mathbf{E} = \left(\mathbf{A}^\intercal {\mathbf{V_y}}^{-1} \mathbf{A} + \tau^2 \mathbf{L}^\intercal \mathbf{L}\right)^{-1} \tag{8.22}$$

In the case of no regularisation ($\tau = 0$), the transformation simplifies to $\mathbf{B} = \mathbf{A}^{-1}$, as expected for matrix inversion.

The covariance matrix before (after) unfolding can be seen on Fig. 8.25 (Fig. 8.26). The input covariance matrix contains only positive entries; off-diagonal events corresponds to correlations among jets coming from the same events. A single-count observable would show purely diagonal covariance matrices; here, since we are measuring a multi-count observable, there are significant non-diagonal contributions, which matter in the unfolding (see Eq. 8.10). The output covariance matrix contains negative entries (which translates into this chess-pattern); indeed, close bins are constrained together and are therefore correlated.

8.B.3 Additional checks

We present here some additional checks to certify the unfolding. We compare systematically the result of the unfolding obtained with the D'Agostini and Tikhonov regularisations.

8.B.3.1 Backfolding

The *backfolding* consists in applying the PM on the particle-level spectrum

$$\mathbf{y}' = \mathbf{A}\mathbf{x} \tag{8.23}$$

The backfolded spectrum $\mathbf{y}'$ can be compared with the measurement $\mathbf{y}$. The difference is expected be of the order of the statistical fluctuations; however, since the backfolded spectrum still keeps track of the regularisation (either from the MC prior with D'Agostini or from the $\mathbf{L}$ matrix with Tikhonov), therefore fluctuations are expectable.

The backfolding after the two algorithms is shown in Fig. 8.27. The curves are compatible with the statistical uncertainties, both for $\hat{n}$ (above) and $\hat{b}$ jets (below). The remaining fluctuations are similar for the two backfolded spectra (with the different algorithms) and for the simulation, and give an estimate of the effect of the regularisation.

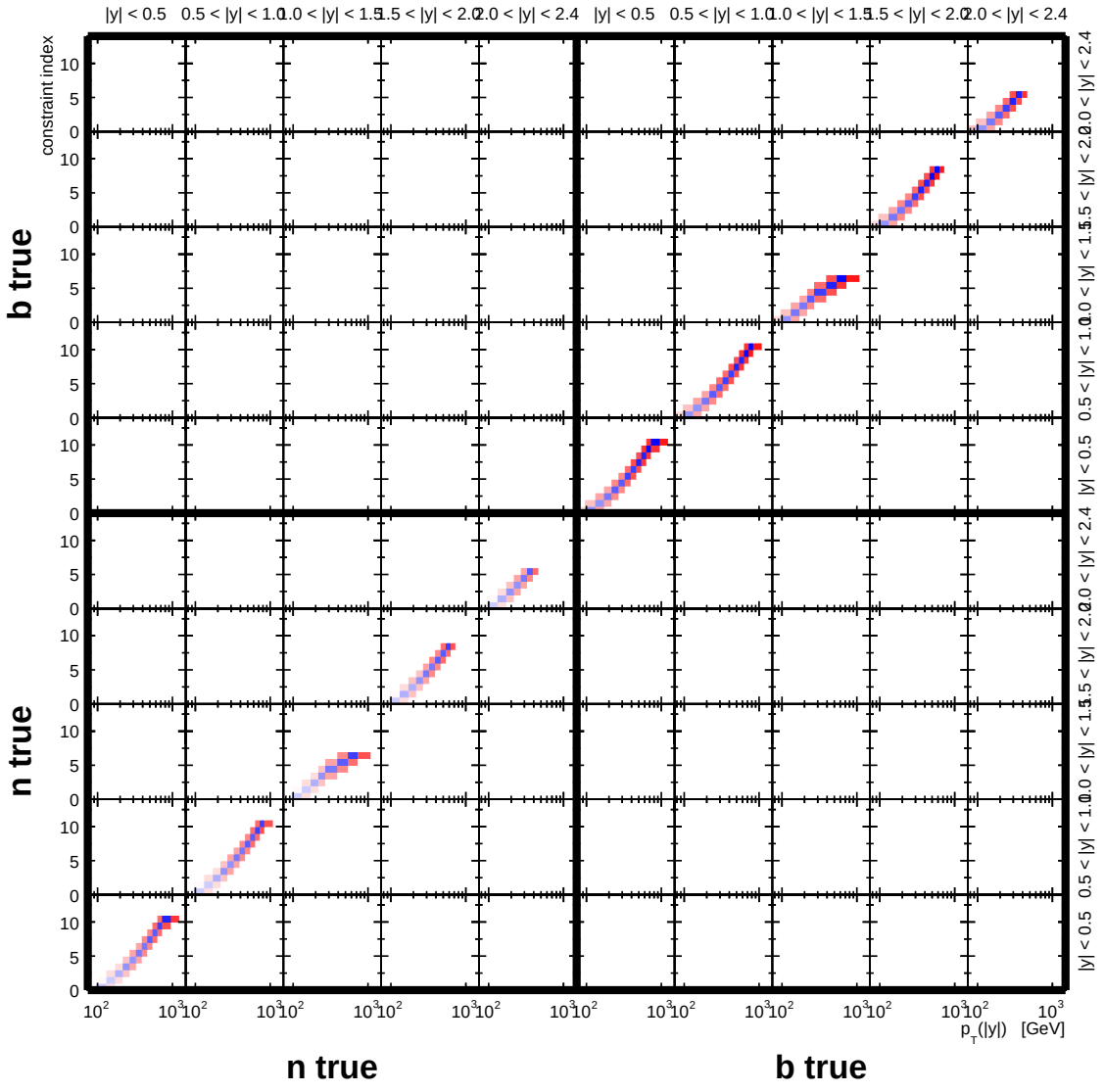


FIGURE 8.23 – **L** matrix constructed from PYTHIA 8. The x axis (y axis) stands for the $(p_T, y, \text{flavour})$ bins (constraint index). The values are given in arbitrary units; the blue (red) entries correspond to positive (negative) entries. The level of transparency is proportional to the absolute value.

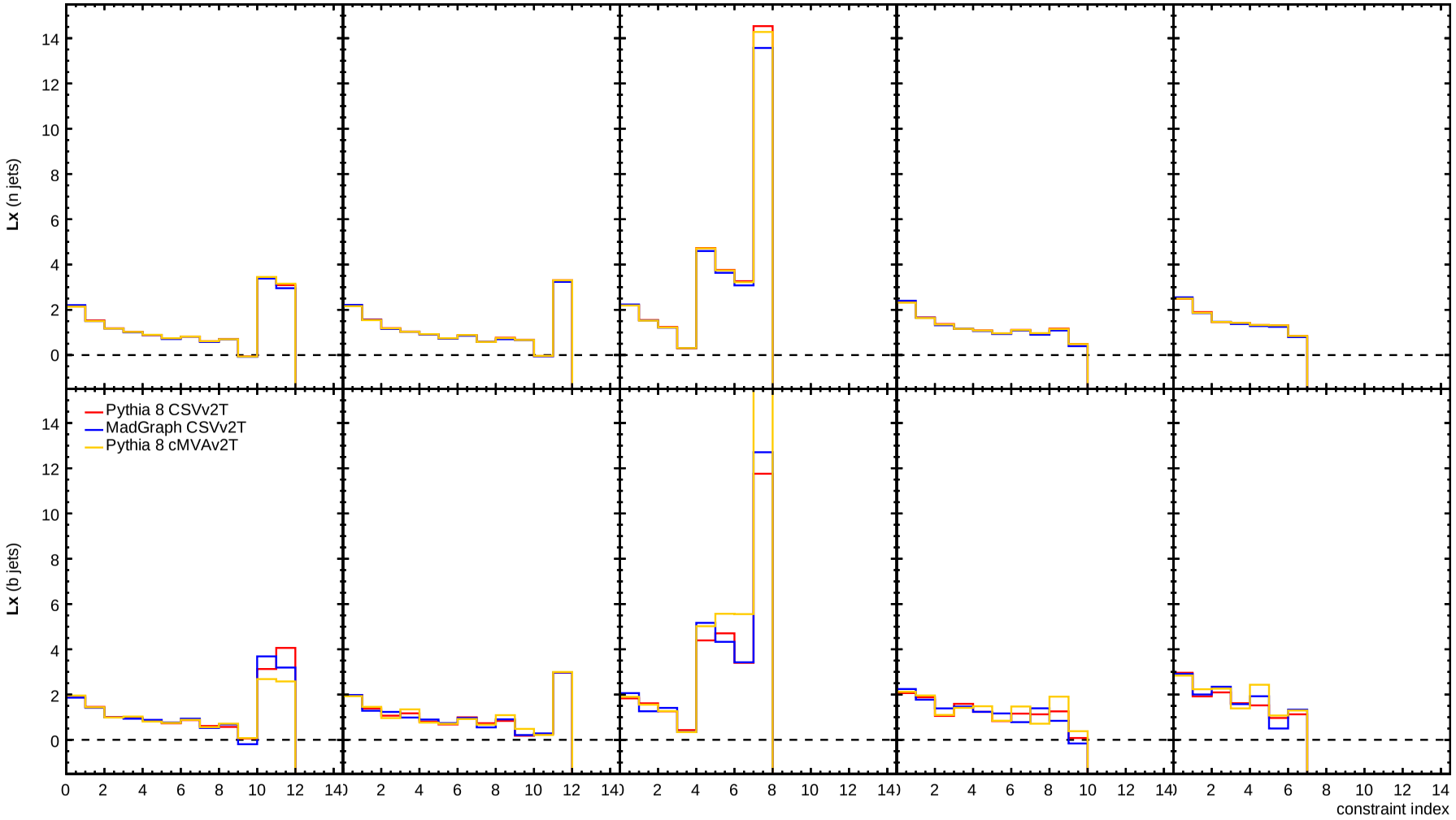


FIGURE 8.24 – The product of regularisation matrix and unfolded result is shown in bins of rapidity (columns) and flavour (rows), shown for the three scenarios.

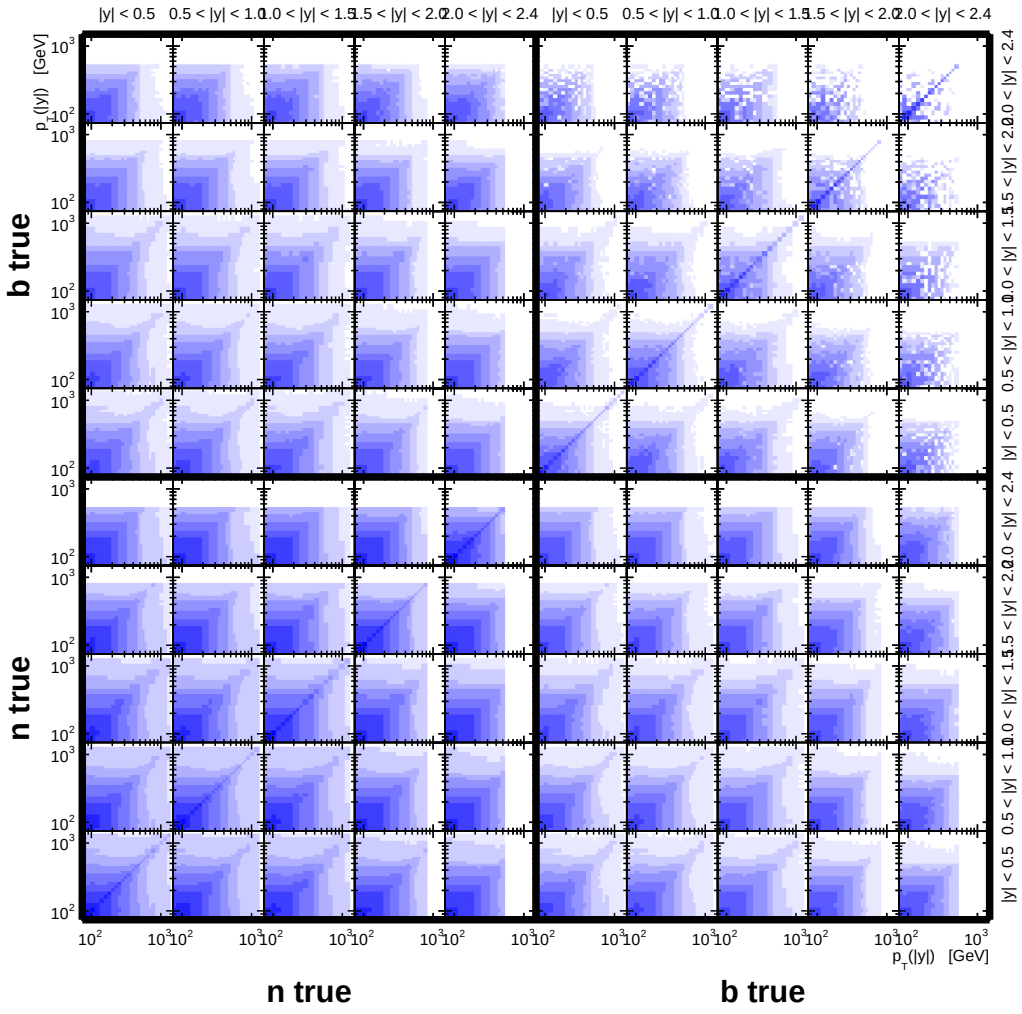


FIGURE 8.25 – Covariance matrix from measurement. The large sectors correspond to the flavour bins, the cells to the rapidity bins and the small matrices to the p_T bins. The level of transparency denotes the magnitude of the content in arbitrary units. All entries are positive. Off-diagonal entries show correlations among jets coming from the same events.

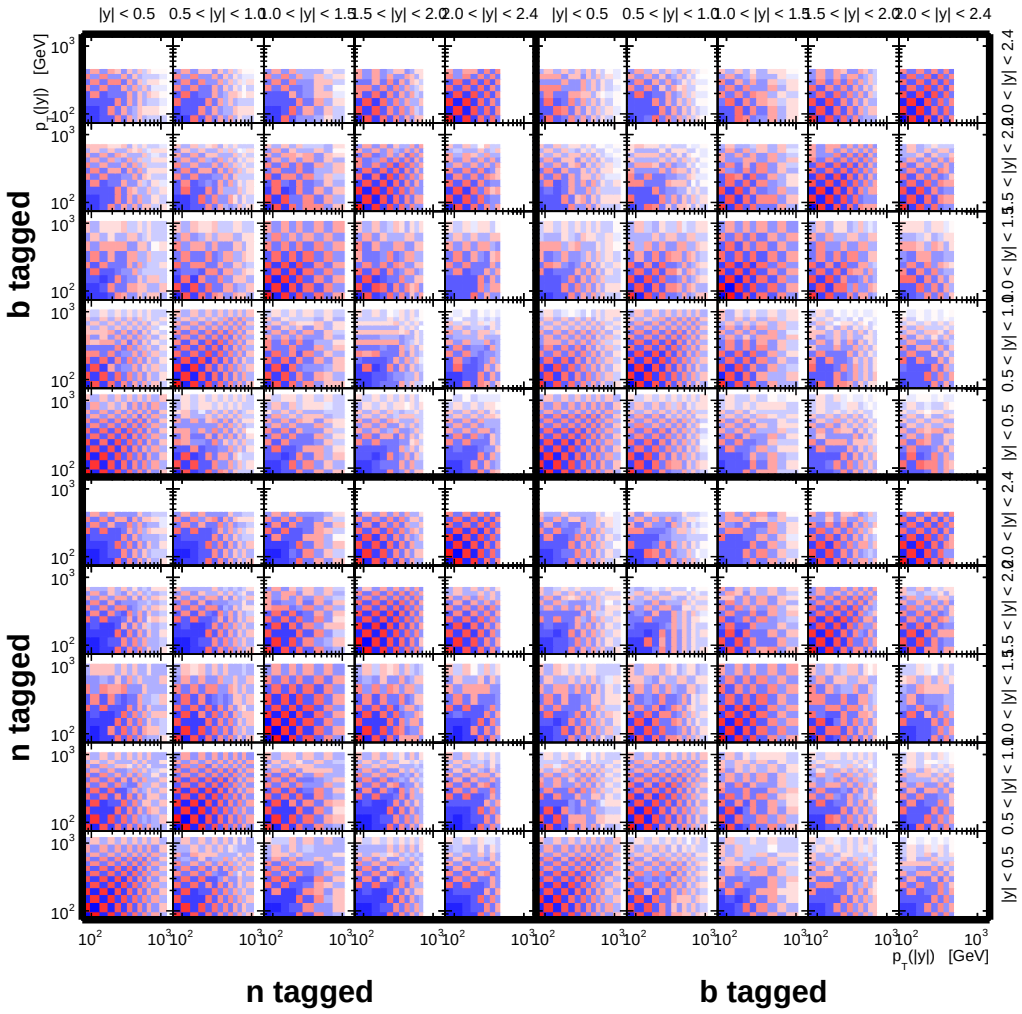


FIGURE 8.26 – Total covariance matrix after unfolding procedure. The large sectors correspond to the flavour bins, the cells to the rapidity bins and the small matrices to the p_T bins. The level of transparency denotes the magnitude of the content in arbitrary units. The positive (negative) entries are coloured in blue (red).

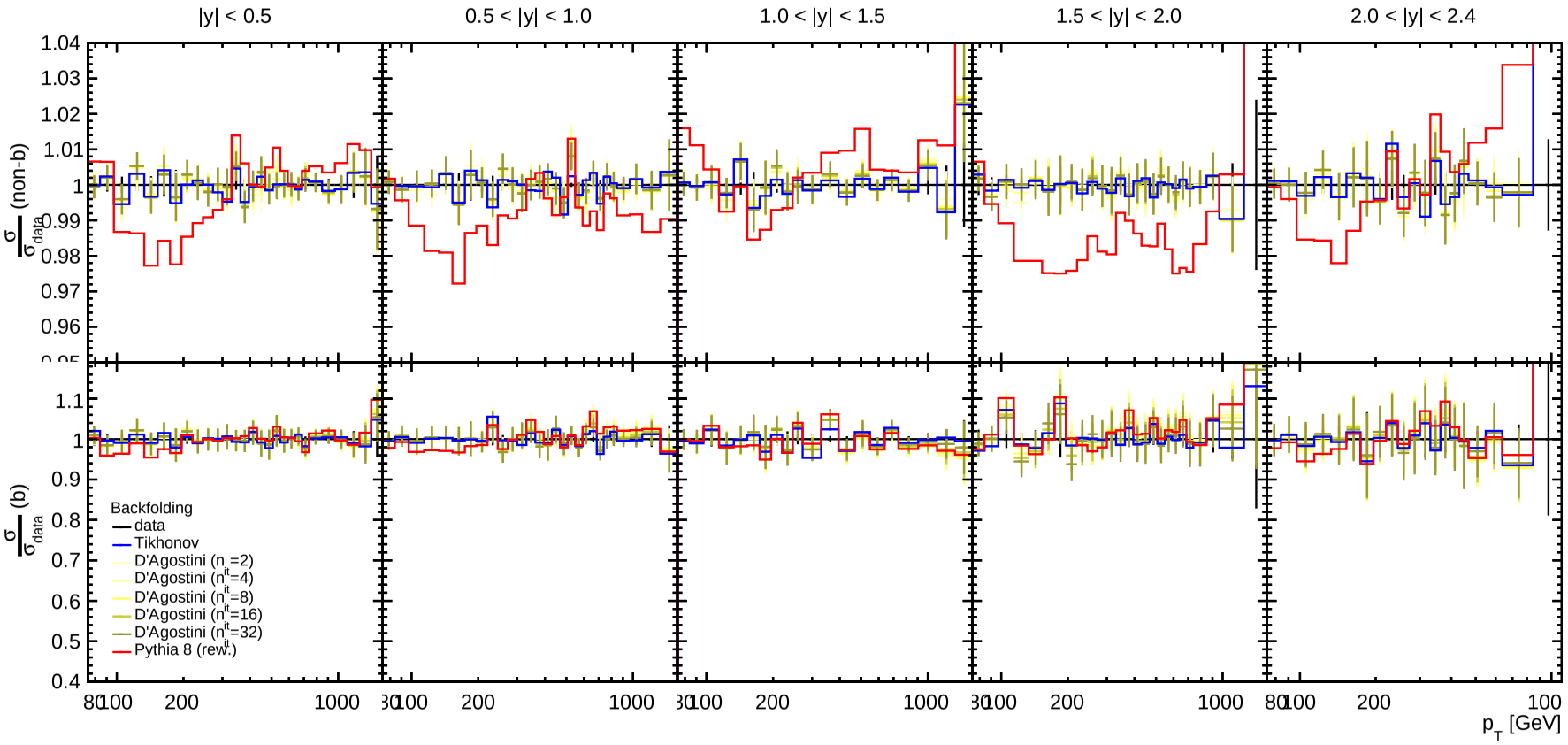


FIGURE 8.27 – The backfolding is compared for Tikhonov (blue) and D'Agostini (yellow) algorithms with the measurement (black, at one) and with PYTHIA 8 using `CSVv2` (red). Different numbers of iterations are shown for D'Agostini. The rows (columns) correspond to the flavour (rapidity).

8.B.3.2 χ^2 of agreement

The χ^2 of agreement is defined in Eq. 8.10. It is shown on the left hand side of Fig. 8.28 for the different iterations and for the unfolding obtained with Tikhonov. One observes the converge to the D'Agostini unfolding to a value close to the one of Tikhonov.

8.B.3.3 χ^2 of change

The χ^2 of change is defined as follows:

$$\chi^2 = (\mathbf{x}_i - \mathbf{x}_{i-1})^\intercal \mathbf{V}^{-1} (\mathbf{x}_i - \mathbf{x}_{i-1}) \tag{8.24}$$

where i denotes the iteration. It is shown on the right hand side of Fig. 8.28, where it is given for the change 2^i to 2^{i+1} iterations. The change is smaller and smaller, indicating the convergence.

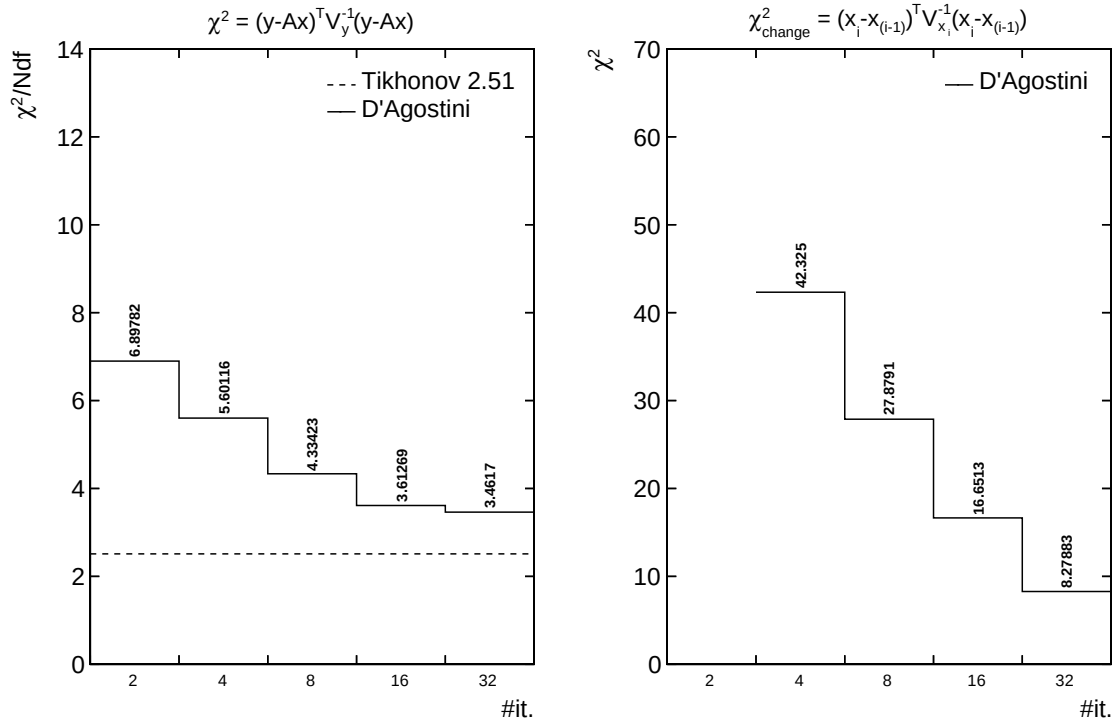


FIGURE 8.28 – χ^2s of agreement (left) and of change (right). The bins corresponds to different numbers of iterations in D'Agostini unfolding. The result obtained with D'Agostini (Tikhonov) unfolding is shown by a continuous (dashed) line.

8.B.3.4 Bottom Line Test

In the Bottom Line Test (BLT), we compare the agreement of simulation and data before unfolding, after unfolding and after backfolding by computing the following χ^2:

$$\chi^2_{\mathrm{BLT}} = (\mathbf{z}_{\mathrm{data}} - \mathbf{z}_{\mathrm{MC}})^\intercal \, \mathbf{V}^{-1}_{\mathrm{data}} \, (\mathbf{z}_{\mathrm{data}} - \mathbf{z}_{\mathrm{MC}}) \tag{8.25}$$

where $\mathbf{z} = \mathbf{y}$ (before) or $\mathbf{z} = \mathbf{x}$ (after) with respective data covariance matrix. One compares the values in the Tikhonov algorithm and for different number of iterations in the D'Agostini algorithm. If the unfolding is correctly performed, *i.e.* if only the effect of

the detector is treated, then one does not expect the agreement to change significantly at the different levels.

The test performed with the PYTHIA 8 sample can be seen in Fig. 8.29, where the value of Eq. 8.25 are shown for the different number of iterations of the D'Agostini unfolding in the bins and with a single line for the Tikhonov unfolding:

— The BLT of the unfolding is shown on the left hand side of the figure. The higher number of iterations does not improve the global agreement of the result obtained with the D'Agostini algorithm. Moreover, the result obtained from D'Agostini has larger uncertainties, and therefore leads to a lower χ^2_{BLT} than with the Tikhonov algorithm.

— The BLT of the backfolding is shown on the right hand side of the figure. In contrast to the BLT at particle level, the χ^2_{BLT} takes different values for D'Agostini, which is likely related to the treatment of the uncertainties. However, it goes to values of the same order as the result obtained with the Tikhonov algorithm and as the measurement. The fact the backfolding after Tikhonov algorithm has a lower χ^2_{BLT} than the measurement is explained by the regularisation; indeed, the backfolded spectrum is still regularised, with respect to the measurement.

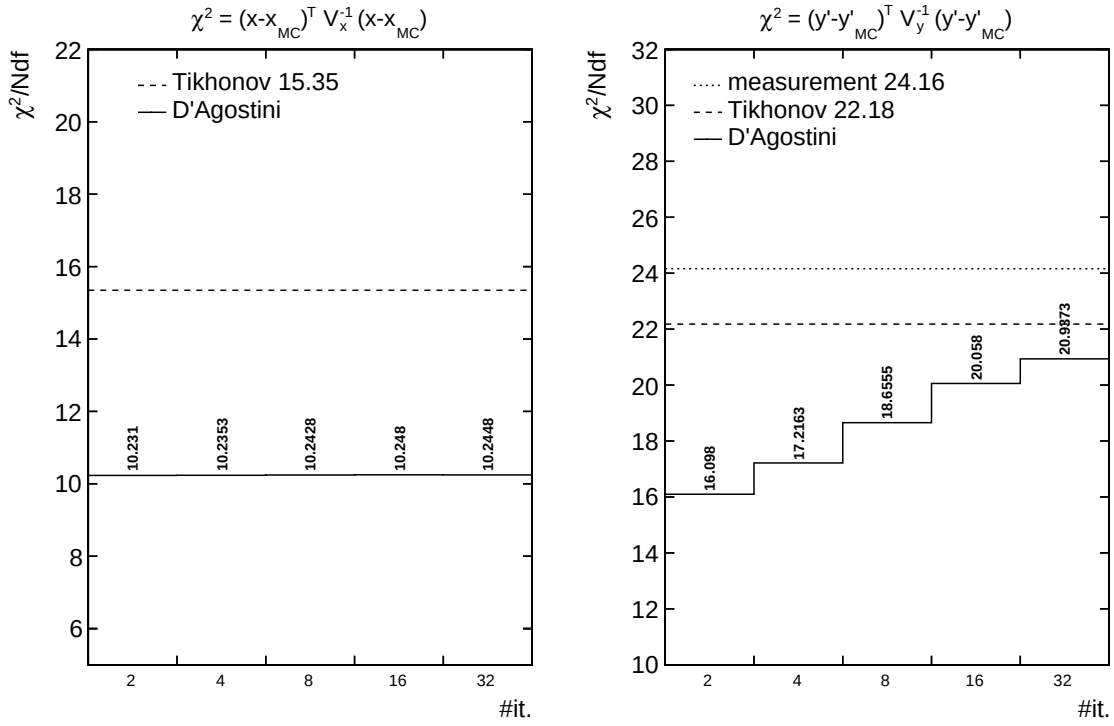


FIGURE 8.29 – Bottom Line Test, on the left (right) at hadron-level (detector-level) for the unfolding (backfolding). The iteration value corresponds to the one used in D'Agostini unfolding (continuous line), while the Tikhonov unfolding (dashed line) and the measurement (dashed line) has only one value.

References

[1] Ivan Marchesini and Kirill Skovpen. *Discussions on calibration with JP*. Private communication.

[2] Kuusela, Mikael. *Introduction to Unfolding in High Energy Physics*. http://smat.epfl.ch/~kuusela/talks/ETH_Jul_2014.pdf. Accessed: 2017-11-30.

[3] Stefan Schmitt. "TUnfold: an algorithm for correcting migration effects in high energy physics". In: *JINST* 7 (2012), T10003. DOI: 10.1088/1748-0221/7/10/T10003. arXiv: 1205.6201 [physics.data-an].

[4] Stefan Schmitt. "Data Unfolding Methods in High Energy Physics". In: *EPJ Web Conf.* 137 (2017), p. 11008. DOI: 10.1051/epjconf/201713711008. arXiv: 1611.01927 [physics.data-an].

[5] Andreas Hocker and Vakhtang Kartvelishvili. "SVD approach to data unfolding". In: *Nucl. Instrum. Meth.* A372 (1996), pp. 469–481. DOI: 10.1016/0168-9002(95)01478-0. arXiv: hep-ph/9509307 [hep-ph].

[6] Stefan Schmitt. *Personal Homepage*. http://www.desy.de/~sschmitt/tunfold.html. Accessed: 2017-09-24.

[7] G. D'Agostini. "A Multidimensional unfolding method based on Bayes' theorem". In: *Nucl. Instrum. Meth.* A362 (1995), pp. 487–498. DOI: 10.1016/0168-9002(95)00274-X.

[8] G. D'Agostini. "Improved iterative Bayesian unfolding". In: *ArXiv e-prints* (Oct. 2010). arXiv: 1010.0632 [physics.data-an].

[9] H.N. Mülthei and B. Schorr. "On an iterative method for the unfolding of spectra". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 257.2 (1987), pp. 371–377. ISSN: 0168-9002. DOI: https://doi.org/10.1016/0168-9002(87)90759-5. URL: http://www.sciencedirect.com/science/article/pii/0168900287907595.

[10] A.E. Shustov and S.E. Ulin. "Matrix of Response Functions for Deconvolution of Gamma-ray Spectra". In: *Physics Procedia* 74.Supplement C (2015). Fundamental Research in Particle Physics and Cosmophysics, pp. 399–404. ISSN: 1875-3892. DOI: https://doi.org/10.1016/j.phpro.2015.09.210. URL: http://www.sciencedirect.com/science/article/pii/S1875389215014091.

[11] Tim Adye. "Unfolding algorithms and tests using RooUnfold". In: arXiv:1105.1160 (May 2011). Comments: 6 pages, 5 figures, presented at PHYSTAT 2011, CERN, Geneva, Switzerland, January 2011, to be published in a CERN Yellow Report, 313–318. 6 p. URL: https://cds.cern.ch/record/1349242.

9

Results

We compare the measurement to theory predictions. We first compare to LO predictions with PYTHIA 8, MADGRAPH and HERWIG++; then we compare to NLO predictions with POWHEG including theoretical uncertainties.

9.1 Comparison to LO predictions

We compare the measurement with three different predictions:

— PYTHIA 8 + `CUETP8M1`

— MADGRAPH + `CUETP8M1`

— HERWIG++ + `CUETHppS1`

Their specifications were given in Chap. 6.

The hadron-level inclusive jet (b jet) double differential cross section as a function of the transverse momentum and rapidity is shown in Fig. 9.1 (Fig. 9.3) and compared to the predictions in Fig. 9.2 (Fig. 9.4).

The spectrum is measured over six order of magnitude, covering a large p_T range from 74 GeV up to the TeV scale. With respect to the previous measurements at 8 and 13 TeV at CMS [1, 2] (also presented in App. C), the inclusive jet cross section covers a similar p_T range, but with higher luminosity. The measurement of inclusive b jet cross section reaches much higher p_T values than the measurements at 7 TeV by CMS [3] and by ATLAS [4] and CMS [3].

In the absolute cross sections in Fig. 9.1 (Fig. 9.3), each curve corresponds to a rapidity bin. The different rapidity bins are rescaled with different factors so as not to overlap. The uncertainties are indicated with the yellow band; the statistical uncertainties are too small to appear on the figures [1].

In the ratio in Fig. 9.2 (Fig. 9.4), the theoretical predictions are divided by the measurement. The rapidity bins are shown in successive panels. The band corresponds to the total relative uncertainty, including the statistical uncertainties, the luminosity uncertainty and the uncertainties described in Sect. 8.2.2.6, all added in quadrature; in addition, the statistical uncertainties are shown with vertical bars (the horizontal ticks separate the contribution from statistical errors of the measurements and the contribution inferred from

1. The correlated uncertainties are shown in the App. 8.B.2.

MC through the unfolding). The relative uncertainty for inclusive jet (inclusive b jet) is below 10% everywhere (around $10-20\%$ up to a few hundred GeV).

We compare the result to the comparison at detector level of the inclusive jet and inclusive b tagged jet spectrum in Figs. 7.1-7.2. For the inclusive jet production, the ratios are similar; for the inclusive b jet production, we observe that the discrepancy at high-p_T has disappeared indeed, since this was due to a mis-calibration of the b tagging. The predictions from PYTHIA 8 and HERWIG++ are usually parallel to the measurements (except at low p_T), which may be only related to the normalisation of the simulation; on the other hand, MADGRAPH even fails at describing even the shape of the measurement.

The fraction of b jets is also given in Figs. 9.5-9.6. In the fraction ratio, the simulation is divided by the measurement. The uncertainties related to jet energy and to pile-up cancel in the fraction; however, the dominant uncertainties are related to the b tagging calibration and to the fit of the purity, and are not vanishing in the fraction. One observes a significant difference among, on one hand, PYTHIA 8 and MADGRAPH, and, on the other hand, HERWIG++, especially for central rapidity and at high transverse momentum. The predictions by PYTHIA 8 and MADGRAPH shows that for $p_T \gg m_b$, the dynamics for b jets do not differ from the dynamics for jets.

9.2 Comparison to NLO predictions

We compare the measurement with POWHEG matched to `CUETP8M1` [5, 6, 7].

First, we describe the simulation and the theoretical uncertainties. Second, we compare data and predictions. Finally, we investigate the different contributions in the simulation.

9.2.1 Theoretical predictions

POWHEG was already presented in Sect. 4.2. It allows to compute predictions at NLO, include a Sudakov factor with splitting functions at NLO and is interfaced with PYTHIA 8 for the PS, the MPI and the hadronisation.

The nominal value and the uncertainties of the theory predictions are derived as follows:

PDF The `NNPDF 3.0` set contains a hundred of replicas. In the generation process, each event is varied according to the replicas; each replica therefore leads to a different cross section. The determination of the nominal value and of the uncertainties is performed bin after bin:

- The nominal value of the cross section is taken as the mean of the variations.
- The replicas are ordered by increasing value, and the 16th (84th) variations are taken as lower (upper) variation, corresponding to a variation of $\pm 1\sigma$ up and down.

Scale variations The renormalisation and factorisation scales are varied with a factor 1/2 down and 2 up. Four variations are considered, corresponding to independent variations of the two scales. The envelope of the scale variations is taken as uncertainty. The joint variation does not change the uncertainty band significantly (not shown here).

PS variations The scale in the PS is also varied with a factor 1/2 down and 2 up.

Tune variations The parameters of `CUETP8M1` are varied, correspondingly to the values shown in Tab. 9.1. Not all parameters are varied, but only the ones proper to `CUETP8M1`, *i.e.* parameters related to MPI (described in Eq. 2.60) and to colour reconnection. The whole list of tune parameters can be found in App. 9.A.

Fragmentation functions Finally, since the b jets are defined at hadron level, an uncertainty on the Bowler factor in the FF (Eq. 2.64) is considered:

$$r_B = 0.895^{+0.184}_{-0.197} \tag{9.1}$$

The uncertainties are then summed in quadrature separately up and down.

Figs. 9.7-9.8 show the composition of the theoretical uncertainties normalised to unity in each bin for the inclusive jet and inclusive b jet productions. The scale uncertainties dominate the uncertainties (lower row); the PS uncertainties get larger and larger at higher and higher p_T; the fragmentation and tune uncertainties (yellow and green) are of the same order as statistical uncertainties (upper row).

9.2.2 Comparison

We show the comparison of POWHEG +PYTHIA 8 (continuous cyan curve) with theory uncertainties (dashed cyan curves) to the measurement in Figs. 9.9-9.14 for the inclusive jet and inclusive b jet productions.

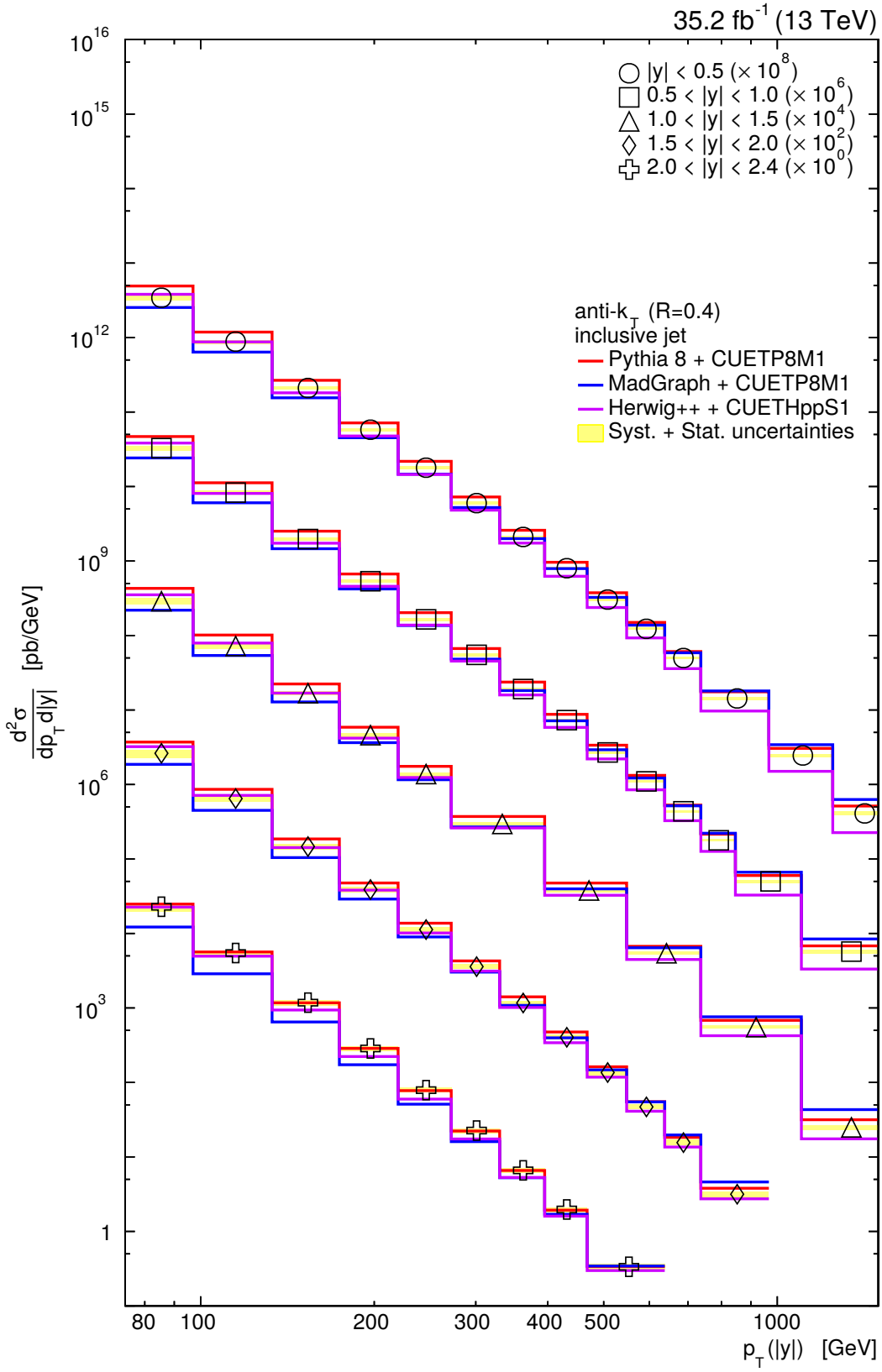


FIGURE 9.1 – Double differential cross section of inclusive jet production for the measurement and for LO simulation obtained with PYTHIA 8+`CUETP8M1`, MADGRAPH +`CUETP8M1`, and HERWIG++ +`CUETHppS1`. The markers correspond to the data; the continuous lines to the MC predictions; the yellow band to the uncertainty. The different rapidity bins have been rescaled with various factors so as no to overlap.

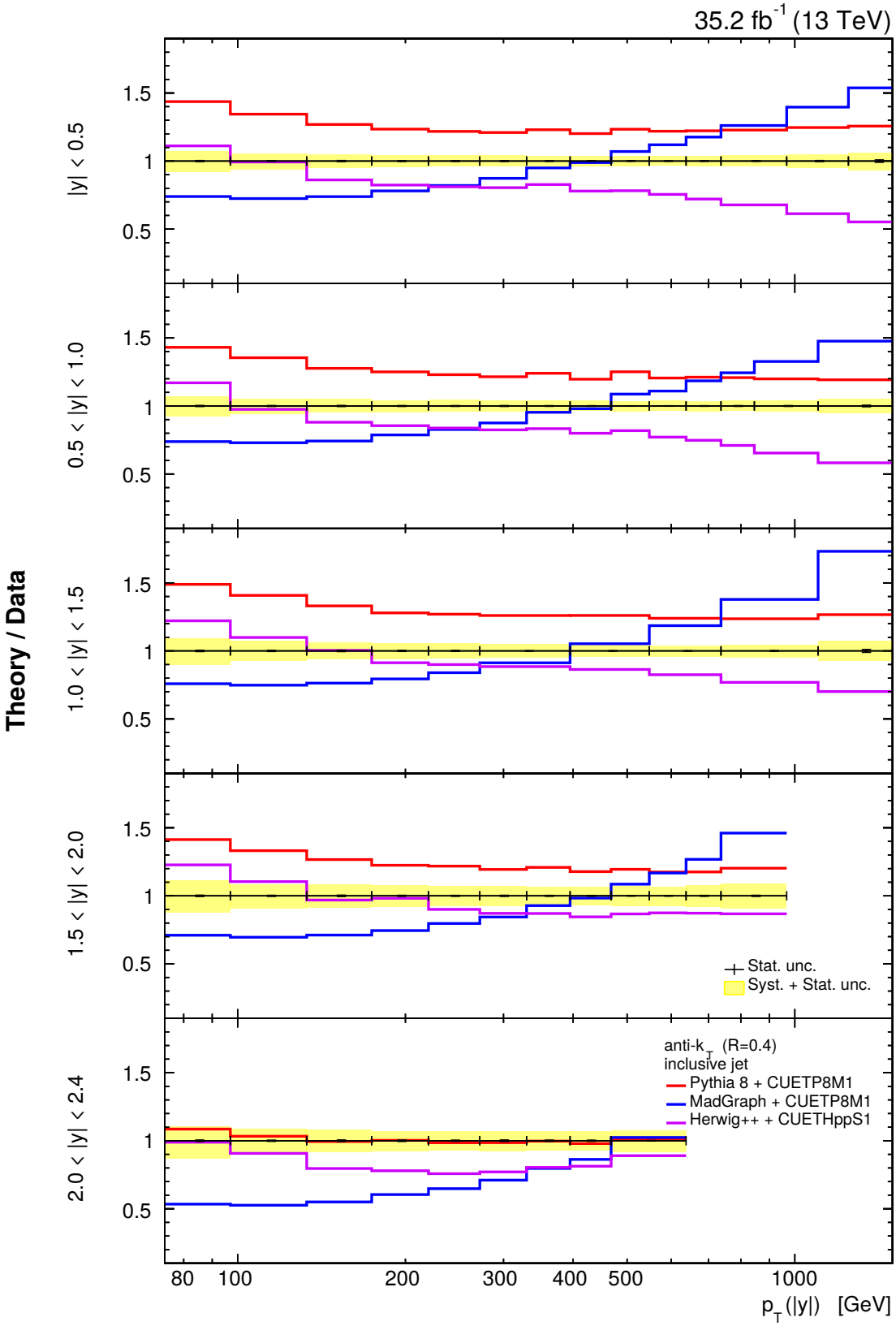


FIGURE 9.2 – Ratio of the double differential cross section of inclusive jet production of the measurement with LO simulation obtained with PYTHIA 8+`CUETP8M1`, MADGRAPH +`CUETP8M1`, and HERWIG++ +`CUETHppS1`. The data is at one; the continuous lines correspond to the MC simulations and the yellow band to the total uncertainty. The statistical uncertainty is indicated with the ticks; the separation between the statistical uncertainties from the data and from the MC is indicated.

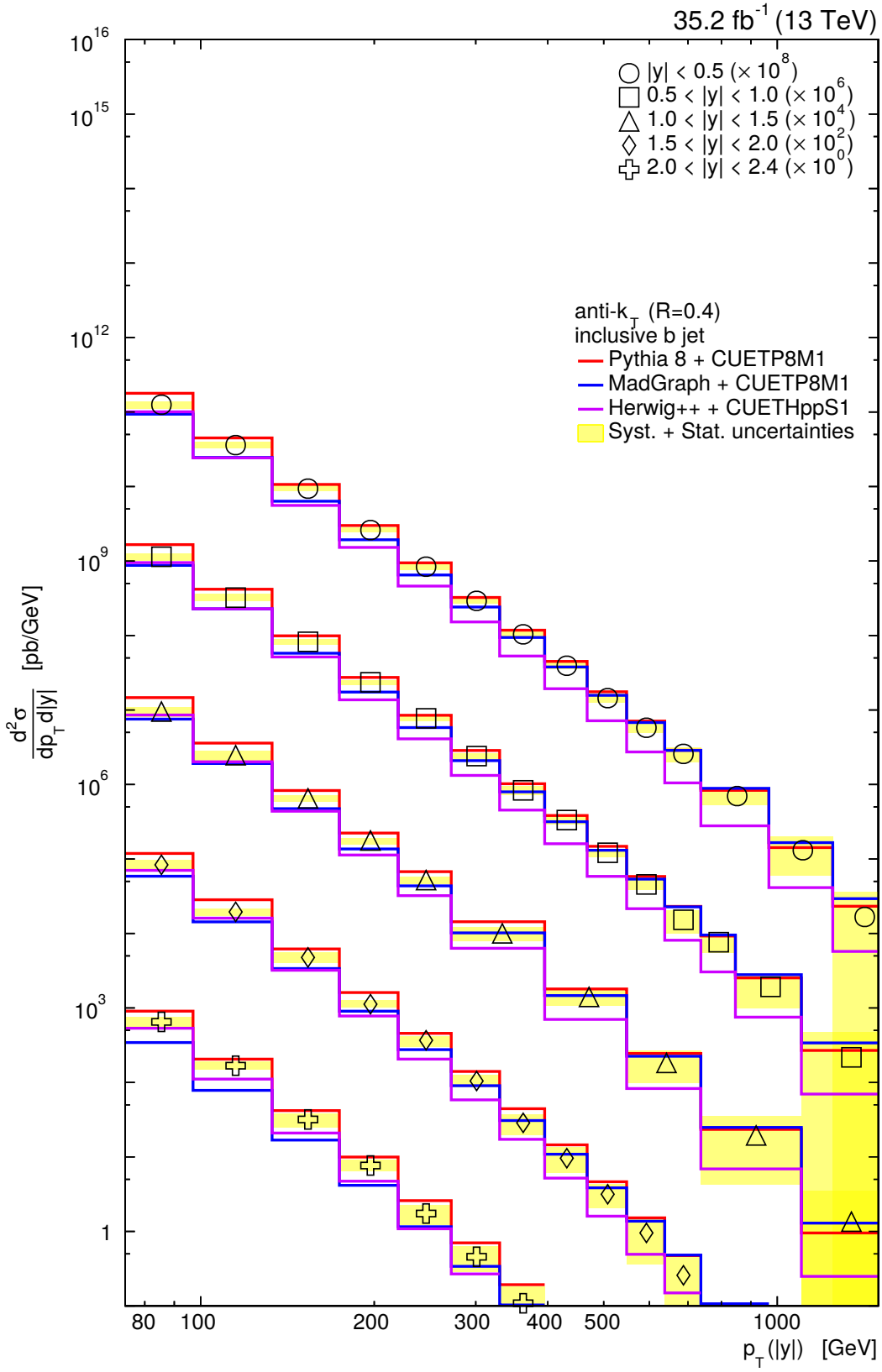


FIGURE 9.3 – Double differential cross section of inclusive b jet production for the measurement and for LO simulation obtained with PYTHIA 8+`CUETP8M1`, MADGRAPH +`CUETP8M1`, and HERWIG++ +`CUETHppS1`. The markers correspond to the data; the continuous lines to the MC predictions; the yellow band to the uncertainty. The different rapidity bins have been rescaled with various factors so as no to overlap.

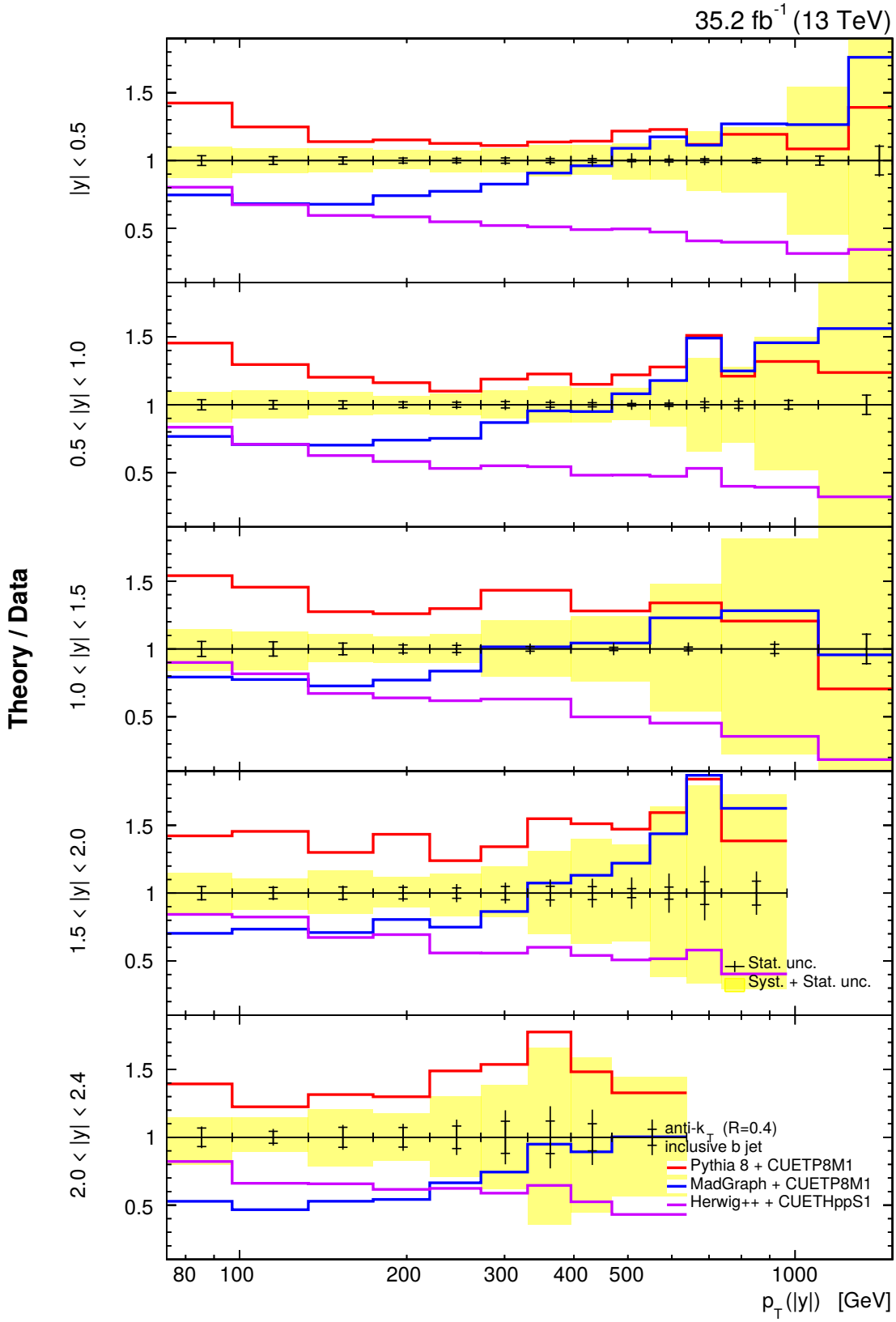


FIGURE 9.4 – Ratio of the double differential cross section of inclusive b jet production of the measurement with LO simulation obtained with PYTHIA 8+`CUETP8M1`, MADGRAPH +`CUETP8M1`, and HERWIG++ +`CUETHppS1`. The data is at one; the continuous lines correspond to the MC simulations and the yellow band to the total uncertainty. The statistical uncertainty is indicated with the ticks; the separation between the statistical uncertainties from the data and from the MC is indicated.

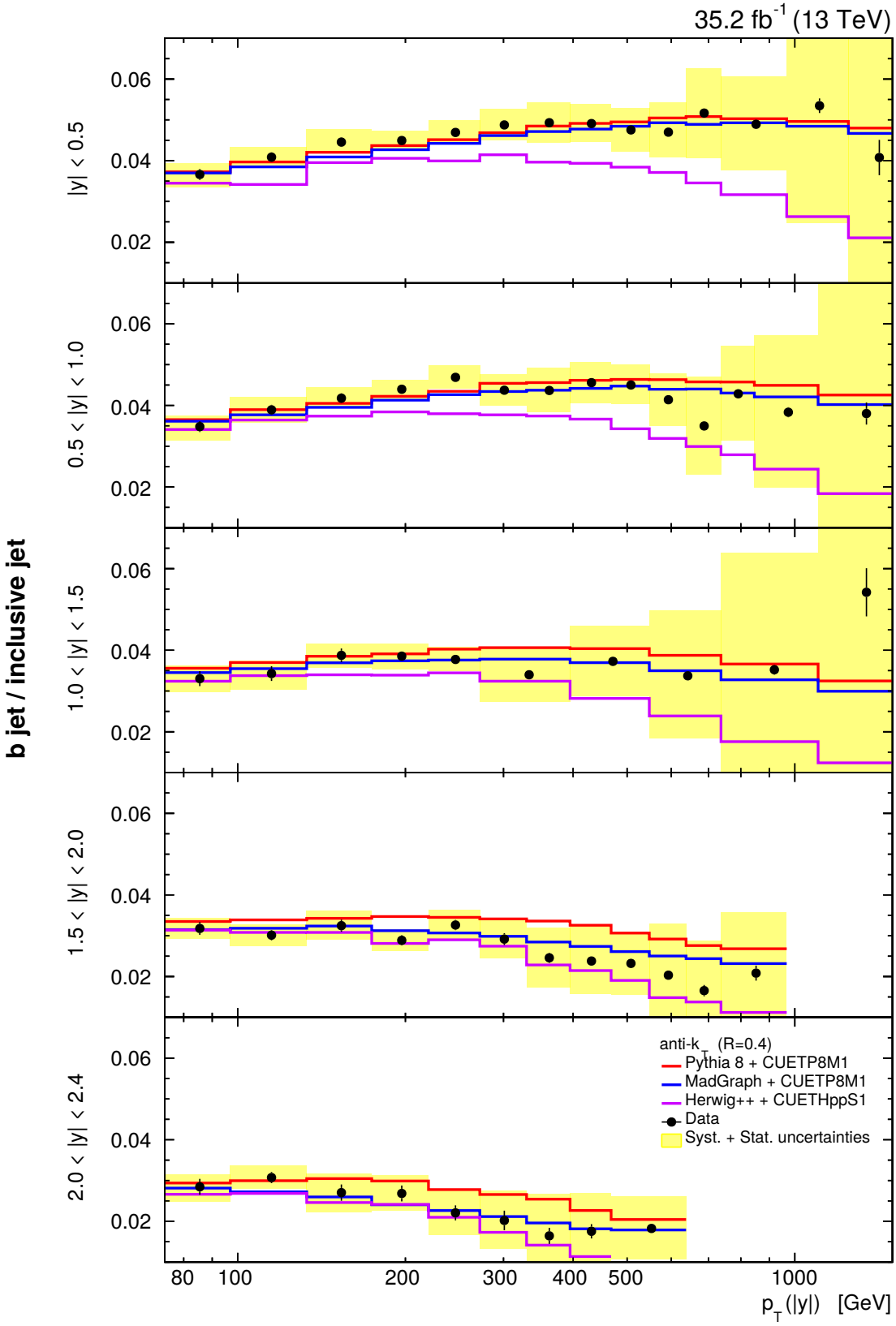


FIGURE 9.5 – Fraction of b jet production in inclusive jet production in the measurement and of LO simulation obtained with PYTHIA 8+`CUETP8M1`, MADGRAPH +`CUETP8M1`, and HERWIG++ +`CUETHppS1`. The markers correspond to the data; the continuous lines to the MC predictions; the yellow band to the uncertainty. The different rapidity bins have been rescaled with various factors so as no to overlap.

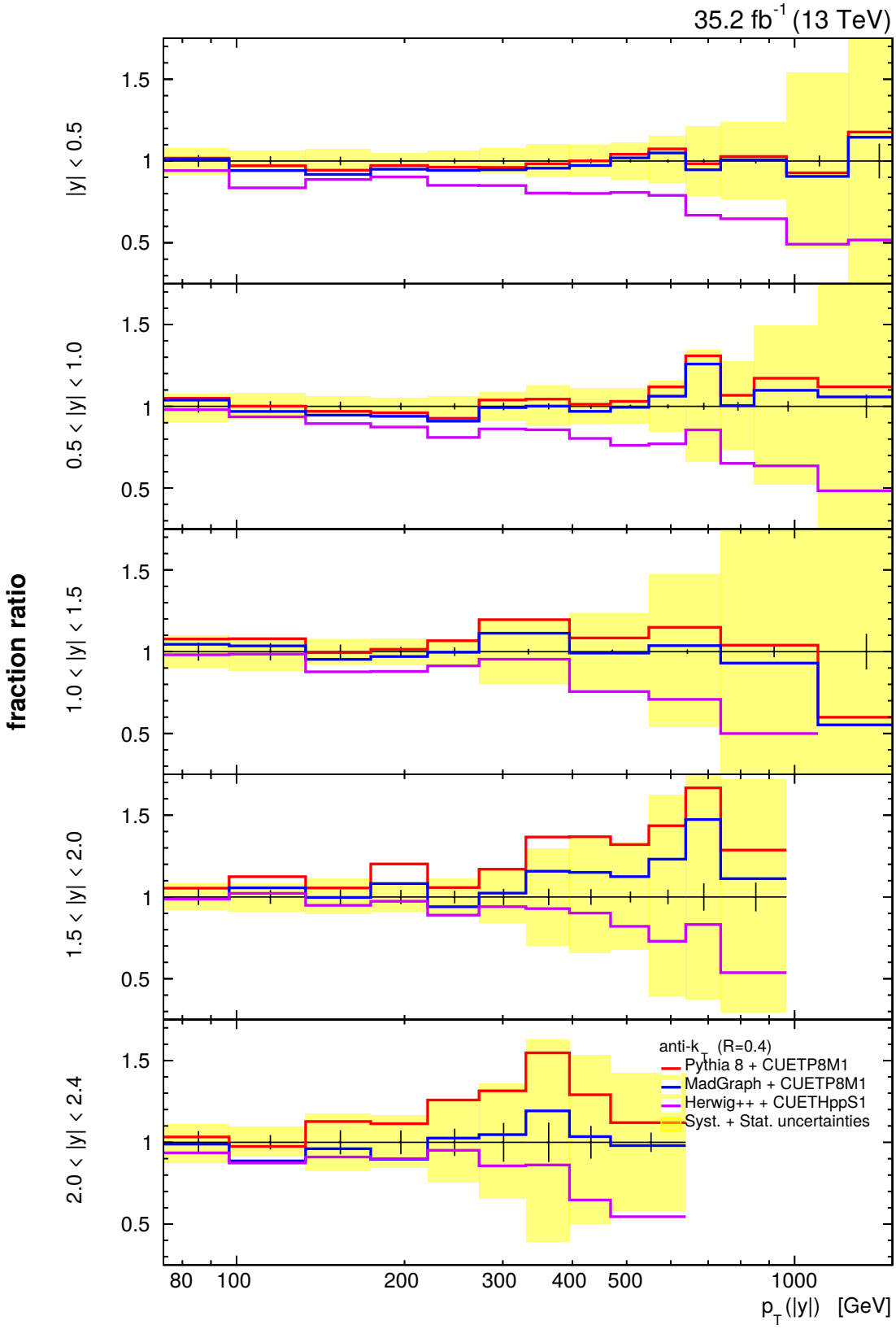


FIGURE 9.6 – Fraction ratio of the b jet production in inclusive jet production in the measurement with LO simulation obtained with PYTHIA 8+`CUETP8M1`, MADGRAPH +`CUETP8M1`, and HERWIG++ +`CUETHppS1`. The data is at one; the continuous lines correspond to the MC simulations and the yellow band to the total uncertainty. The statistical uncertainty is indicated with the ticks.

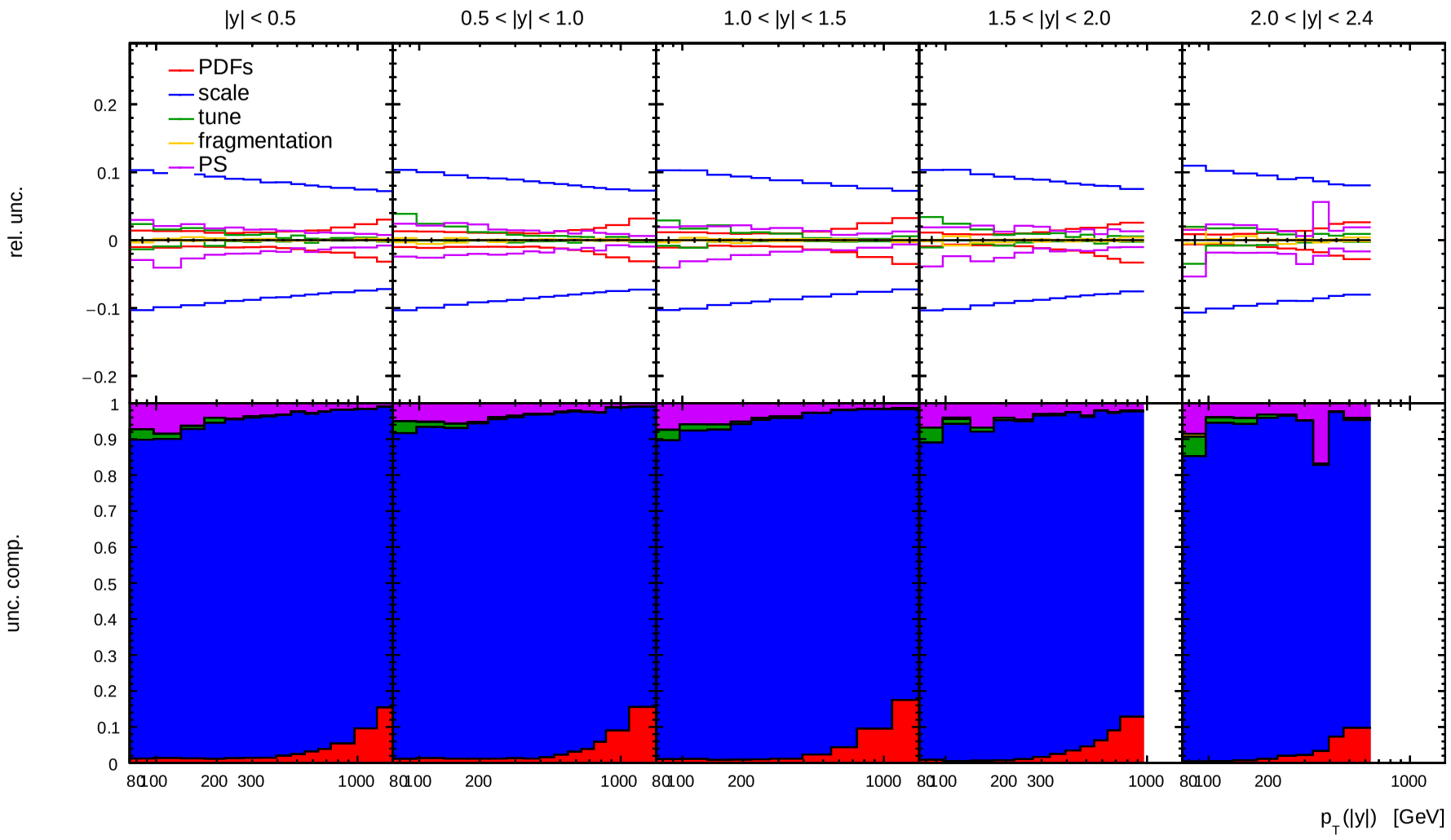


FIGURE 9.7 – Composition of the theory uncertainties for the POWHEG + `CUETP8M1` prediction of the inclusive jet production. The uncertainties are stacked and the sum is normalised to unity, bin after bin. Above (below) is the relative uncertainties (is the stack of the relative uncertainties normalised to one bin by bin). The five bins of rapidity are shown in columns.

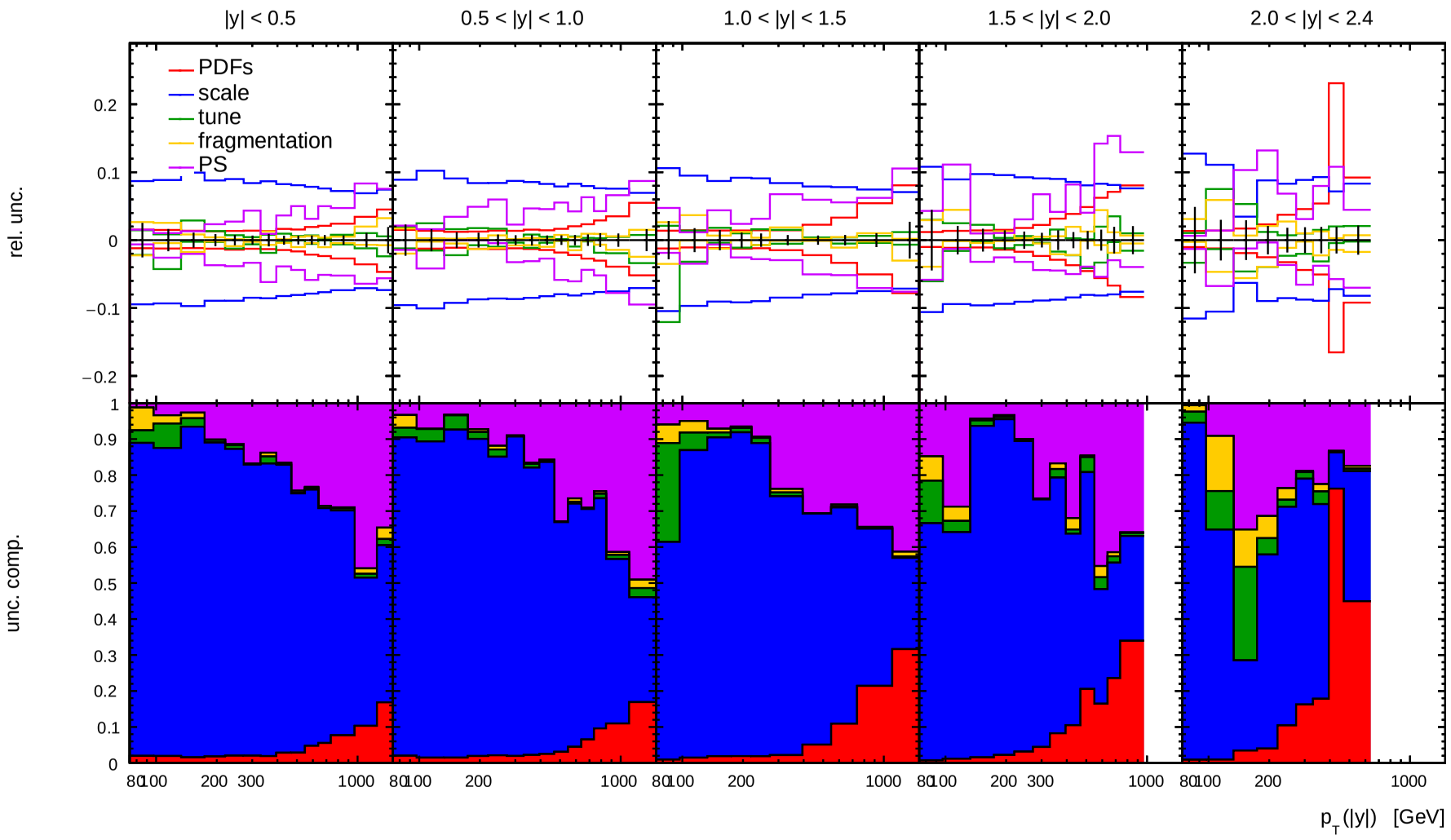


FIGURE 9.8 – Composition of the theory uncertainties for the POWHEG + `CUETP8M1` prediction of the inclusive b jet production. The uncertainties are stacked and the sum is normalised to unity, bin after bin. Above (below) is the relative uncertainties (is the stack of the relative uncertainties normalised to one bin by bin). The five bins of rapidity are shown in columns.

	variable	nominal	up	down
MPI	MPI cut-off p_{T0}^{ref}	2.4024	1.8238	2.60468
	Exponent ϵ	0.25208	2.5208	0.25208
	Reference energy $\sqrt{s_0}$	1.6	3.2	1.5
colour reconnection	Free parameter R	1.80	7.60	4.20

TABLE 9.1 – Variations of the tune.

The figures are organised in the same way as in the first section: first, the inclusive jet cross section is presented in Figs. 9.9-9.10. The agreement of data and simulation is excellent; the data has smaller uncertainty bands than the predictions. Then, the inclusive b jet cross section is presented in Figs. 9.11-9.12. Data and simulation also agree within the uncertainties; fluctuations in data are likely due to the irregularities in the performance of the taggers.

The comparison to the fraction in the inclusive jet production is shown in Figs. 9.13-9.14. The theory agrees everywhere within the theoretical and experimental uncertainties; however, in this case, the theory uncertainties are smaller than the experimental uncertainties.

9.2.3 Contributions to the prediction

In Figs. 9.15-9.16, we investigate the different contributions to the predictions:

— hard process + PS + MPI + hadronisation

— hard process + PS + hadronisation

— hard process + hadronisation

The two figures correspond to the cases of the inclusive jet and inclusive b jet productions. The uncertainties correspond to the scale and PDF uncertainties (described in the next section).

From the figures, one concludes that MPI has mainly an effect at low p_T (from red to blue), both for jets and b jets. However, the PS has a different effect for jets and for b jets. The PS "unsmears" the spectrum, since high p_T partons are likely to radiate; however, the b jet spectrum is significantly reduced, since a significant fraction of the b quarks come from gluon splitting. This effect was already investigated in Chap. 6 (Figs. 6.1-6.2).

9.A List of tune parameters

`CUETP8M1` is a tune developed by and for CMS, based on the Monash 2013 tune for PYTHIA 8.1 [8].

The Monash tune gathers the parameters from e^+e^- measurements at LEP and SLD[2] (especially all parameters related to hadronisation and FSR), and parameters from pp and $p\bar{p}$ measurements at TEVATRON and LHC (FSR, MPI, etc.).

`CUETP8M1` essentially contains a retuning of MPI, which is expected to be more significant at higher energy in the centre-of-mass.

All parameters, as used in the POWHEG prediction, are shown in Tab. 9.2.

2. SLAC Large Detector

References

[1] Vardan Khachatryan et al. "Measurement of the double-differential inclusive jet cross section in proton – proton collisions at $\sqrt{s} = 13\,\text{TeV}$". In: *Eur. Phys. J.* C76.8 (2016), p. 451. DOI: 10.1140/epjc/s10052-016-4286-3. arXiv: 1605.04436 [hep-ex].

[2] Vardan Khachatryan et al. "Measurement and QCD analysis of double-differential inclusive jet cross sections in pp collisions at $\sqrt{s} = 8$ TeV and cross section ratios to 2.76 and 7 TeV". In: *JHEP* 03 (2017), p. 156. DOI: 10.1007/JHEP03(2017)156. arXiv: 1609.05331 [hep-ex].

[3] Serguei Chatrchyan et al. "Inclusive b-jet production in pp collisions at $\sqrt{s} = 7$ TeV". In: *JHEP* 04 (2012), p. 084. DOI: 10.1007/JHEP04(2012)084. arXiv: 1202.4617 [hep-ex].

[4] Georges Aad et al. "Measurement of the inclusive and dijet cross-sections of b^- jets in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector". In: *Eur. Phys. J.* C71 (2011), p. 1846. DOI: 10.1140/epjc/s10052-011-1846-4. arXiv: 1109.6833 [hep-ex].

[5] Simone Alioli et al. "Jet pair production in POWHEG". In: *JHEP* 1104 (2011), p. 081. DOI: 10.1007/JHEP04(2011)081. arXiv: 1012.3380 [hep-ph].

[6] Simone Alioli et al. "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX". In: *JHEP* 06 (2010), p. 043. DOI: 10.1007/JHEP06(2010)043. arXiv: 1002.2581 [hep-ph].

[7] P. A Nason. "Recent developments in POWHEG". In: *Proceedings, 9th International Symposium on Radiative Corrections: Applications of quantum field theory to phenomenology. (RADCOR 2009).* Vol. RADCOR2009. 2010, p. 018. arXiv: 1001.2747 [hep-ph].

[8] Peter Skands, Stefano Carrazza, and Juan Rojo. "Tuning PYTHIA 8.1: the Monash 2013 Tune". In: *Eur. Phys. J.* C74.8 (2014), p. 3024. DOI: 10.1140/epjc/s10052-014-3024-y. arXiv: 1404.5630 [hep-ph].

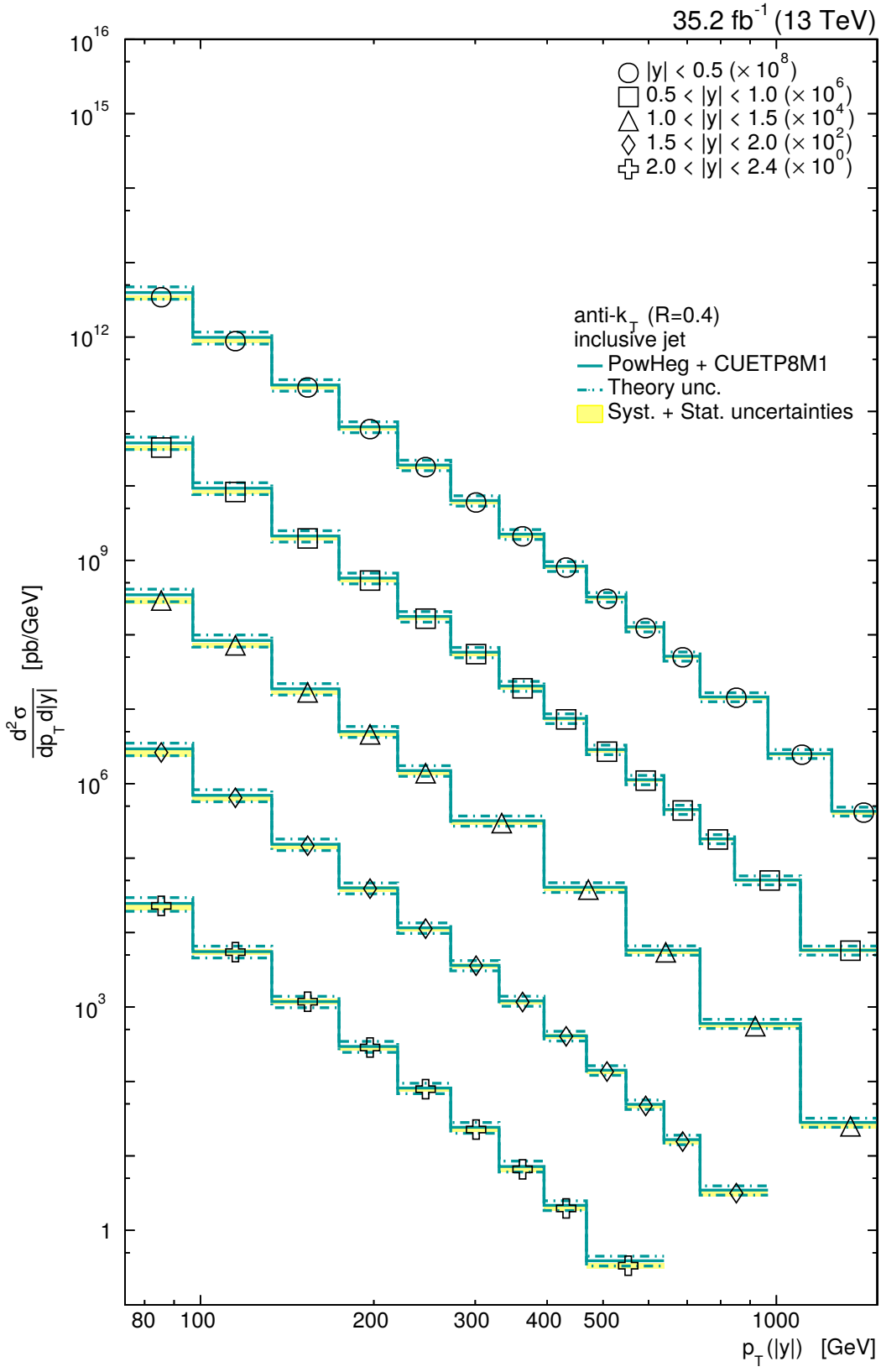


FIGURE 9.9 – Double differential cross section of inclusive jet production for the measurement and for NLO simulation obtained with POWHEG +`CUETP8M1`. The markers correspond to the data; the continuous lines to the MC predictions; the yellow band to the uncertainty. The different rapidity bins have been rescaled with various factors so as no to overlap.

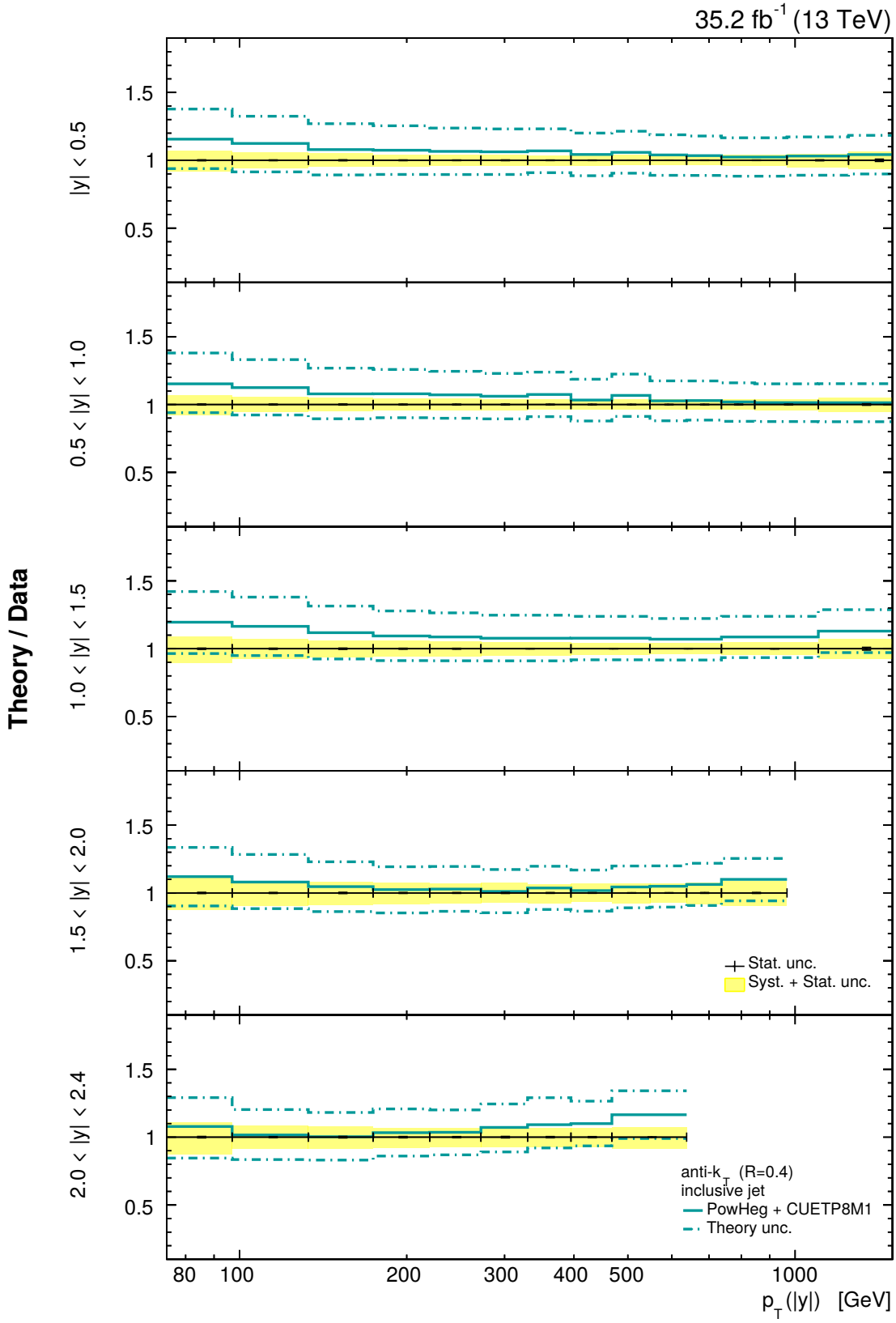


FIGURE 9.10 – Ratio of the double differential cross section of inclusive jet production of the measurement with NLO simulation obtained with POWHEG +`CUETP8M1`. The data is at one; the continuous lines correspond to the MC simulations and the yellow band to the total uncertainty. The statistical uncertainty is indicated with the ticks; the separation between the statistical uncertainties from the data and from the MC is indicated.

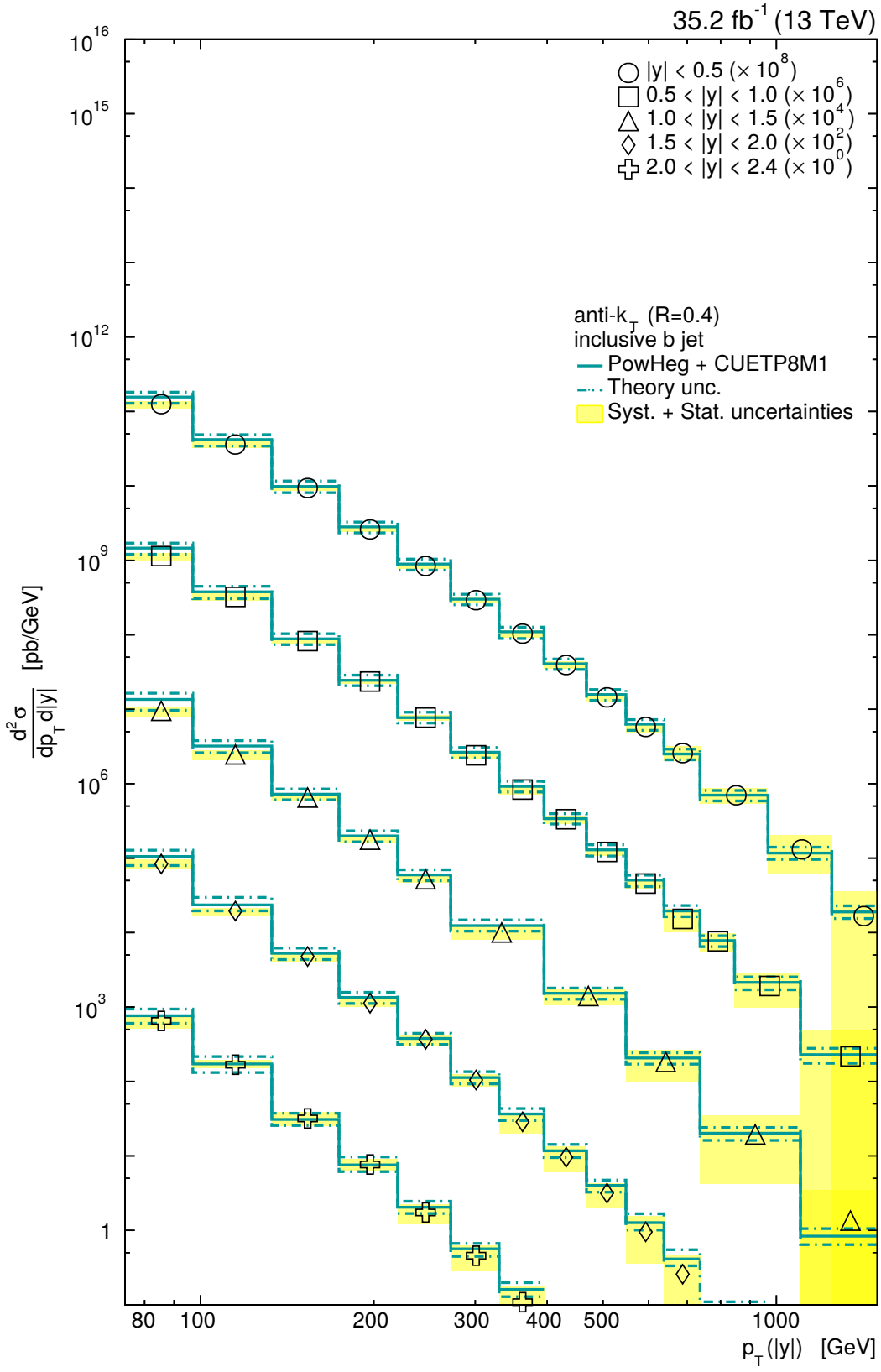


FIGURE 9.11 – Double differential cross section of inclusive b jet production for the measurement and for NLO simulation obtained with POWHEG +`CUETP8M1`. The markers correspond to the data; the continuous lines to the MC predictions; the yellow band to the uncertainty. The different rapidity bins have been rescaled with various factors so as no to overlap.

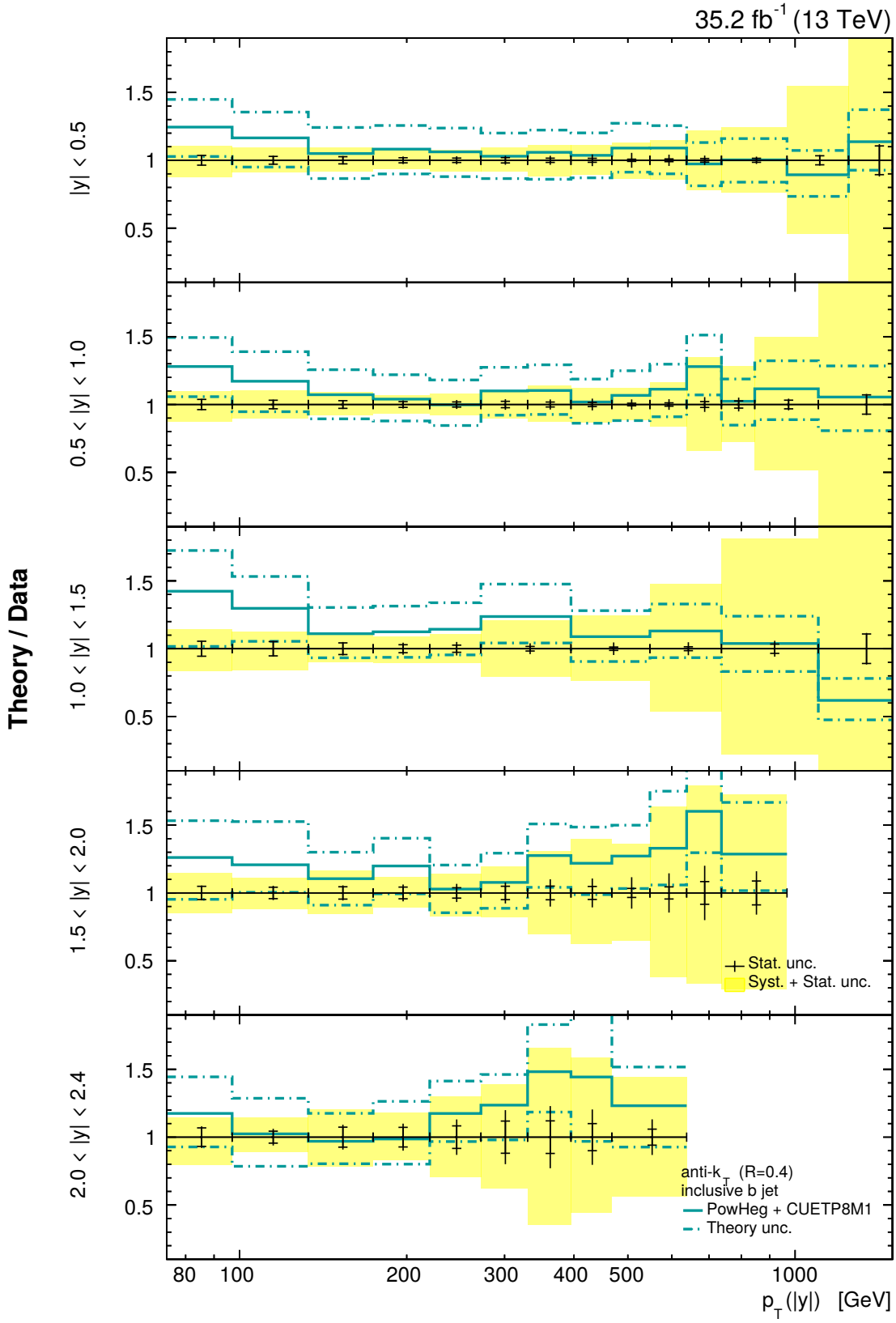


FIGURE 9.12 – Ratio of the double differential cross section of inclusive$b$ jet production of the measurement with NLO simulation obtained with POWHEG +`CUETP8M1`. The data is at one; the continuous lines correspond to the MC simulations and the yellow band to the total uncertainty. The statistical uncertainty is indicated with the ticks; the separation between the statistical uncertainties from the data and from the MC is indicated.

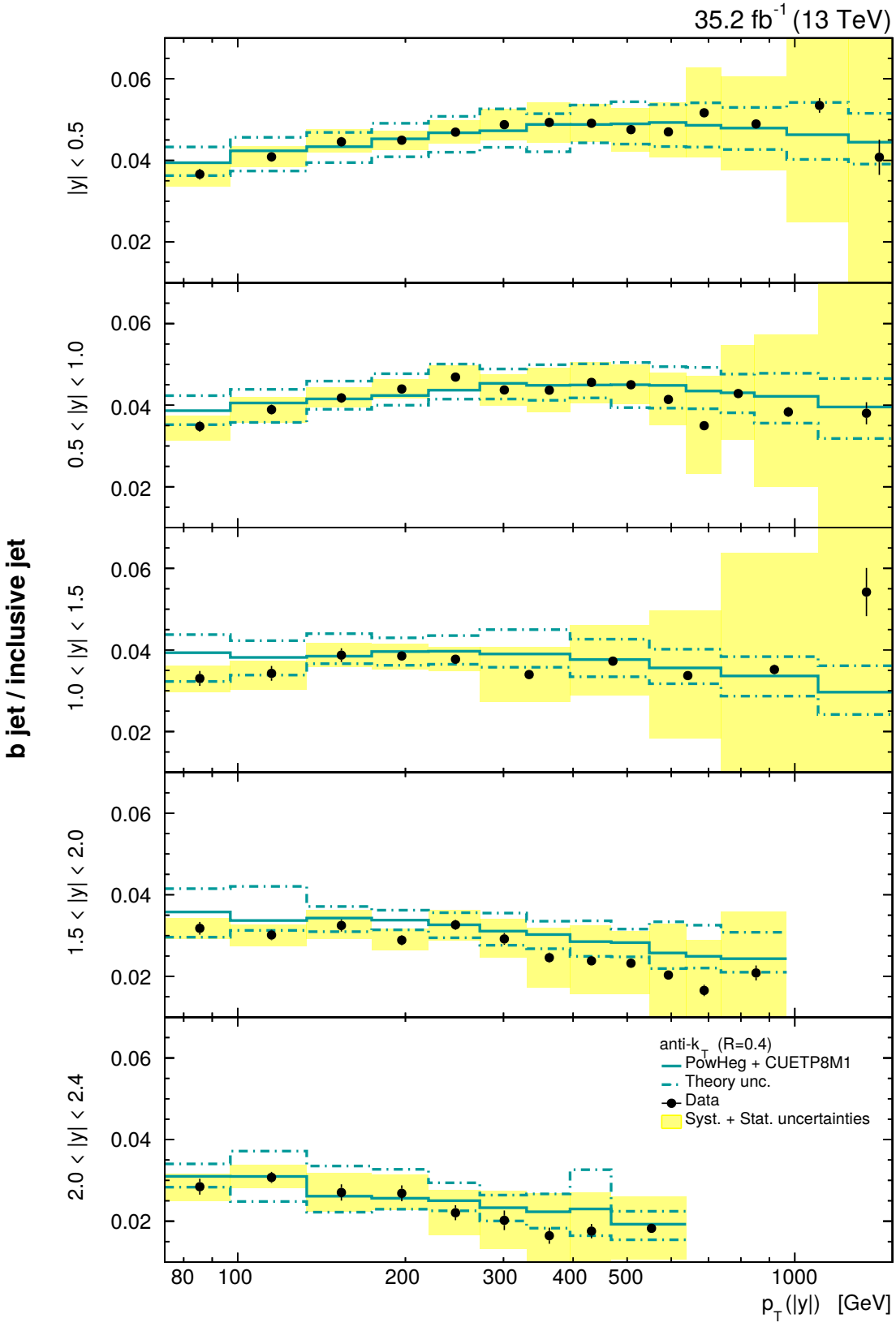


FIGURE 9.13 – Fraction of the b jet production in the inclusive jet production in the measurement and of NLO simulation obtained with POWHEG +`CUETP8M1`. The markers correspond to the data; the continuous lines to the MC predictions; the yellow band to the uncertainty. The different rapidity bins have been rescaled with various factors so as no to overlap.

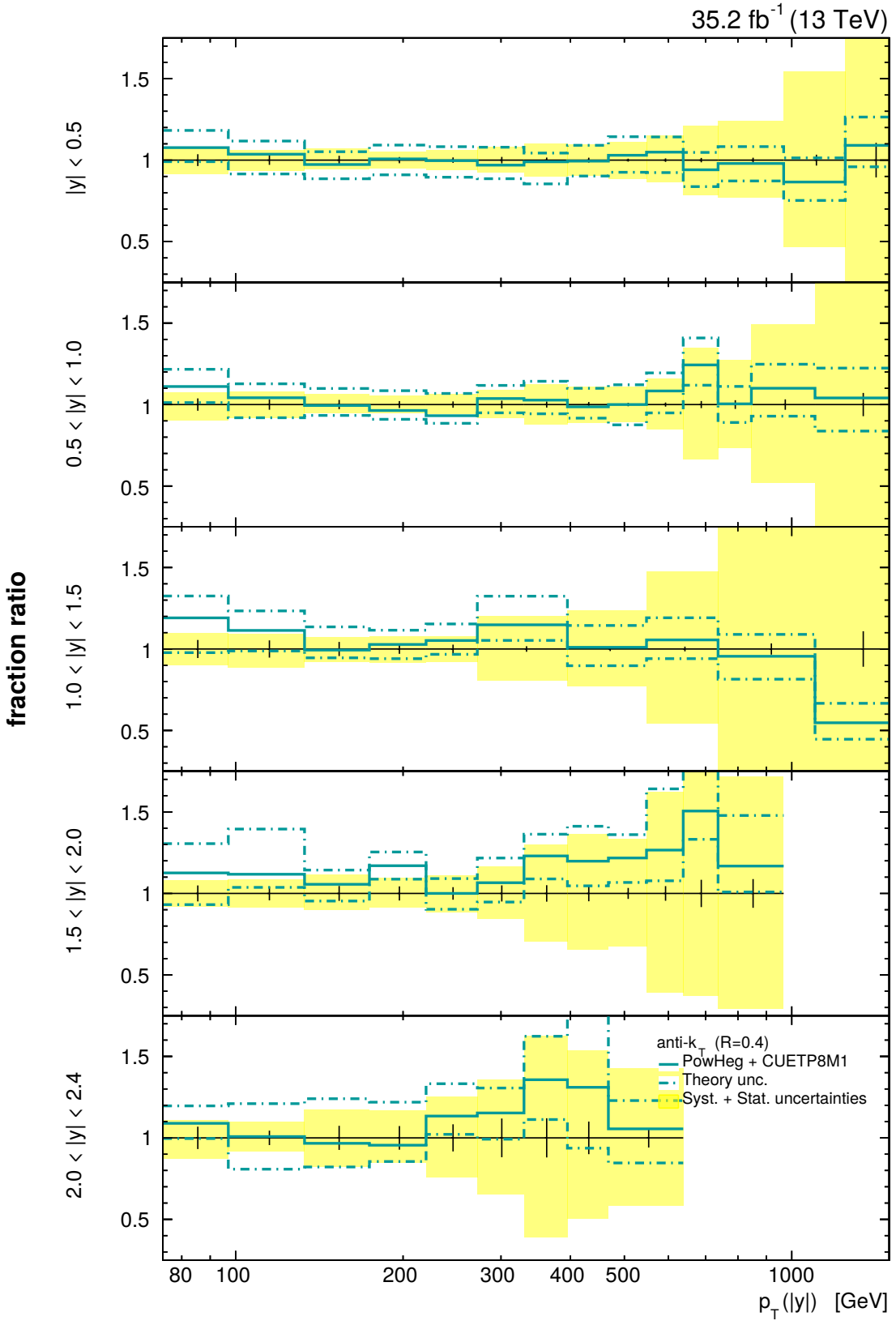


FIGURE 9.14 – Fraction ratio of the b jet production in the inclusive jet production in the measurement with NLO simulation obtained with POWHEG +`CUETP8M1`. The data is at one; the continuous lines correspond to the MC simulations and the yellow band to the total uncertainty. The statistical uncertainty is indicated with the ticks.

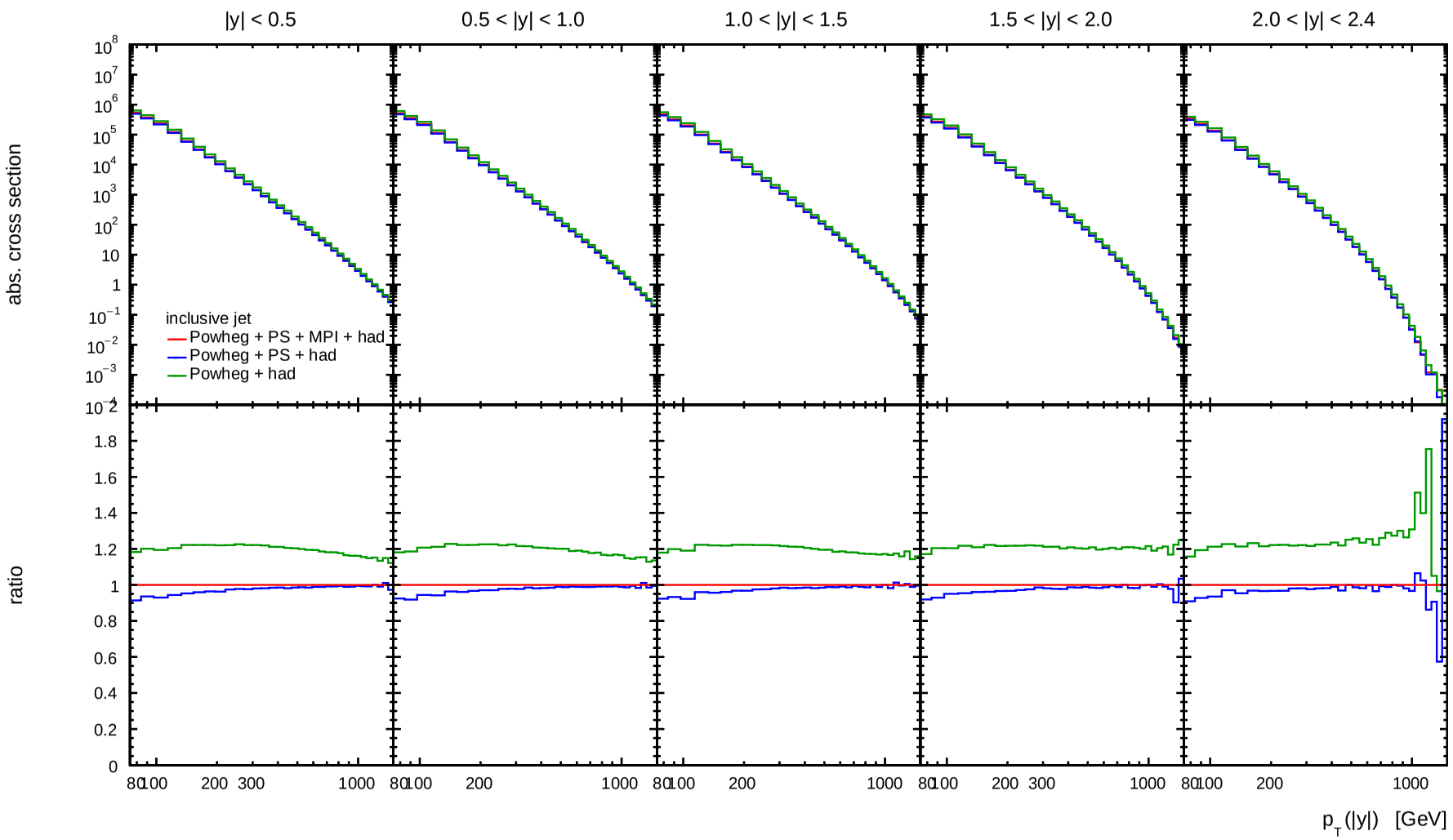


FIGURE 9.15 – Contributions to the prediction of the inclusive jet production by POWHEG + `CUETP8M1`. The columns correspond to the rapidity bins. Above (below) is the absolute cross section as predicted by POWHEG + `CUETP8M1` (the ratio to the prediction including all contributions).

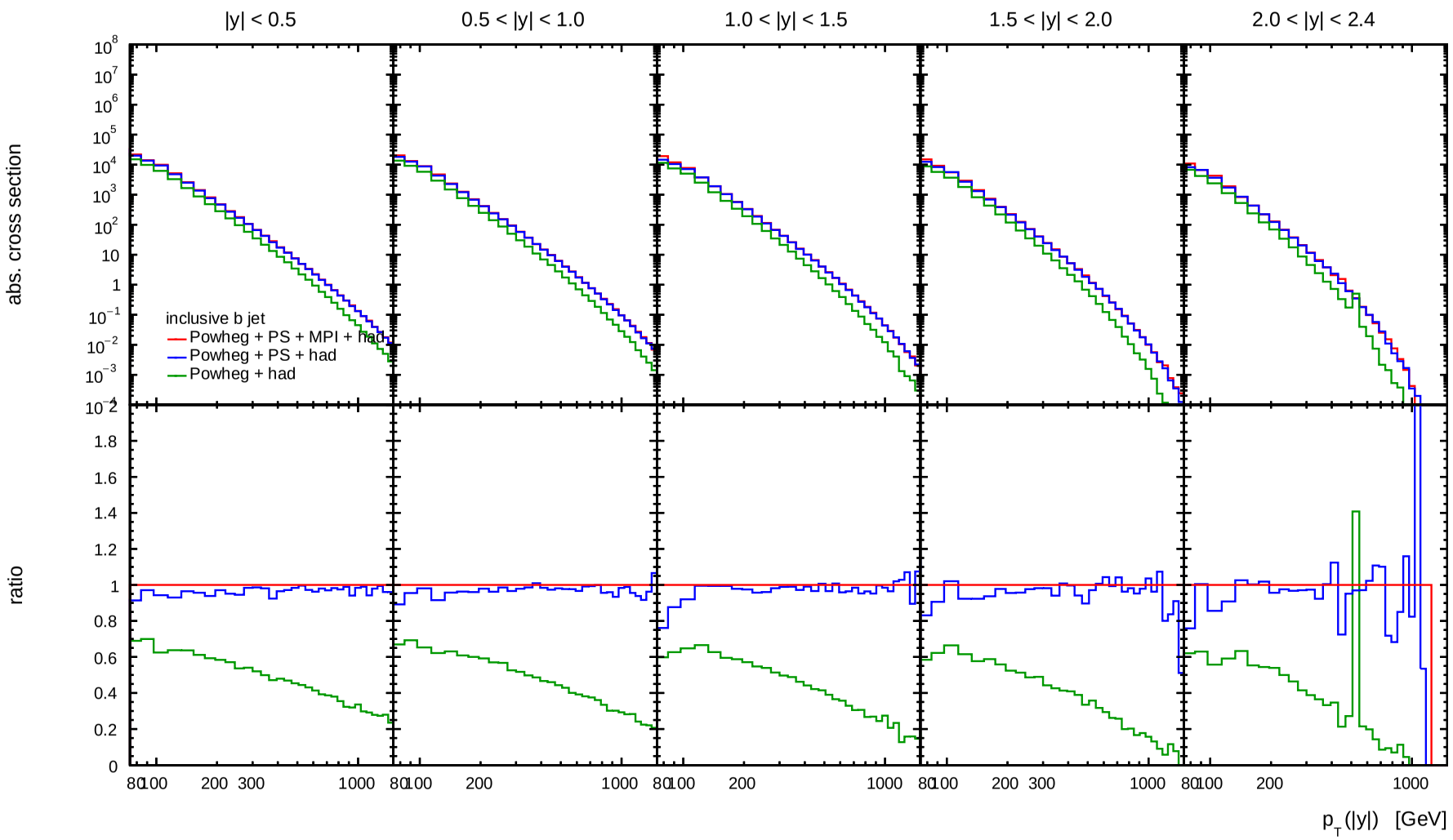


FIGURE 9.16 – Contributions to the prediction of the inclusive b jet production by POWHEG + `CUETP8M1`. The columns correspond to the rapidity bins. Above (below) is the absolute cross section as predicted by POWHEG + `CUETP8M1` (the ratio to the prediction including all contributions).

tune	parameter	value
ee tune 7	StringFlav:probStoUD	0.217
	StringFlav:probQQtoQ	0.081
	StringFlav:probSQtoQQ	0.915
	StringFlav:probQQ1toQQ0	0.0275
	StringFlav:mesonUDvector	0.50
	StringFlav:mesonSvector	0.55
	StringFlav:mesonCvector	0.88
	StringFlav:mesonBvector	2.20
	StringFlav:etaSup	0.60
	StringFlav:etaPrimeSup	0.12
	StringFlav:popcornSpair	0.90
	StringFlav:popcornSmeson	0.50
	StringFlav:suppressLeadingB	false
	StringZ:aLund	0.68
	StringZ:bLund	0.98
	StringZ:aExtraSquark	0.00
	StringZ:aExtraDiquark	0.97
	StringZ:rFactC	1.32
	StringPT:sigma	0.335
	StringPT:enhancedFraction	0.01
	StringPT:enhancedWidth	2.0
	TimeShower:alphaSvalue	0.1365
	TimeShower:alphaSorder	1
	TimeShower:alphaSuseCMW	false
	TimeShower:pTmin	0.5
	TimeShower:pTminChgQ	0.5
pp tune 14	PDF:pSet	NNPDF
	SigmaProcess:alphaSvalue	0.130
	SigmaTotal:zeroAXB	true
	SigmaDiffractive:dampen	true
	SigmaDiffractive:maxXB	65.0
	SigmaDiffractive:maxAX	65.0
	SigmaDiffractive:maxXX	65.0
	Diffraction:largeMassSuppress	4.0
	TimeShower:dampenBeamRecoil	true
	TimeShower:phiPolAsym	true
	SpaceShower:alphaSvalue	0.1365
	SpaceShower:alphaSorder	1
	SpaceShower:alphaSuseCMW	false
	SpaceShower:samePTasMPI	false
	SpaceShower:pT0Ref	2.0
	SpaceShower:ecmRef	7000.0
	SpaceShower:ecmPow	0.0
	SpaceShower:pTmaxFudge	1.0
	SpaceShower:pTdampFudge	1.0
	SpaceShower:rapidityOrder	true
	SpaceShower:rapidityOrderMPI	true
	SpaceShower:phiPolAsym	true
	SpaceShower:phiIntAsym	true
	MultipartonInteractions:alphaSvalue	0.130
	MultipartonInteractions:bProfile	3
	MultipartonInteractions:expPow	1.85
	MultipartonInteractions:a1	0.15
	BeamRemnants:primordialKTsoft	0.9
	BeamRemnants:primordialKThard	1.8
	BeamRemnants:halfScaleForKT	1.5
	BeamRemnants:halfMassForKT	1.0
	ColourReconnection:mode	0
	ColourReconnection:range	1.80
CUETP8M1	MultipartonInteractions:pT0Ref	2.4024
	MultipartonInteractions:ecmPow	0.25208
	MultipartonInteractions:expPow	1.6
	StringZ:rFactB	0.895

TABLE 9.2 – Parameters of the tune for UE in PYTHIA 8.

Part III

Conclusions

10

Summary, Conclusions and Perspectives

The measurement that has been presented in this thesis is a textbook measurement. It is one of the most important observables in proton-proton collisions.

In this chapter, we summarise the analysis. In addition, we discuss additional results that can be obtained thanks to the techniques explained in this thesis. In this section, limitations and successes of these techniques are discussed and prospects for future measurements are given.

10.1 Summary

In this thesis, we have presented a new measurement of the inclusive b jet production at $\sqrt{s} = 13\,\text{TeV}$.

After a long introduction to situate the context of the measurement, we showed how to obtain the signal with the CMS experiment, and we corrected it to parton level. For this, we applied an advanced method of unfolding to disentangle the b jets from the light and c jets. Then we compared it to predictions at LO and at NLO; in addition, we also measured and compared the fraction of b jets in the inclusive jet production.

We also investigated the mechanisms of productions of b jets. The contribution from PS dominate at high p_T.

The cross section is measured over six orders of magnitude, with a large coverage of the phase space. The transverse momentum goes from 100 GeV to 1 TeV and the rapidity from -2.4 to 2.4. The agreement with the theory is a great success.

Its validation at LHC opens up many new measurements. For instance, the improvement of the modelling of $b\bar{b}$ background could lead to the measurement the Higgs boson in the $H \to b\bar{b}$ channel.

10.2 Perspectives

Many issues still need to be solved in QCD. As we saw, despite the numerous successes of the SM, difficulties remain in the description of the MPI, or in general, of most non-perturbative aspects.

We review here some perspectives of new measurements and new techniques of measurement with b jets and in which circumstances they may be of interest.

10.2.1 New measurements with Run-II data

We discuss here HF measurements that can be done with $\sqrt{s} = 13\,\text{TeV}$ data at LHC.

10.2.1.1 Cross section for $b\bar{b}$ production

The measurement of $b\bar{b}$ is the natural next step, after the measurement of inclusive b jet production.

First, we can measure the mass spectrum of the pair, and investigate evolution effects by reconstructing extra radiations.

Then, we can measure the triple differential cross section of the $b\bar{b}$ dijet system as a function of the average transverse momentum, of the rapidity boost and of the rapidity separation. A similar measurement was performed at 8 TeV at CMS, as discussed in App. C. The interest of this measurement would be to constrain PDFs; it is also the textbook measurement to describe the ME of production of $b\bar{b}$.

In addition, we can study the azimuthal correlations among b jets, and also with additional jet production. While the inclusive b jet measurement is more sensitive to the ME, the azimuthal correlations with additional radiations having significant p_T is more sensitive to the UE. One could study effect more related to the PS and MPI.

10.2.1.2 Inclusive c jet cross section

The inclusive c jet cross section would be another good test of the flavour democracy of QCD. A similar measurement has never been done at LHC and would be a premiere.

However, although it looks similar to the inclusive b jet measurement, as was done in this thesis, it is slightly more complicated. Indeed, a c tagger consists of a double tagger: an "anti-light" tagger plus an "anti-b" tagger. At the time of writing the thesis, recent developments have been achieved regarding c tagging [1]. While with the first versions of the tools, the analysis could not be reproduced with c jets, the performance is significantly improving.

The strategy to measure a c jet cross section would be to disentangle simultaneously light, c and b jets For this, we can use the double tagger or `DeepCSV`, and apply similar techniques as what has been described in this thesis, extending the matrix in Eq. 8.2 to a 3×3 matrix.

10.2.2 Techniques of measurements

The use of Deep Learning techniques is getting more and more widespread in the context of b tagging. The successor of `DeepCSV`, including more observables and more sophisticated, is already under development and will be available for 2017 and 2018 data, called `DeepFlavour`.

In addition, a novel technique of tagging for silicon detectors has recently been published and would be particularly adapted to CMS [2]. Currently, as we described in this volume, the tagging techniques are mostly based on tracking and vertexing. In this new method, instead of losing efficiency at tracking and vertexing to perform b tagging, patterns of hits in the tracker are directly investigated. This technique only starts to be efficient for

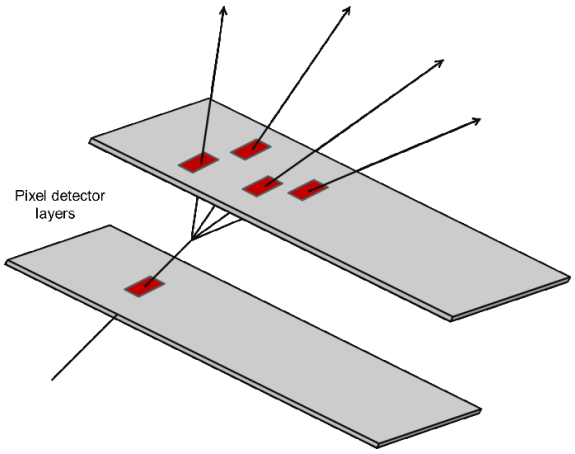


FIGURE 10.1 – Sketch of the decay of a B hadron between to layers of the tracker. The gray planes represents silicon modules; the arrows represent tracks; the hits represent hits. Taken from [2].

boosted B's, and should be included in the techniques of Deep Learning used in current taggers. This could significantly improve the purity at high p_T.

Moreover, in the coming decade, the CMS tracker should extended down to $y = 4.0$; with, in addition, the ever increasing pile-up conditions, this technique could significantly help to perform b tagging in an environment where tracks are not evident to reconstruct.

10.2.3 Prospects at longer term

At longer term, the LHC will undergo substantial upgrades in order to deliver higher luminosities [3], aiming at an integrated luminosity of $3000\,\text{fb}^{-1}$. Discussions are also ongoing, whether LHC could be upgraded for higher energies in the centre-of-mass system, with $\sqrt{s} \approx 33\,\text{TeV}$. This implies that the statistics at the TeV scale will become significantly larger, and that multi-differential precision measurements will be possible in this region of the phase space.

In the context of b jets, this has several implications: first, it is important to pursue effort in b tagging techniques. Second, as extra radiations are extremely important at high p_T, the modelling of PS has to be refined further, including additional orders.

But the most exciting comes in terms of physics. It will also become possible to test the flavour democracy by comparing bottom and top productions, since the scale will be much larger than the mass of the top quark. Indeed, any deviations from the flavour democracy might be a hint to new physics.

References

[1] Albert M Sirunyan et al. "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV". In: (2017). arXiv: 1712.07158 [physics.ins-det].

[2] B. Todd Huffman, Charles Jackson, and Jeff Tseng. "Tagging b quarks at extreme energies without tracks". In: *J. Phys.* G43.8 (2016), p. 085001. DOI: 10.1088/0954-3899/43/8/085001. arXiv: 1604.05036 [hep-ex].

[3] Apollinari G. et al. *High-Luminosity Large Hadron Collider (HL-LHC): Technical Design Report V. 0.1.* CERN Yellow Reports: Monographs. Geneva: CERN, 2017. URL: https://cds.cern.ch/record/2284929.

Appendices

A

Tracker Alignment

As described previously in Chap. 3, the CMS tracker consists of several superimposed layers of silicon modules (as of 2016). In order to achieve reconstruction of tracks with optimal precision despite the finite fabrication tolerances of the large structures, the modules of the tracking system need to be *aligned*. Corrections for the position, orientation and curvature need to be computed for every single sensor, possibly changing with time.

The strategy to align the modules is based on the reconstruction of tracks. One technique consists in determining the parameters of the modules and of the tracks simultaneously, involving up to $O(10^6)$ parameters.

In the context of this thesis, it is important to mention the alignment is crucial for the good performance of b tagging, since it relies on the reconstruction of tracks and SVs.

In this chapter, we explain how to solve the challenging problem of alignment with the track-based approach. We first describe the general strategy adopted at CMS (Sect. A.1); then we present the *Legacy Alignment* of the 2016 data (Sect. A.2).

A significant part of the work spent at DESY for this thesis and for the CMS collaboration was devoted to the alignment of the tracker. This chapter is the written version of a talk given at the TIPP 2017 conference in Beijing.

A.1 Introduction

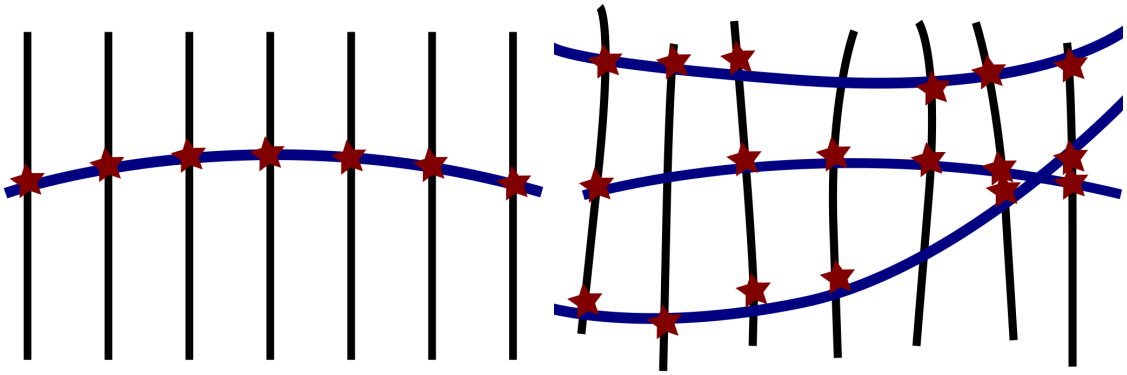
The purpose of a silicon tracker is to reconstruct the tracks from their hits in the successively traversed modules; the basic principles are illustrated in Fig. A.1a. At CMS, the algorithm for the track reconstruction is the *Kalman filter*, as described in Sect. 3.2.2.1.

At the mounting of the tracker, the precision of the mechanical alignment is typically of $\mathcal{O}(1\,\text{mm})$. At this stage, the uncertainty on the alignment is much larger than the uncertainty on the resolution of reconstruction of the hit:

$$\sigma_{\text{align}} \gg \sigma_{\text{hit}} \tag{A.1}$$

While a random misalignment only degrades the accuracy of the measurement, a systematic misalignment may lead to biased physics results. Therefore, it is important to improve the alignment to a resolution of the same order as the hit resolution:

$$\sigma_{\text{align}} \approx \sigma_{\text{hit}} \tag{A.2}$$



(A) Ideal picture. (B) Realistic picture.

FIGURE A.1 – Track reconstruction in the context of alignment. The black, straight lines represent the silicon modules, seen transversally; the dark blue, curved line represents tracks; the red stars represent the track hits.

Acronym	Full name	substructures	# layers/discs
PXB	PiXel Barrel	2 half cylinders	3
PXF	PiXel Forward	2×2 half disks	2
TIB	Tracker Inner Barrel	2 half barrels	4
TOB	Tracker Outer Barrel	2 half barrels	6
TID	Tracker Inner Disk	2 full discs	3
TEC	Tracker End-Caps	2 full discs	9

TABLE A.1 – Structure of the CMS tracker and characteristics of the mechanical structures. (Complementary diagram in Fig. A.2).

This is achieved by applying a correction in the track reconstruction. This procedure will be described in this chapter, as well as data-driven methods to validate the alignment.

A.1.1 Tracker alignment at CMS

The subdivision of the CMS tracker into mechanical structures is described in Tab. A.1 and Fig. A.2. The tracker can be aligned at different levels of precision:

1. large mechanical structures,
2. layers and discs,
3. ladders and blades,
4. or sensors.

These objects are called *alignables*. Typically, the alignment of the large mechanical structures (sensors) corresponds to corrections of $\mathcal{O}(1\,\text{mm})$ ($\mathcal{O}(10\ \mu\text{m})$). The degree of precision, *i.e.* how many alignables are considered, is related to size of the sample of tracks that will be used to perform the alignment.

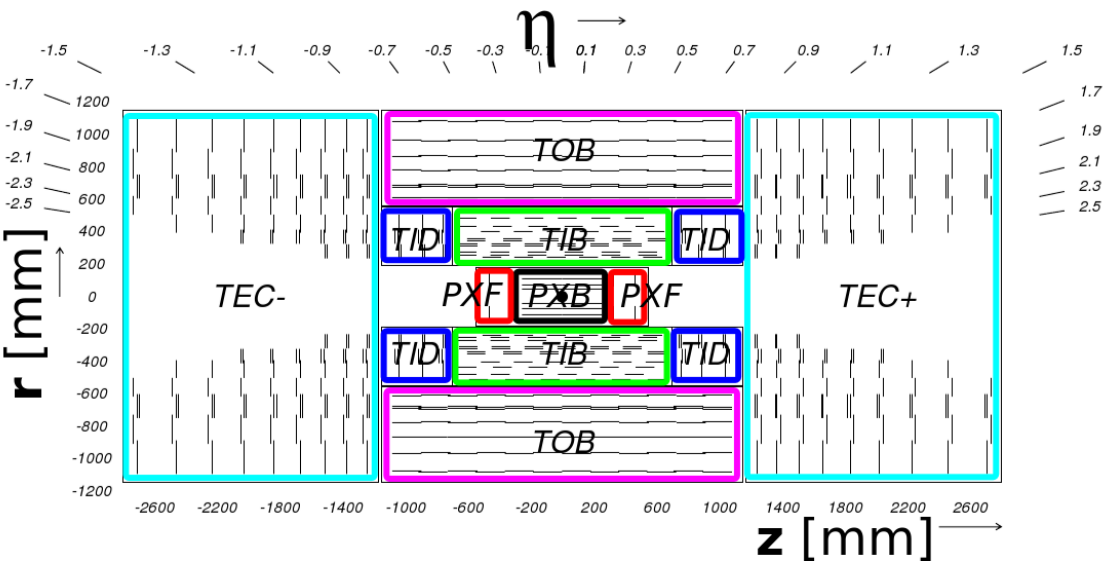


FIGURE A.2 – Structure of the tracker as of 2016. The pixel (strip) tracker is the innermost (outermost) part. The pixel tracker is made of two layers in the forward region (`PXF`) and of **three layers in the barrel region** (`PXB`); the strip tracker is made of four layers in the inner barrel (`TIB`), of three layers in the inner disk (`TID`) on each side, of six layers in the outer barrel (`TOB`) and of nine layers in the end-caps (`TEC`) on each side. Figure modified from [1].

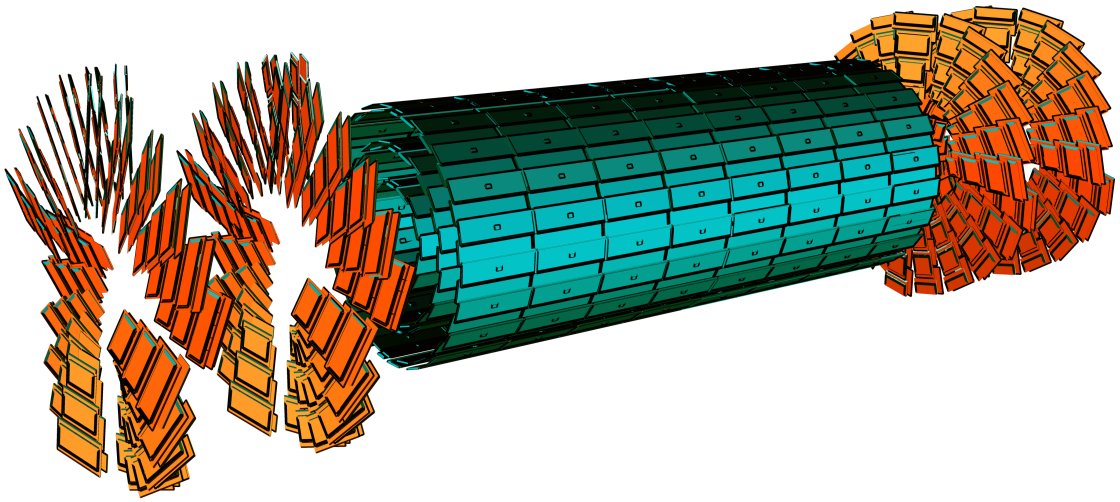


FIGURE A.3 – The ladders (barrel region) and the blades (end-caps) in the pixel tracker [2].

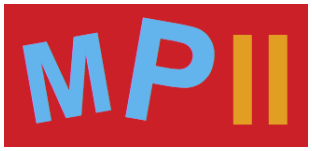


FIGURE A.4 – MILLEPEDE-II logo.

A.1.2 Track-based approach

The approach chosen at CMS is the track-based approach [1]. It consists in a least-square minimisation of the following fit [4, 5]:

$$\chi^2(\mathbf{p},\mathbf{q}) = \sum_{j}^{\text{tracks}} \sum_{i}^{\text{hits}} \left(\frac{m_{ij} - f_{ij}(\mathbf{p},\mathbf{q}_j)}{\sigma_{ij}} \right)^2 \tag{A.3}$$

where

- $\mathbf{p}$ stands for the alignment parameters and $\mathbf{q}$ for the track parameters,
- $\mathbf{m}$ stands for the measurements and $\mathbf{f}$ for the predictions,
- and σ stands for the uncertainties from the measurement.

The difficulty consists in minimising this χ^2 for a number of parameters of $O(10^5)$.

At CMS, two algorithms are used: MILLEPEDE-II [6] and HIPPY (previously HIP) [7]. Both are based on a linearisation of the χ^2:

$$\chi^2(\mathbf{p}_0+\Delta\mathbf{p},\mathbf{q}_0+\Delta\mathbf{q}) = \sum_{j}^{\text{tracks}} \sum_{i}^{\text{hits}} \left(\frac{m_{ij} - f_{ij}(\mathbf{p}_0,\mathbf{q}_{0j}) - \Delta\mathbf{p} f_{ij}(\mathbf{p}_0,\mathbf{q}_{0j}) \Delta\mathbf{q}_j f_{ij}(\mathbf{p}_0,\mathbf{q}_{0j})}{\sigma_{ij}} \right)^2 \tag{A.4}$$

A.1.2.1 MillePede-II

MILLEPEDE-II is a project developed in Hamburg (jointly at DESY and at the University) [6]. It is external to CMS, and is also used in other experiments like Belle [8].

The approach consists in performing a global fit, at the same time determining the correction to the alignment of the modules and refitting the tracks. Therefore, all correlations are treated in a mathematically rigorous way. The linearisation of the χ^2 allows to treat the problem with linear algebra.

$$\mathbf{C} \times (\mathbf{p}\ \mathbf{q}) = \mathbf{b} \tag{A.5}$$

Then the large number of parameters can be treated thanks to the special structure of the matrix $\mathbf{C}$. Indeed, it can be partitioned into blocks for *local* and *global* parameters (related to tracks and modules, respectively), reducing drastically the size of the matrix to invert:

$$\mathbf{C}_j\mathbf{q}_j = \mathbf{b}_j \qquad \text{local parameters} \tag{A.6}$$

$$\mathbf{C}'\mathbf{p} = \mathbf{b}' \qquad \text{global parameters} \tag{A.7}$$

1. The laser-based method has also been performed [3]. However, it can only align large mechanical structures.

where $\mathbf{b}'$ can be determined from $\mathbf{q}_j$ and $\mathbf{C}'$ from $\mathbf{C}_j^{-1}$ and some additional blocks in C describing correlations between local and global parameters.

In practice, "MILLEPEDE = Mille + Pede":

Mille determination of all the values needed to calculate the global χ^2 *i.e.* $\mathbf{p}$, $\mathbf{q}$, $\mathbf{m}$, σ, local $\mathrm{d}\mathbf{f}/\,\mathrm{d}\mathbf{q}$ and global $\mathrm{d}\mathbf{f}/\,\mathrm{d}\mathbf{p}$ parameters

Pede determination of local (track) refits to construct the linear equation system and determination of global (alignment) parameters

As a global-fit approach, it can be quite demanding in terms of memory and needs dedicated machines. Typically, a full alignment of the CMS tracker requires a running time of the order of a day.

A.1.2.2 HipPy

The HIPPY project was the first adopted technique of alignment of the CMS detector and is based on a different idea [7]. The dependence on the track parameters is removed in the χ^2 (local-fit approach); to compensate from this assumption, the procedure is iterated. Moreover, since it is not demanding in terms of memory, it can be used with very large statistics.

At DESY, the activity related to alignment are only focused on MILLEPEDE-II; therefore, HIPPY is mainly mentioned for completeness, but will not be discussed further.

A.1.3 Samples

A natural choice consists in taking Minimum Bias (MB) tracks, in order to scan most regions of the phase space. Additional types of tracks may however be considered.

First, *isolated muons* are used to fill regions of higher transverse momentum in the phase space, where the MB statistics is too low; in other words, since the transverse momentum is higher, the tracks are straighter, which allows to constrain modules in different series.

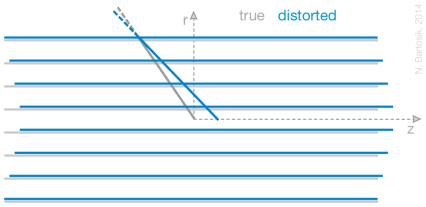
In the next subsection, we shall see why and how two other types of tracks are crucial to perform the alignment:

1. tracks of *cosmic rays*,
2. tracks of muons coming from the decay of a Z^0 boson (*i.e.* $Z^0 \longrightarrow \mu\mu$).

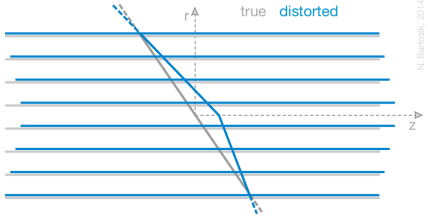
A.1.4 Weak modes

A Weak Mode (WM) refers to any transformation such that $\Delta\chi^2 \approx 0$. Equivalently, it is a transformation that changes a set of *valid* tracks into another set of *valid* tracks. This may happen in our case for two reasons:

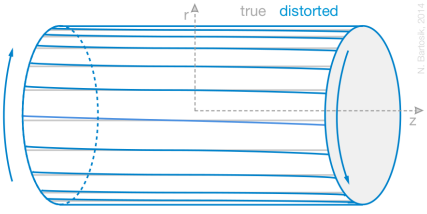
1. all collision tracks come from the center of the detector, and the collision are performed in the centre-of-mass frame with identical protons;
2. the detector is symmetric around the beam axis, with respect to the transverse plane passing through the center of the detector, and with respect to the center of the detector.



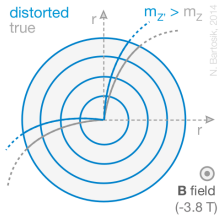
(A) Telescope without cosmic rays.



(B) Telescope with cosmic rays.



(C) Twist without $Z^0 \longrightarrow \mu\mu$.



(D) Twist with $Z^0 \longrightarrow \mu\mu$.

FIGURE A.5 – Two WMs appearing when only MB tracks are used are illustrated here, together with a picture of how other track topologies may solve the problem (plots from [9]).

A.1.4.1 Types

There are different types of WMs; for instance:

Telescope longitudinal deformations:

$$\Delta y = C \times z \tag{A.8}$$

Twist azimuthal deformations:

$$\Delta \phi = C \times z \tag{A.9}$$

Sagitta radial deformations:

$$\Delta y = C \times r \tag{A.10}$$

z ***rescaling*** longitudinal deformations:

$$\Delta z = C \times z \tag{A.11}$$

The telescope and twist WMs are illustrated in Fig. A.5.

A.1.4.2 Fix with tracks of various topologies

The solution consists in using tracks of cosmic rays and tracks of muons decaying from a Z^0 boson.

Cosmic rays. These tracks have a different topology than the collision tracks. Therefore, they break the centered symmetry of the problem. They are particularly useful to solve the telescope and z rescaling WMs.

$Z^0 \longrightarrow \mu\mu$ data. The two muons coming from a Z^0 boson introduce a momentum scale in the alignment. Moreover, they connect modules in different directions[2]. They are particularly useful to solve the twist and z rescaling WMs.

A.1.5 Time variations

The alignment is a very fine calibration that can also vary over time. This may be a difficulty, as datasets (especially for cosmic rays) may have very limited statistics.

There are different reasons for these time variations.

A.1.5.1 Types and reasons for variations

Magnet cycles. Ideally, the magnet should be turned on once for all. However, for maintenance reasons, even in data-taking periods, it may be temporarily turned off[3]. This mostly affects the large mechanical structures, for $\mathcal{O}(1\,\text{mm})$.

Temperature variations. In data-taking mode, the tracker is cooled down to -15^0 C. Also for maintenance purposes, like a long shut-down of the detector, the cooling operations may affect the alignment, this time at the level of the modules, for $\mathcal{O}(10\,\mu\text{m})$.

Ageing of the modules. As we discussed in Sect. 3.2.2.1, the modules operate in a high-radiation environment; therefore, their performance vary with time. In particular, the Lorentz drift, *i.e.* the drift of the released particles in the modules, is significantly affected and requires dedicated calibration. The Lorentz angle depends on several parameters:

1. the magnetic field,
2. the electric field,
3. the mobility of the charge carriers,
4. the thickness of the active zone.

While the magnetic field is kept constant over time, the three latter parameters are changing with time, due to the high radiation, *i.e.* the measurement of the position of the hit changes with $\mathrm{d}x \approx \tan\theta_{\text{Lorentz}}$, as is illustrated on Fig. A.6. At CMS, modules are mounted pointing either inward or outward (see Fig. A.3), therefore this change over time applies with different signs in the measurement of the position of the hits. The effect is continuous over time, and for a long period of data taking, it has to be taken into account. There is also a dedicated calibration of the pixel modules to compensate for the variation of the Lorentz drift, but it was found in practice that it is optimally corrected when treated complementarily in the alignment procedure, as will be shown later on.

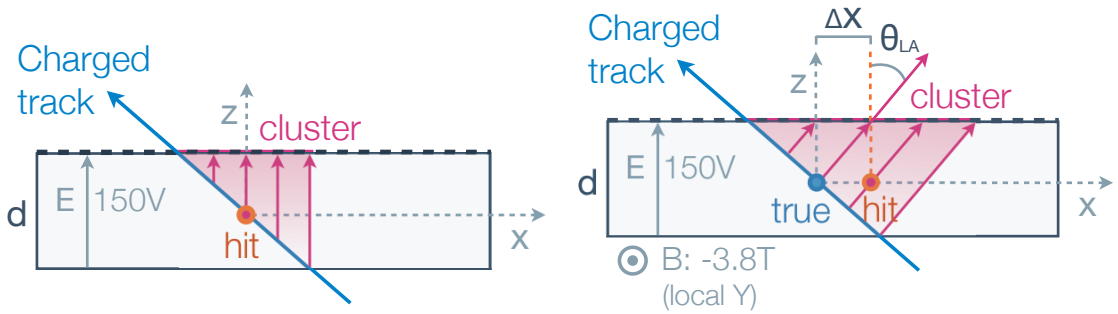
A.1.5.2 Strategy

Align separately:

— *absolute* positions of the High-Level Structures (HLSs) *with time-dependence*;

2. Similarly, one can also include $\Upsilon \longrightarrow \mu\mu$ data: not only they also introduce a scale, but they also connect modules in additional directions, as the muons decaying from a Z^0 are mostly back to back, while muons decaying from a Υ meson are rather making a right angle; they may therefore solve additional WMs. In 2016, they were not ready in time to be used by MILLEPEDE but were used with HIPPY.

3. Especially in 2015, the magnet was suffering of a severe issue, and was turned on and off quite frequently, forcing the alignment crew to perform the alignment on an almost daily basis.



(A) At $B = 0\,\mathrm{T}$.

(B) At $B = 3.8\,\mathrm{T}$.

FIGURE A.6 – When a high-energy particle goes through a silicon module, it releases charge carriers that are drifted by the ambient electric field. The *Lorentz drift* denotes the additional drift due to the magnetic field of the CMS detector. As modules get significantly aged over the data-taking period, this effects varies over time (from [9]).

— *relative* positions of the modules to the HLSs *without time-dependence*.

The HLS can be

— either the large mechanical structures (already described in Fig. A.2),

— or disks and layer (also represented in Fig. A.2),

— or the ladders and blades (sketched in Fig. A.3).

This strategy turns out to be a good compromise to include time dependence while keeping large statistics, especially for cosmic rays and dimuon data. The choice of the HLSs is motivated by the treatment of the Lorentz drift.

A.2 2016 Legacy Alignment

We present the performance of the alignment in 2016. First we detail the configuration; then we plot the performance of the detector for a given interval of time.

A.2.1 Configuration

The period of data taking is divided into 36 intervals of time. It is performed at the level of the sensors, which is possible thanks to high statistics (see Table A.2).

Several iterations are performed with MILLEPEDE-II and HIPPY in order to cope with the linearisation of the problem in case of large deviations. The choice of the HLSs was set to ladders and blades in order to correct for the non-constant Lorentz angle (since a ladder contains modules pointing in the same direction, the correction to its position can absorb the Lorentz drift).

A.2.2 Performance

The difference between two geometries can be checked on a *Geometry Comparison Plot*. In Fig. A.7, we compare the geometry of the detector during the data-taking and the one after the alignment procedure. Each point represents a module; the colour is related to the high-level structure consistently with Fig. A.2. One can see the movement $Y(\Delta r, \Delta z, r\Delta\phi)$ of a module initially at position $X(r, z, \phi)$.

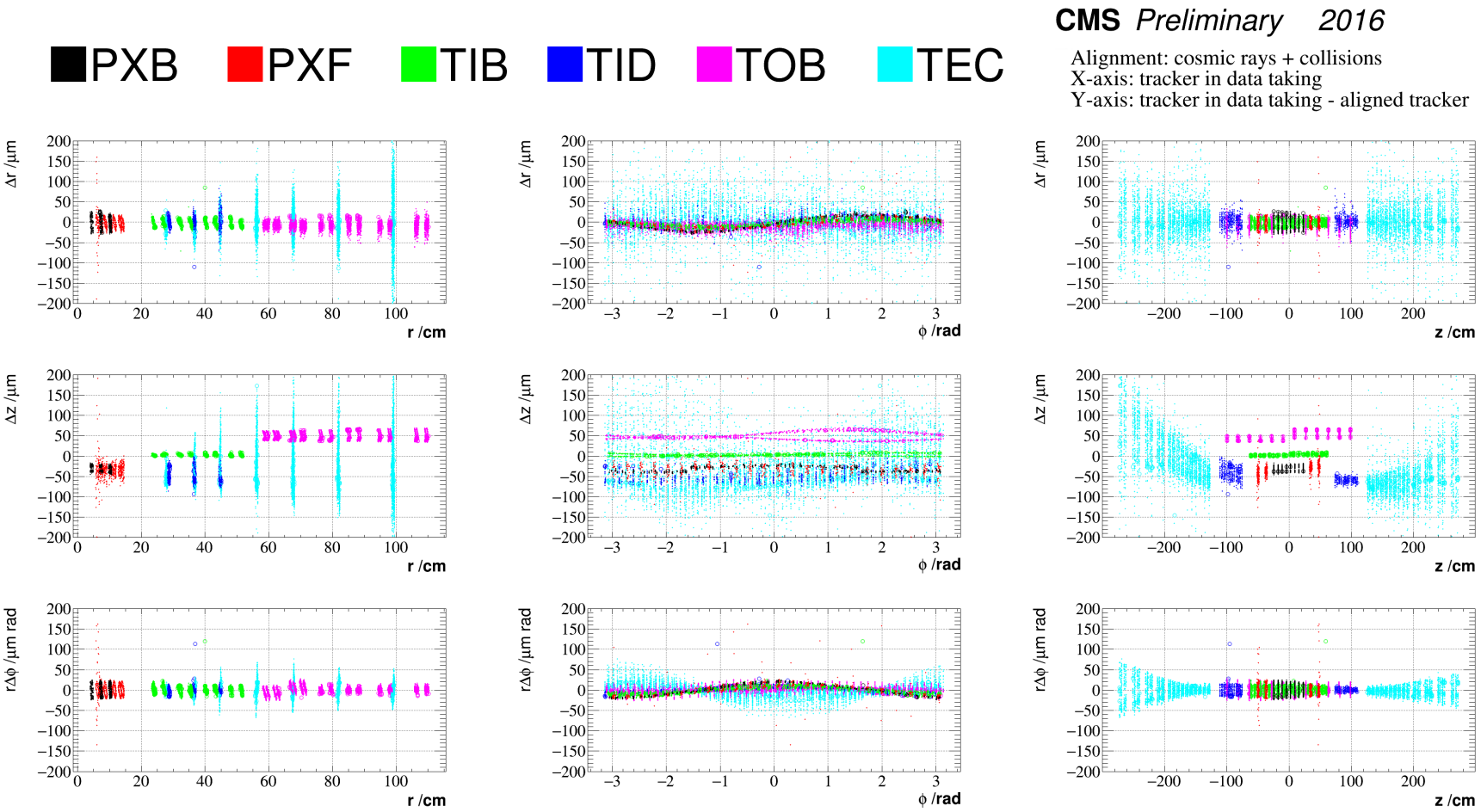


FIGURE A.7 – *Geometry comparison plot* between the geometry during data-taking and the geometry after 2016 Legacy Alignment.

dataset	#tracks	weight
MB tracks	$13M$	0.2 – 0.3
isolated muons	$53M$	0.25
$Z \longrightarrow \mu\mu$	$32M$	1.0
cosmic rays	$3M$	2.5

TABLE A.2 – Statistics in use for MILLEPEDE-II. Note: this configuration and this statistics require 150 GB of RAM and around 30 h per iteration.

In the current case, clear differences between the tracker in data-taking and after re-alignment are seen, but are not sufficient to tell whether the alignment has improved. Data-driven quantities are however checked in order to compare the quality of different alignments. Here, we present the following ones:

— distributions of the median of the residuals,

— performance of the reconstruction of the PV as a function of the track kinematics,

— reconstruction of the mass of the Z^0 boson as a function of the kinematics of the outgoing muons.

As the alignment is performed in 36 interval of times, the performance of the alignment can be studied in each of them. Apart from the survey of the Lorentz drift, only one interval of time with large statistic is sufficient to attest the global performance of the alignment.

For more readability, the geometries in data taking and after realignment are respectively in red and blue in the different validation plots. In addition, for reference, a geometry without misalignment is displayed in green.

A.2.2.1 Distribution of the medians of the residuals

The *distributions of the medians of the residuals* are a measure of the local precision.

Let us consider a set of tracks:

— Each track is reconstructed for different geometries.

— The hit prediction x'_{pred} for each module is obtained from all other track hits.

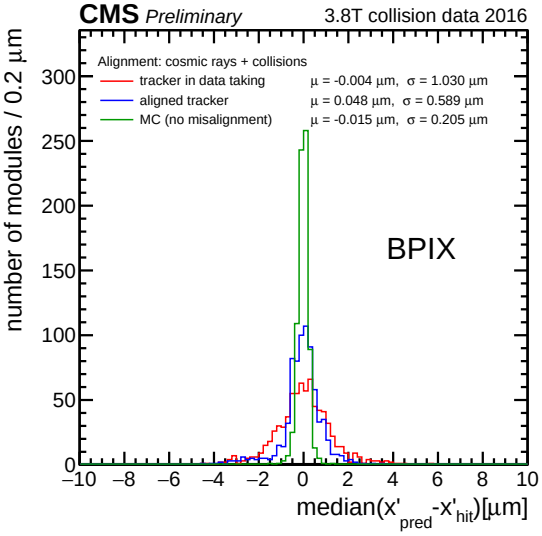
— Then the residuals $x'_{\text{pred}} - x'_{\text{hit}}$ is histogrammed for each module.

— Finally, for each large mechanical structure, the median of these residuals is plotted.

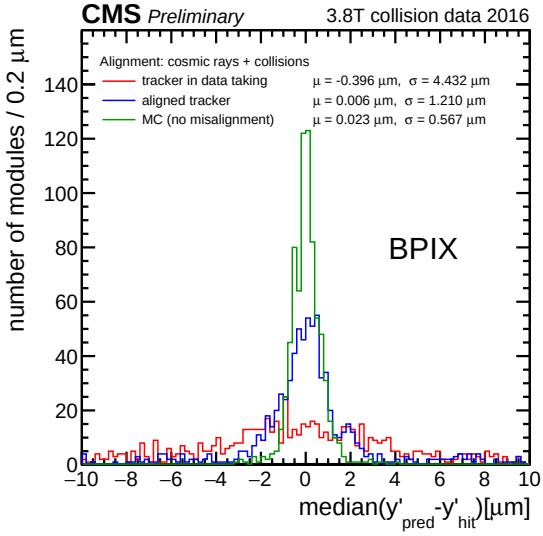
This validation can be seen in our case on Figs. A.8-A.9. Note that in order to avoid statistical correlations, we use independent samples for alignment and validation.

An optimally aligned detector should be peaked around zero, as the Monte Carlo reference suggests. The width has also a component related to the statistics, which is why the Monte Carlo reference, though ideal, still presents a width. Deviations from 0 for the mean indicate systematic biases.

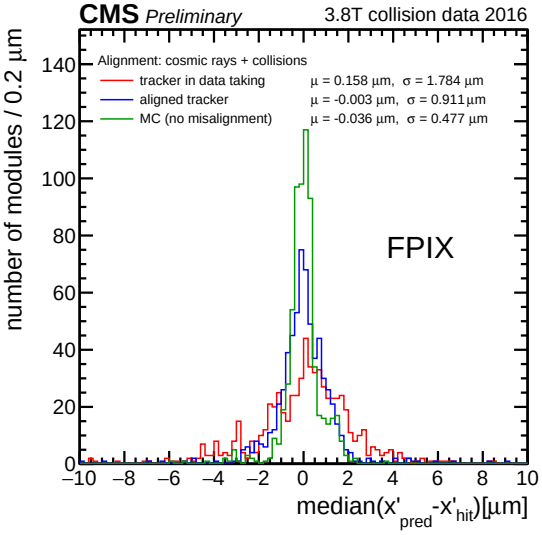
Lorentz drift. The variation of the Lorentz drift can be investigated over time from the distributions of the median of the residuals. In particular, they are produced for each interval of time independently for inward and outward pointing modules. For a wrong correction of the Lorentz angle (typically in the case of a constant geometry), two peaks ($\Delta\mu \neq 0$) will appear in the distributions, as the local x shows in different directions for



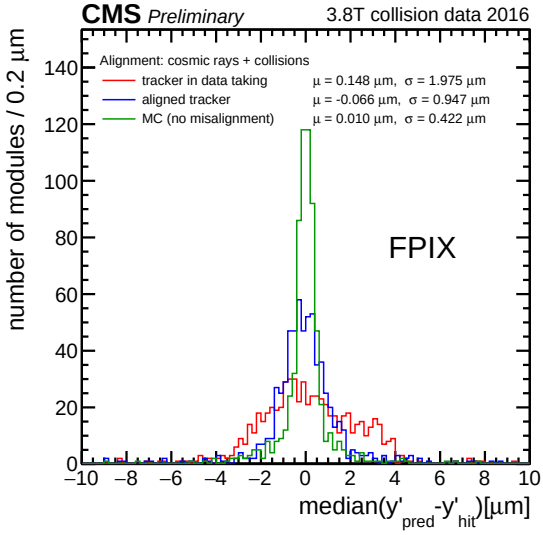
(A) Pixel barrel in the X-direction.



(B) Pixel barrel in the Y-direction.

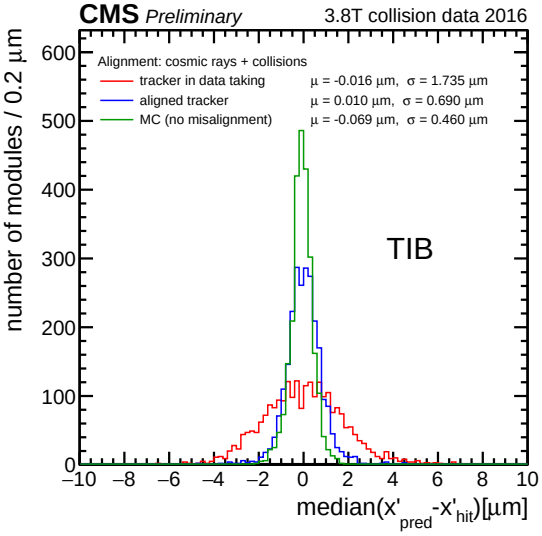


(C) Pixel forward in the X-direction.

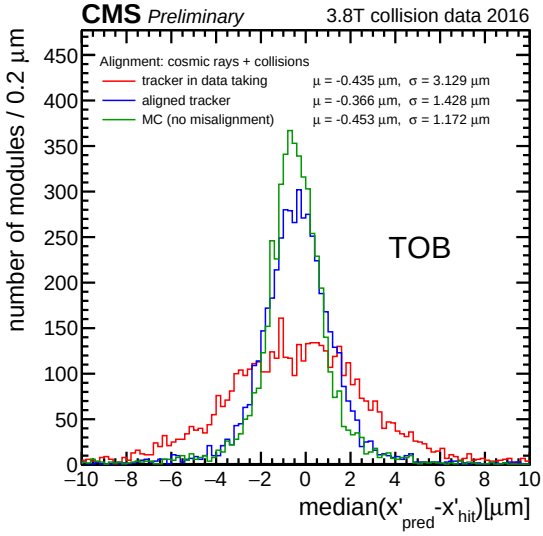


(D) Pixel forward in the Y-direction.

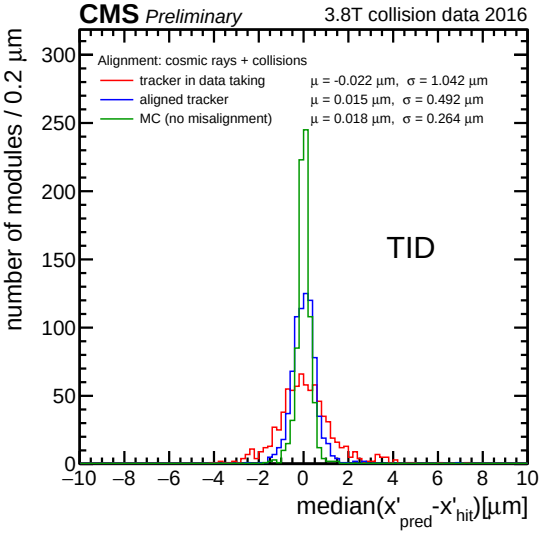
FIGURE A.8 – Distributions of the medians of the residuals in the pixel part of the tracker [10].



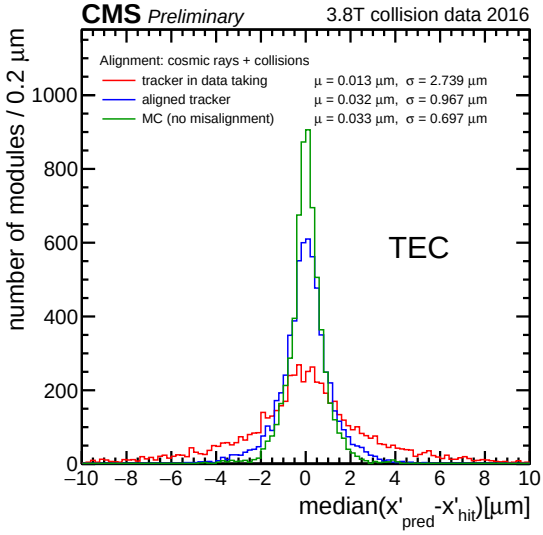
(A) Tracker end-caps.



(B) Tracker end-caps.



(C) Tracker end-caps.



(D) Tracker end-caps.

FIGURE A.9 – Distributions of the medians of the residuals in the strip part of the tracker for 2016 Legacy Alignment [10].

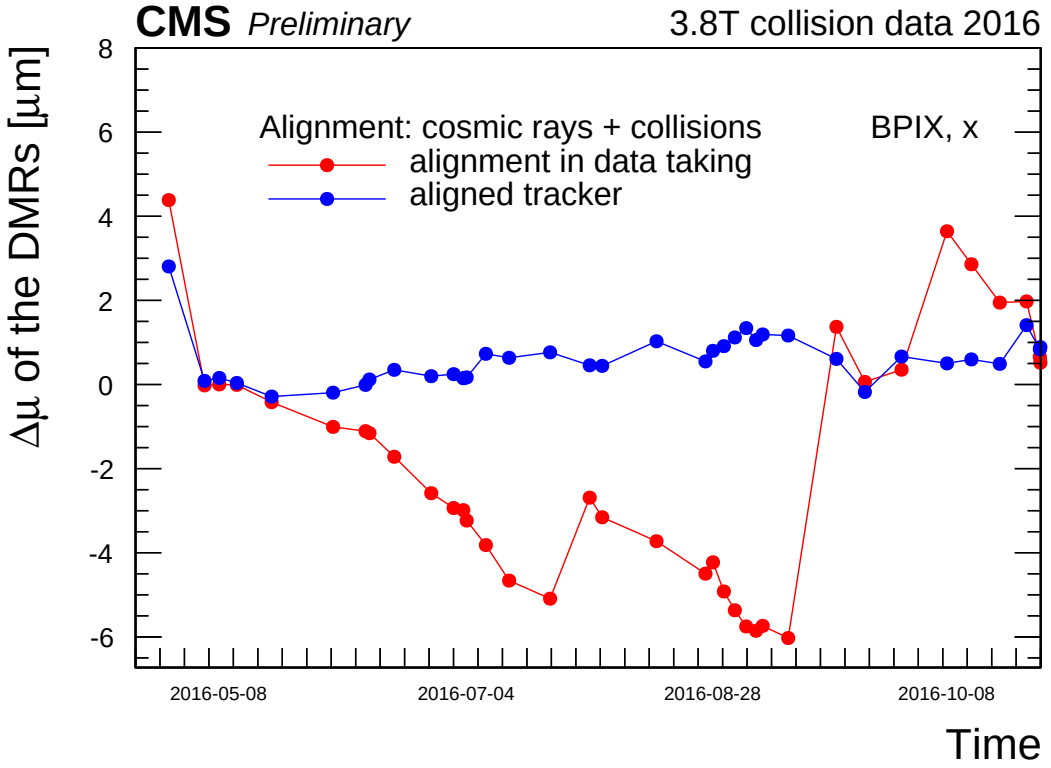


FIGURE A.10 – Correction of Lorentz drift for the 2016 Legacy Alignment [10]. Difference of means in the distributions of the medians of the residuals over time for inward- and outward-pointing modules in barrel pixel part of the detector. The jumps in the red curve correspond to updates of the calibration of the pixel.

different module orientations; on the other hand, with a new alignment, this would be absorbed by the geometry, as ladders are considered as high-level structures, in which case no double-peak structure should be seen ($\Delta\mu = 0$) [4]. This is illustrated on Fig. A.10.

A.2.2.2 Primary-vertex validation

The PV tests the alignment by looking at the reconstruction of the vertices.

Let us consider a vertex reconstructed from N tracks. Then

1. consider one tracks;
2. refit the vertex with the $N-1$ other tracks with the usual reconstruction;
3. and check the impact parameter of the track under scrutiny with respect to the vertex as a function of the direction of the tracks.

Random misalignments translate into an increase of the spread, and simply lead to lower precision, while systematic misalignment translate to biases in the mean (the exact pattern depends on the misalignment), and may lead to systematic biases in the measurement.

To perform the PV validation, we consider tracks from MB events satisfying the following requirements:

Vertex — at least four degrees of freedom in the vertex fit,

4. It is also possible to disentangle the Lorentz angle from the alignment by using data with magnetic field on and off, but this was not done in 2016.

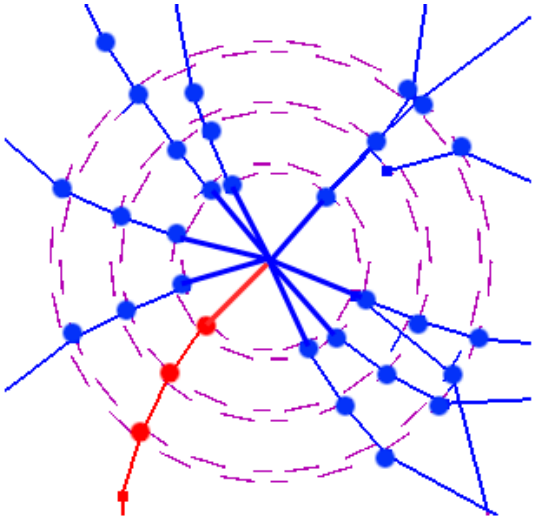


FIGURE A.11 – Picture of the PV validation. Given a vertex composed of N tracks, it is refitted with $N-1$ tracks; the impact parameter of the Nth track is then studied [11].

Tracks — at least six hits in the tracker, of which at least two in the pixel detector,
— at least one hit in the first layer of the Barrel Pixel or the first disk of the `PXF`
— $\chi^2_{\text{track}}/\text{n.d.f.} < 5$

The case of the 2016 Legacy Alignment is shown in Fig. A.12. In particular, one can see that the modulation in ϕ in the geometry during data-taking is cured after realignment.

A.2.2.3 $Z \to \mu\mu$ validation

Distortions in the geometry may degrade the kinematics of the two outgoing muons coming from the decay of a Z^0 boson. The reconstruction of the Z^0 boson is thus investigated by measuring its mass as a function of the kinematics of the muons.

The Z^0-boson mass is reconstructed with a Voigtian function [5] with fixed decay width for the Breit-Wigner component, while the background is reconstructed with a exponential function. The mass is then estimated from the mean of the Voigtian function as a function of different variables:

— the azimuthal angles $\phi_{\mu\pm}$ of each of the muons,
— the rapidity separation $\eta_{\mu+} - \eta_{\mu-}$,
— the cosine of the angle of the boson $\cos\theta_{\text{CS}}$ in the Collins-Soper frame.

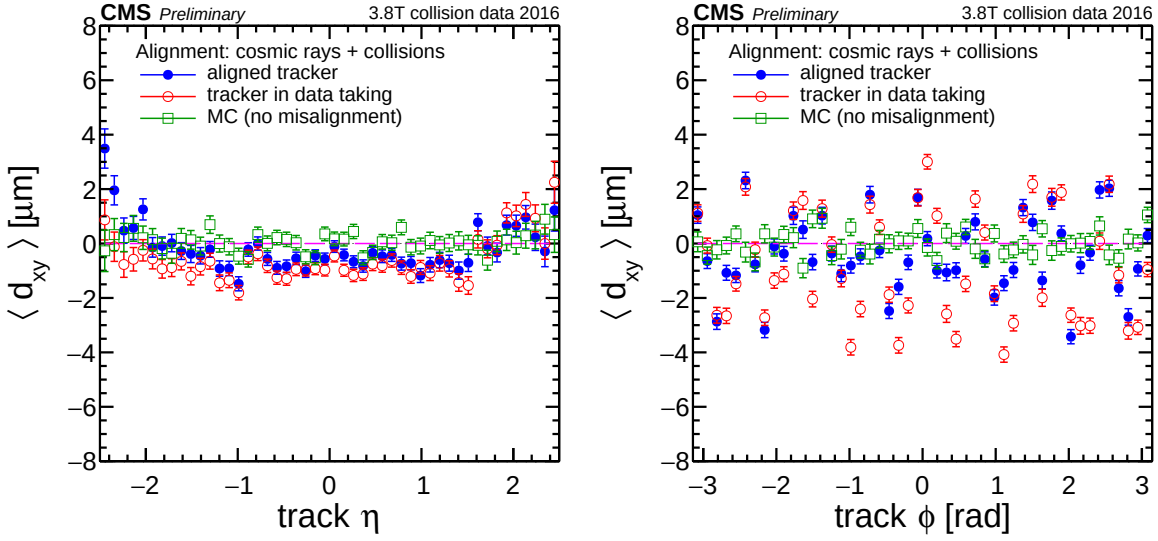
This is shown in Fig. A.13. In addition, a fit of the mass is performed for each geometry.

The selected muons must satisfy the following requirements:

— $p_T > 20\,\text{GeV}$
— $|\eta| < 2.4$
— $80\,\text{GeV} < M_{\mu\mu} < 120\,\text{GeV}$

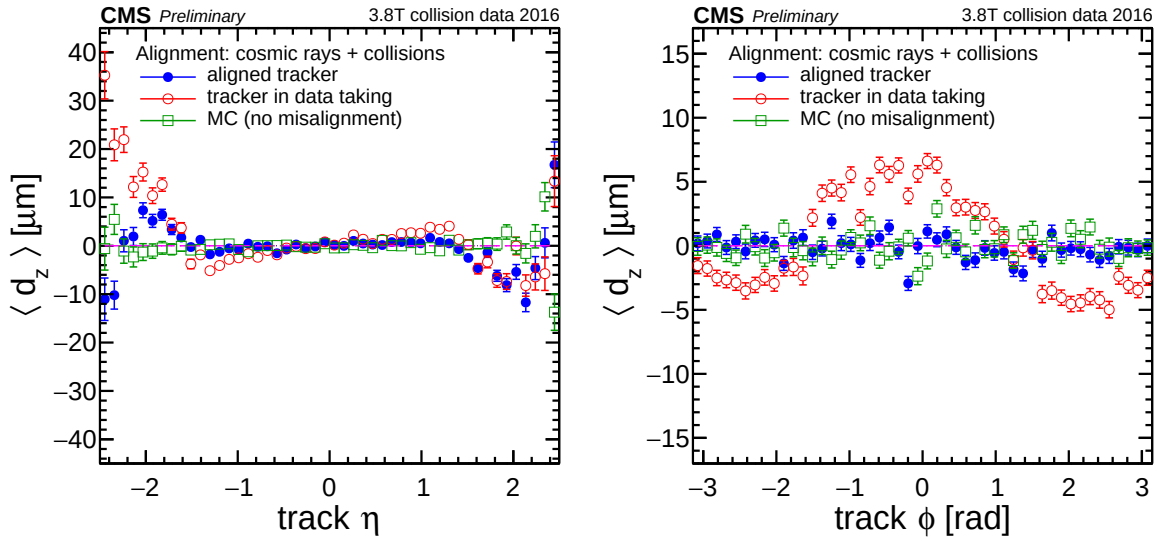
5. Convolution of Gaussian and Lorentzian functions:

$$V(x) = \int_{-\infty}^{+\infty} \exp\left(-a\left((x-t)-x_0\right)^2\right) \times \frac{1}{b+(t-x_0)^2}\,\mathrm{d}t \tag{A.12}$$



(A) $< d_{xy} >$ vs. η

(B) $< d_{xy} >$ vs. ϕ



(C) $< d_z >$ vs. η

(D) $< d_z >$ vs. ϕ

FIGURE A.12 – PV validation for the 2016 Legacy Alignment, performed with PV tracks [10]. The impact parameter of the track under scrutiny of a given vertex is plotted along the beam axis $< d_z >$ and in the transverse plane $< d_{xy} >$ as a function of its azimuthal angle ϕ and pseudorapidity η. Here, the ϕ modulation and high-η deformations in $< d_z >$ have clearly been fixed.

Note that muons are reconstructed with both the tracker and the muon system, but only the geometry of the tracker is updated in the validation.

A.3 Summary & Conclusions

The alignment of the tracker is a crucial step in the procedure of calibration of the detector; in the hierarchy, it is one of the first to perform, and many other calibrations rely on it: calorimeter alignment, muon alignment, calorimeter calibration, beamspot calibration, etc. Given the large size of the silicon tracker, dedicated techniques needed to be developed; these were concisely presented. As an illustration, the alignment campaign of the 2016 data was presented in this appendix; it is one of the most precise campaigns ever performed at CMS so far, not only in terms of data, but also in terms of configuration.

Other aspects of the alignment, *e.g.* online automated calibration or other campaigns, are also covered in Ref. [12].

References

[1] Serguei Chatrchyan et al. "Description and performance of track and primary-vertex reconstruction with the CMS tracker". In: *JINST* 9.10 (2014), P10009. DOI: 10.1088/1748-0221/9/10/P10009. arXiv: 1405.6569 [physics.ins-det].

[2] S Chatrchyan et al. "Commissioning and Performance of the CMS Pixel Tracker with Cosmic Ray Muons". In: *JINST* 5 (2010), T03007. DOI: 10.1088/1748-0221/5/03/T03007. arXiv: 0911.5434 [physics.ins-det].

[3] Albert M Sirunyan et al. "Mechanical stability of the CMS strip tracker measured with a laser alignment system". In: *JINST* 12.04 (2017), P04023. DOI: 10.1088/1748-0221/12/04/P04023. arXiv: 1701.02022 [physics.ins-det].

[4] CMS Collaboration. "Alignment of the CMS silicon tracker during commissioning with cosmic rays". In: *Journal of Instrumentation* 5.03 (2010), T03009. URL: http://stacks.iop.org/1748-0221/5/i=03/a=T03009.

[5] The CMS collaboration. "Alignment of the CMS tracker with LHC and cosmic ray data". In: *Journal of Instrumentation* 9.06 (2014), P06009. URL: http://stacks.iop.org/1748-0221/9/i=06/a=P06009.

[6] Volker Blobel and Claus Kleinwort. "A New Method for the High-Precision Alignment of Track Detectors". In: *Proceedings of the Conference on Adcanced Statistical Techniques in Particle Physics* (2002).

[7] CMS Collaboration. *The HIP Algorithm for Track Based Alignment and its Applicatino to the CMS Pixel Detector*. http://lib.tkk.fi/Diss/2007/isbn9789521037115/article6.pdf. Accessed: 2018-02-07.

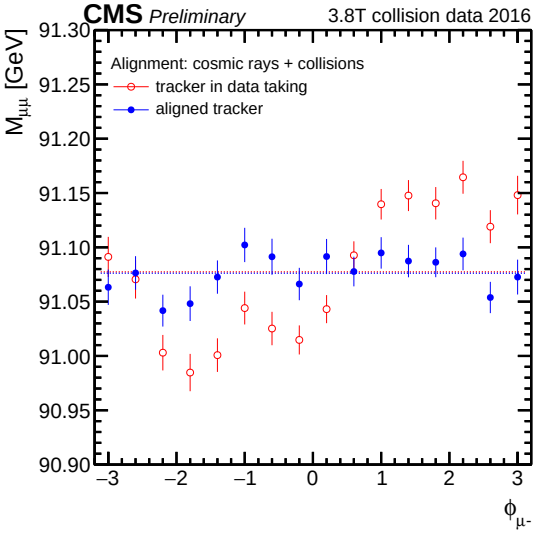
[8] T. Abe et al. "Belle II Technical Design Report". In: (2010). arXiv: 1011.0352 [physics.ins-det].

[9] Nazar Bartosik. "Associated top-quark-pair and b-jet production in the dilepton channel at $\sqrt{s} =$ 8 TeV as test of QCD and background tt+Higgs production". PhD thesis. Hamburg: U. Hamburg, Dept. Phys., 2015. DOI: 10.3204/DESY-THESIS-2015-035. URL: http://bib-pubdb1.desy.de/search?cc=Publication+Database&of=hd&p=reportnumber:DESY-THESIS-2015-035.

[10] CMS Collaboration. *CMS Tracker Alignment Performance Results 2016*. https://twiki.cern.ch/twiki/bin/view/CMSPublic/TkAlignmentPerformanceEOY16. Accessed: 2018-02-08.

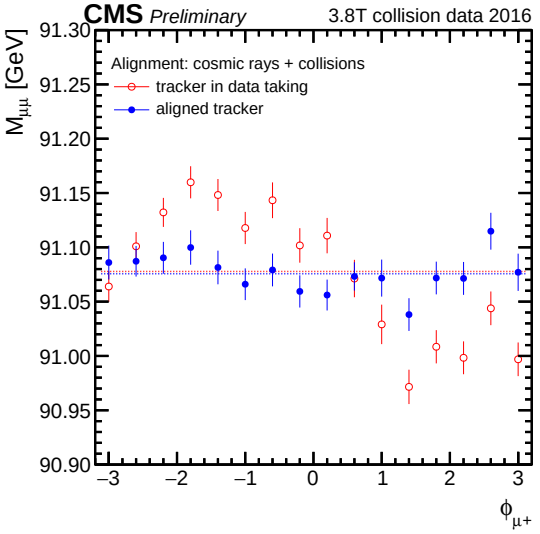
[11] Marco Musich. *Primary-vertex validation*. Private communication.

[12] Gregor Mittag and CMS Collaboration. "Alignment of the CMS Tracker: Latest Results from LHC Run-II". In: *Journal of Physics: Conference Series* 898.4 (2017), p. 042014. URL: http://stacks.iop.org/1742-6596/898/i=4/a=042014.



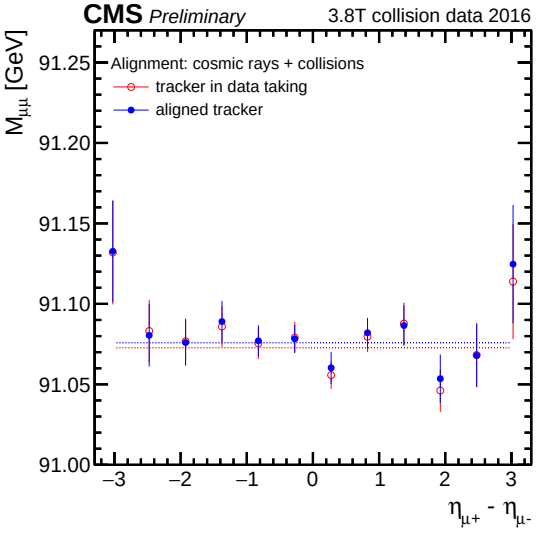
	χ^2/n.d.f.	p-value
tracker in data taking	15.99	< 0.01
aligned tracker	1.39	0.14

(A) Z^0 mass as a function of the negatively charged muon.



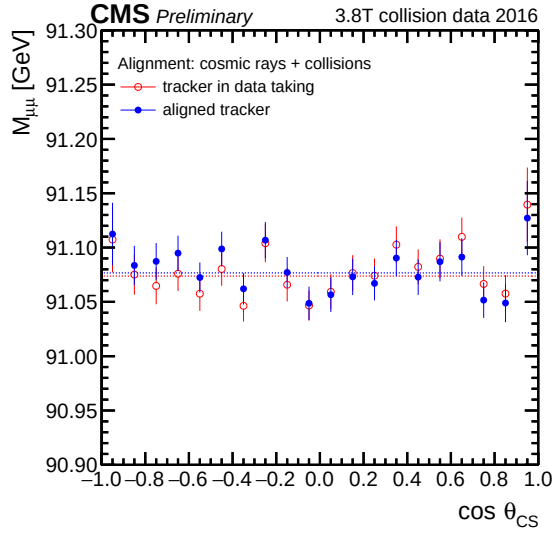
	χ^2/n.d.f.	p-value
tracker in data taking	15.76	< 0.01
aligned tracker	1.33	0.17

(B) Z^0 mass as a function of the positively charged muon.



	χ^2/ndf	p-value
tracker in data taking	1.31	0.22
aligned tracker	0.80	0.61

(C) Z^0 mass as a function of the pseudo-rapidity separation in the Collins-Soper frame.



	χ^2/ndf	p-value
tracker in data taking	1.43	< 0.09
aligned tracker	1.25	0.21

(D) Z^0 mass as a function of the cosine of the angular separation in the Collins-Soper frame.

FIGURE A.13 – $Z^0 \to \mu\mu$ validation for the 2016 Legacy Alignment [10]. The mass of the Z^0 boson is fitted as a function of different kinematic parameters of the outgoing muons. Any deviations indicates systematic misalignment of the tracker.

B

TMD Plotter

In parallel of the work presented in this thesis, a dedicated tool to plot collinear as well as TMD densities. This work has been presented at the DIS16 conference in Hamburg and at the REF16 workshop in Antwerp. The proceedings of the DIS conference are given in the next pages.

Figure B.1 – Logo of the DIS 2016 conference.

PROCEEDINGS OF SCIENCE

TMDlib 1.0.8 and TMDplotter 2.1.1

Patrick L.S. Connor*
DESY, Germany
E-mail: patrick.connor@desy.de

Hannes Jung
DESY, Germany
Universiteit Antwerpen, Belgium
E-mail: hannes.jung@desy.de

Francesco Hautmann
Oxford University, UK
E-mail: f.hautmann1@physics.ox.ac.uk

Johannes Scheller
Eberhard Karls Universität Tübingen, Germany
DESY, Germany
E-mail: johannes.scheller@desy.de

The latest versions of TMDlib and TMDplotter are presented. Parameterisations of TMDs in TMDlib are illustrated. New features of TMDplotter include plotting as a function of the momentum fraction, the transverse momentum, or the evolution scale, as well as integration of TMDs and comparison to sets from LHAPDF. Luminosity for collinear PDFs is also described. The tool is available at http://tmdplotter.desy.de.

XXIV International Workshop on Deep-Inelastic Scattering and Related Subjects
11-15 April, 2016
DESY Hamburg, Germany

*Speaker.

© Copyright owned by the author(s) under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (CC BY-NC-ND 4.0).

1. Introduction

The *Transverse-Momentum Dependent* (TMD) *Parton Density Functions* (PDF) $\mathscr{A}(x,k_t,p)$ extend collinear PDFs $f(x,p)$ by taking into account the transverse momentum k_t of the parton, where x is the momentum fraction and p the evolution scale. TMDs are necessary for a proper description of multi-scale processes [1].

TMDlib and TMDplotter are two utility tools available on internet [2, 3] to help study TMDs quantitatively [4]: the former is a `C++` library that centralises many TMD parameterisations in a common framework, similarly to the LHAPDF library [5], and the latter is an online tool making use of TMDlib to plot and compare TMDs among them, or to integrate and compare them with collinear PDFs; it also allows to compute *partonic luminosities*.

In order to illustrate the tool, we will use `ccfm-JH-2013` sets [6].

2. TMDlib

TMDlib is a `C++` library for working with TMDs. A precise description of the theoretical background and of the tool can be found in [4].

The list of available sets can be found on `http://tmdlib.hepforge.org`. The initialisation and the call are done on a unified way. Storage and interpolation are done automatically and do not require any action from the user.

3. TMDplotter

TMDs as a function of the momentum fraction x, the transverse momentum k_t and the evolution scale p can be plotted directly from TMDlib using TMDplotter.

Integrated TMDs are obtained from TMDs by integration over k_t (using the QAGS algorithm from the GNU Scientific Library [7]), with integration limits $k_t^{\max}$ and $k_t^{\min}$ to be given by the user:

$$xf(x,p) = \int_{k_{t,\min}}^{k_{t,\max}} dk_t^2\, x\mathscr{A}(x,k_t,p) \tag{3.1}$$

TMDplotter can also compute and plot the *partonic luminosity*, both for collinear PDFs and for integrated TMDs. This quantity describes the contribution from the PDFs to the hadronic cross section. The partonic luminosity reads [8]:

$$\frac{d\mathscr{L}_{ij}}{d\tau} = \frac{1}{1+\delta_{ij}} \int_0^1 dx_1 \int_0^1 dx_2\, f_i(x_1,p)\, f_j(x_2,p)\, \delta(\tau - x_1x_2) + (i \leftrightarrow j) \tag{3.2}$$

with $\tau = x_1x_2$ the product of the momentum fractions.

Similarly, the TMD luminosity can also be defined from a TMD:

$$\begin{aligned}\frac{d^2\mathscr{L}_{ij}}{d\tau dq_t} =& \frac{1}{1+\delta_{ij}} \int_0^1 dx_1 \int_0^1 dx_2 \int \frac{d^2\mathbf{k_{t1}}}{\pi} \int \frac{d^2\mathbf{k_{t2}}}{\pi}\, \mathscr{A}_i(x_1,k_{t1},p)\, \mathscr{A}_j(x_2,k_{t2},p) \times \\ & \times \delta(\tau - x_1x_2)\, \delta^{(2)}(\mathbf{q_t} - (\mathbf{k_{t1}} + \mathbf{k_{t2}})) + (i \leftrightarrow j)\end{aligned} \tag{3.3}$$

where $\mathbf{q_t}$ is the sum of the transverse momenta.

4. Application

The two sets of `ccfm-JH-2013` are extracted for gluons from the combined measurements of proton's structure functions F_2 in deeply inelastic scattering at HERA collider. The `set1` is determined from the fit to the inclusive structure function only, whilst the `set2` is determined from both the inclusive and charm structure functions. A precise description of the extraction of the sets can be found in [6].

Using those sets, we show some applications of TMDplotter: in Fig. 1, the two sets are plotted as a function of x and k_t; and in Fig. 2 the integral of the second set is shown for different boundaries. From Fig. 2 we conclude that the lower integration limit does not play a large rôle, as soon as it is low enough, while the upper integration has a significant influence.

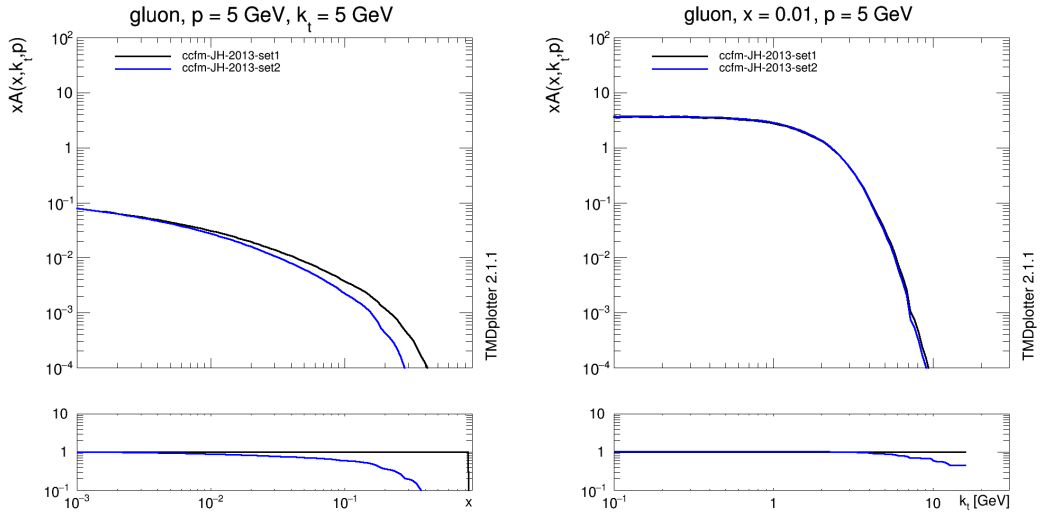


Figure 1: The `ccfm-JH-2013` sets are plotted as a function of x (left) and k_t (right). The lower panel shows the ratio with the first curve.

In Fig. 3, the partonic luminosity is plotted for two collinear PDFs for gluons and photons.

5. Summary

The two tools TMDlib and TMDplotter for working with TMDs have been presented. In particular, TMDplotter (`http://tmdplotter.desy.de`) has been augmented with many new features.

References

[1] R. Angeles-Martinez et al. Transverse Momentum Dependent (TMD) parton distribution functions: status and prospects. *Acta Phys. Polon.*, B46(12):2501–2534, 2015.

[2] F. Hautmann, H. Jung, P.L.S. Connor, M. Krämer, P. J. Mulders, E. R. Nocera, T. C. Rogers, and A. Signori. TMDlib. `tmdlib.hepforge.org`.

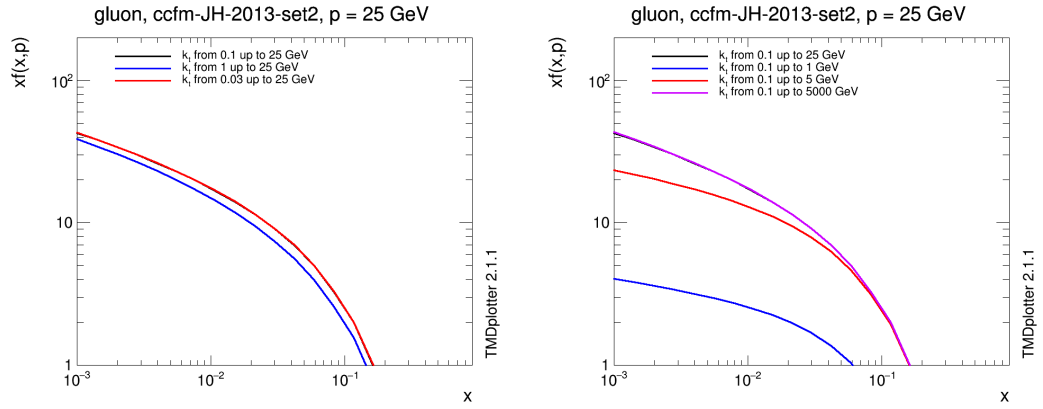


Figure 2: `ccfm-JH-2013` sets are integrated. On the left (right), the lower (upper) integration boundary is varied.

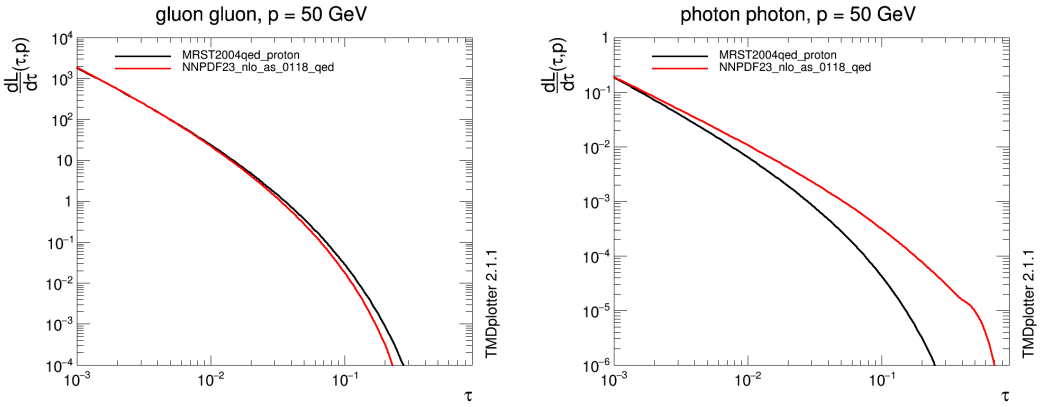


Figure 3: The partonic luminosities of two colliding gluons (left) and of two colliding photons (right) are shown for the collinear PDFs `MRST2004qed_proton` and `NNPDF23_nlo_as_0118_qed`.

[3] F. Hautmann, H. Jung, P.L.S. Connor, M. Krämer, and J. Scheller. TMDplotter. `tmdplotter.desy.de`.

[4] F. Hautmann, H. Jung, M. Krämer, P. J. Mulders, E. R. Nocera, T. C. Rogers, and A. Signori. TMDlib and TMDplotter: library and plotting tools for transverse-momentum-dependent parton distributions. *Eur. Phys. J.*, C74:3220, 2014.

[5] A. Buckley. New access to pdf data via lhapdf6. In *Les Houches 2013: Physics at TeV Colliders: Standard Model Working Group Report*, 2014.

[6] F. Hautmann and H. Jung. Transverse momentum dependent gluon density from DIS precision data. *Nucl. Phys. B*, 883:1 – 19, 2014.

[7] M. Galassi et al. *GNU Scientific Library Reference Manual*. Network Theory Ltd., 3rd edition edition, 1 2009.

[8] Richard Keith Ellis, William James Stirling, and Bryan R. Webber. *QCD and collider physics*, volume 8 of *Cambridge Monographs on Particle Physics, Nuclear Physics and Cosmology*. Cambridge University Press, December 2003.

C

Latest jet measurements at CMS

A review of the latest jet measurements at CMS was presented at the 2017 edition of the LHCP conference at the Shanghai Jiaotong university. The proceedings of this conference are given in the few next pages.

Figure C.1 – Logo of the LHCP 2017 conference.

Proceedings of the Fifth Annual LHCP
CMS CR-2017/358
November 2, 2017

Measurement of jet properties in CMS

Patrick L.S. Connor[1]

DESY, Notkestrasse 87, 22607 Hamburg, Germany

ABSTRACT

We present measurements of the inclusive jet production at centre-of-mass energies of 8 and 13 TeV, and of multijets at 8 TeV. These measurements allow to constrain PDFs and the strong coupling constant. Two measurements of the azimuthal correlations at 8 and 13 TeV are also presented, testing higher order QCD calculations.

PRESENTED AT

The Fifth Annual Conference
on Large Hadron Collider Physics
Shanghai Jiao Tong University, Shanghai, China
May 15-20, 2017

[1]on behalf of the CMS Collaboration

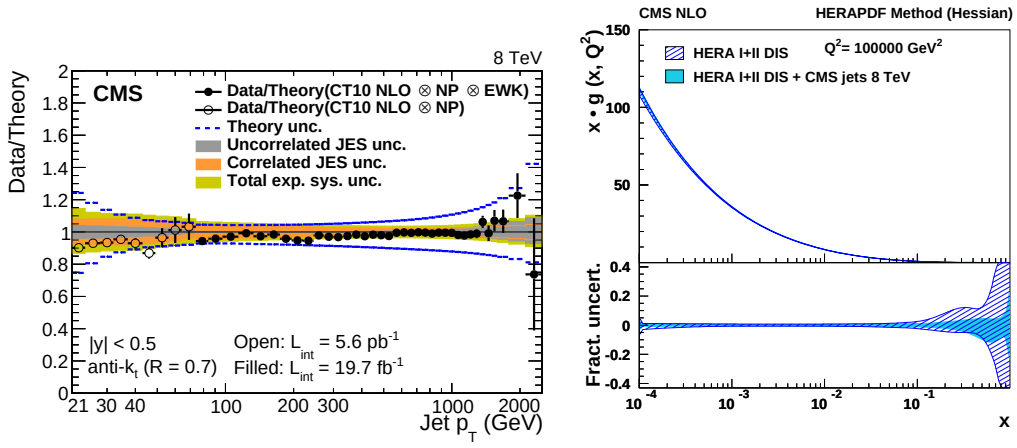


Figure 1: Left: comparison to fixed-order parton-level calculations of the inclusive jet production at 8 TeV. Right: gluon PDF before and after inclusion of inclusive jet data [2].

1 Introduction

We summarise the six most recent jet measurements in CMS [1] at 8 and 13 TeV. Analog measurements at 8 and 13 TeV are presented together.

2 Inclusive jet analyses at 8 and 13 TeV

The double differential cross section as a function of the transverse momentum and the rapidity is given by

$$\frac{\mathrm{d}^2\sigma}{\mathrm{d}p_T\mathrm{d}y} = \frac{1}{\epsilon\mathcal{L}_{\mathrm{int}}^{\mathrm{eff}}}\frac{N_{\mathrm{jets}}}{\Delta p_T \Delta|y|} \tag{1}$$

At 8 TeV [2], the luminosity for the high p_T region is 19.7 fb^{-1}, while for the low p_T region with dedicated low-pile-up runs, it is 5.6 pb^{-1}. The measurement is performed for large cone size jets only with the anti-k_T algorithm [8]. The rapidity coverage is $0 < |y| < 4.7$.

A similar measurement was performed with the first data at 13 TeV [3]. The luminosity is 71 pb^{-1} in the central region ($|y| < 3.0$) and 44 pb^{-1} in the forward region ($3.2 < |y| < 4.7$). The rapidity coverage and binning are kept the same as for the analysis at 8 TeV. The measurement at 13 TeV is performed for two cone size radii: $R = 0.4$ and $R = 0.7$.

The measurements are compared to predictions from MC event generators and from fixed-order parton-level calculations. In general, the uncertainty related to the correction of the jet energy scale is of the order of a percent.

The fixed-order calculation includes the electroweak and non-perturbative QCD corrections. We observe a very good agreement over the two orders of magnitude. The gluon PDF can be constrained with the inclusive jet data, and α_S can be extracted together with the fit of the PDF, as an additional parameter, with a value of $0.1164^{+0.0093}_{-0.0073}$. The measurement and the PDF fit are illustrated in Fig. 1.

The comparison to fixed-order parton-level calculations (not shown here) agrees better for a large cone size radius than for a small cone size radius. This is understood as being related to missing higher order calculations.

The comparison of the measurement with predictions from POWHEG + P8 (NLO + PS) shows very good agreement for both cone sizes at 13 TeV.

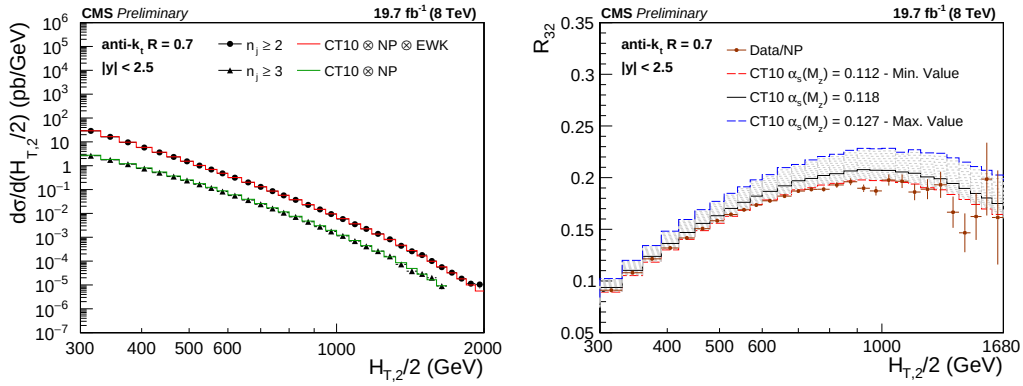


Figure 2: Measurement of the di-jet and tri-jet production at 8 TeV [4].

3 Multijet analysis at 8 TeV

The measurement of the multijet production was performed as a function of $H_{T,n} = \sum_{i=1}^{n} p_{T,i}$ with all the jets of an event in the considered phase space [4]:

$$\frac{\mathrm{d}\sigma}{\mathrm{d}(H_{T,2}/2)} = \frac{1}{\epsilon \mathcal{L}_{\mathrm{int}}^{\mathrm{eff}}} \frac{N_{\mathrm{events}}}{\Delta(H_{T,2}/2)} \tag{2}$$

Since $R_{mn} = \frac{\sigma_{m-\mathrm{jet}}}{\sigma_{n-\mathrm{jet}}} \propto \alpha_S^{m-n}$, α_S can be extracted from the ratio of 3 and 2-jet measurements. Some uncertainties cancel in the ratio. The selection consists of jets of $p_T > 150$ GeV in $|y| < 2.5$, clustered with the anti-k_T algorithm with $R = 0.7$. Events where the leading jet would be in the forward region are vetoed. The measurement is shown in Fig. 2 and $\alpha_S = 0.1150^{+0.0088}_{-0.0038}$ is extracted.

4 Dijet triple differential cross section at 8 TeV

The triple differential cross section of the di-jet production at 8 TeV consists in the following measurement [5]:

$$\frac{\mathrm{d}^3\sigma}{\mathrm{d}p_{T,\mathrm{avg}}\mathrm{d}y^*\mathrm{d}y_b} = \frac{1}{\epsilon \mathcal{L}_{\mathrm{int}}^{\mathrm{eff}}} \frac{N_{\text{di-jet events}}}{\Delta p_{T,\mathrm{avg}} \Delta y^* \Delta y_b} \tag{3}$$

where

- $p_{T,\mathrm{avg}} = \frac{1}{2}(p_{T,1} + p_{T,2})$ is the average transverse momentum of the di-jet system;
- $y_b = \frac{1}{2}|y_1 + y_2|$ is the *rapidity boost* of the di-jet system;
- $y^* = \frac{1}{2}|y_1 - y_2|$ is the *rapidity separation* of the di-jet system.

Anti-k_T jets with $R = 0.7$ are selected with $p_{T,\mathrm{jet}} > 50$ GeV, $|y_{\mathrm{jet}}| < 3.0$ and $p_{T,\mathrm{avg}} > 133$ GeV.

The different regions of the phase space are then exploited to extract the strong coupling and to constrain PDFs: the central region (small y_b and small y^*) is most suited for α_S extraction at high energy scales; in the boosted region (higher y_b but small y^*), the high-x region of PDFs can be better constrained; finally, in the region of large rapidity separation (small y_b but higher y^*), PDF and detector effects can be better disentangled.

The measurement is compared first to NLO parton-level calculations as well as predictions from MC event generators (Herwig 7 and Pythia 8). In general, predictions are slightly overestimated at high transverse momentum in the boosted region. Eventually, the di-jet data may be used to extract gluon PDFs. This is illustrated in Fig. 3.

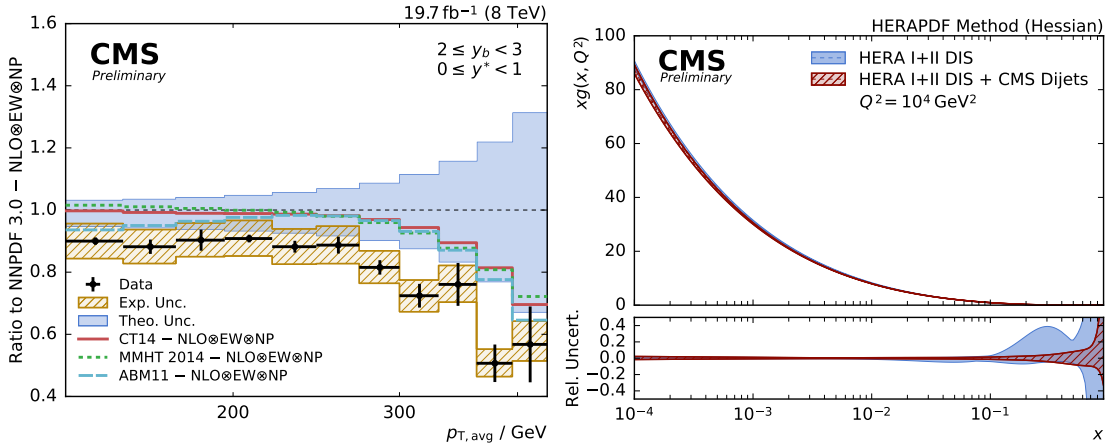


Figure 3: Left: di-jet triple differential cross section in the boosted region. Right: comparisons of the gluon PDF before and after inclusion of di-jet data [5].

5 Azimuthal correlations at 8 and 13 TeV

The azimuthal correlations allow to investigate higher-order QCD corrections: the more extra radiations, the less correlated the two leading jets.

CMS has published two measurements:

1. the measurement of the azimuthal correlations between the two leading jets at 8 TeV with a jet reconstructed with the anti-k_T algorithm of $R = 0.7$ [6];
2. the analog measurement at 13 TeV for a radius of 0.4 [7]; in addition, the same measurement is also performed with minimum jet multiplicities of 3 and 4 jets in order to be more sensitive to higher order effects; finally, correlations between the subleading and the closest third- or fourth-leading jet are measured, also in order to increase the sensitivity to parton showers.

Here, given the large amount of results reported in these analyses, we only show corresponding measurements at 8 and 13 TeV in Fig. 4.

6 Conclusions

In general, the QCD predictions describe well the measurements. With jet data, gluon PDFs can be significantly improved at high x. The value of α_S is also extracted. The azimuthal correlations in multijet events illustrate the importance of higher-order QCD corrections.

References

[1] CMS Collaboration, The CMS experiment at the CERN LHC, JINST, 3:S08004, 2008.

[2] CMS Collaboration, Measurement and QCD analysis of double-differential inclusive jet cross sections in pp collisions at $\sqrt{s} = 8$ TeV and cross section ratios to 2.76 and 7 TeV, JHEP, 03 (2017) [arXiv:1609.05331 [hep-ex]].

[3] CMS Collaboration, Measurement of the double-differential inclusive jet cross section in protonproton collisions at $\sqrt{s} = 13$ TeV, Eur. Phys. J. **C76**, 8 (2016)

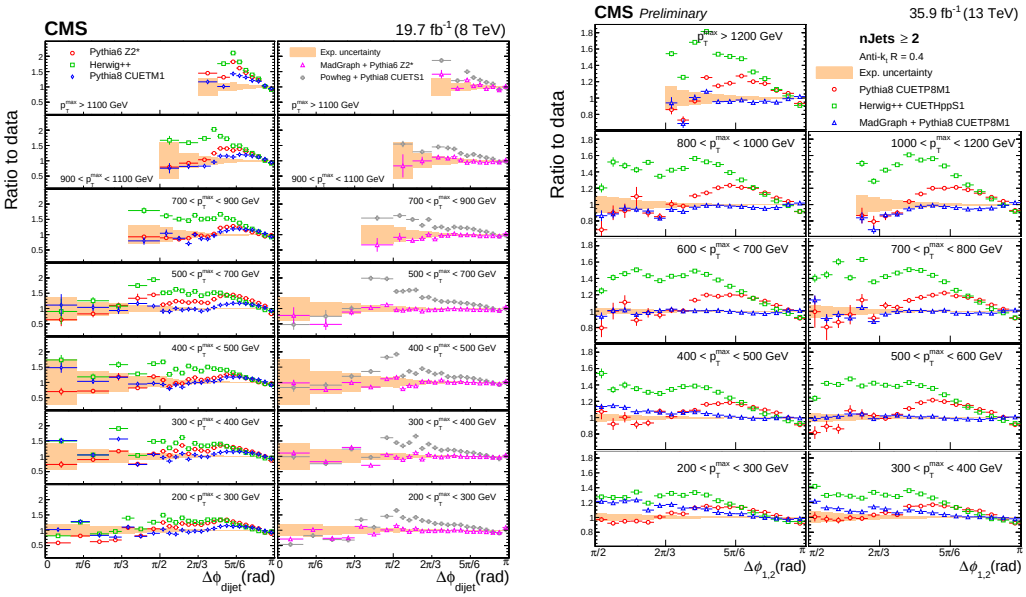


Figure 4: Analog measurements at 8 and 13 TeV (with respective cone size radii) for the azimuthal correlations [7].

[4] CMS Collaboration, Determination of the strong coupling constant from the measurement of inclusive multijet event cross sections in pp collisions at $\sqrt{s} = 8$ TeV, CMS-PAS-SMP-16-008 (CDS 2253091)

[5] CMS Collaboration, Measurement of Triple-Differential Dijet Cross Sections at $\sqrt{s} = 8\,\text{TeV}$ with the CMS Detector and Constraints on Parton Distribution Functions, CMS-PAS-SMP-16-011 (CDS 2235163)

[6] CMS Collaboration, Measurement of dijet azimuthal decorrelation in pp collisions at $\sqrt{s} = 8\,\text{TeV}$, Eur. Phys. J. **C76**, 10 (2016) [arXiv:1602.04384 [hep-ex]].

[7] CMS Collaboration, Measurements of inclusive 2-jet, 3-jet and 4-jet azimuthal correlations in pp collisions at $\sqrt{s} = 13$ TeV, CMS-PAS-SMP-16-014 (CDS 2257685)

[8] Cacciari, Matteo and Salam, Gavin P. and Soyez, Gregory, The Anti-k(t) jet clustering algorithm, JHEP, 04 (2008) [arXiv:0802.1189 [hep-ph]]

List of Figures

1.1 Bubble chamber print . . . 13
1.2 Comparison of Thomson's and Rutherford's models . . . 15
1.3 Artistic view of the content of the proton . . . 16

2.1 Strong running coupling . . . 30
2.2 HERA PDFs . . . 38
2.3 CMS event displays of tri- and four-jet final-state events . . . 39
2.4 Jet clustering algorithms . . . 41
2.5 Overview of a pp collisions . . . 43
2.6 Sketch of PS with colour flow . . . 47
2.7 Interleaved, stronly ordered Parton Shower . . . 49
2.8 Hadronisation process . . . 51
2.9 Hadronisation models . . . 51
2.10 Diagram of the evolution equations . . . 53
2.11 Young experiment with electrons . . . 56
2.12 Stern-Gerlach experiment . . . 56
2.13 Coordinates . . . 58

3.1 CERN logo . . . 67
3.2 CERN accelerator complex . . . 68
3.3 Phase space at LHC compared to previous experiences . . . 69
3.4 LHC vs. Geneva Airport . . . 71
3.5 Radiofrequency cavities . . . 72
3.6 Beam focusing at the interaction point . . . 73
3.7 Luminosity at CMS in 2016 . . . 76
3.8 Instantaneous luminosity . . . 77
3.9 CMS logo . . . 77
3.10 Picture of the CMS detector . . . 78
3.11 Transversal diagram of the CMS detector . . . 81
3.12 Pictures of silicon modules . . . 83
3.13 Illustration of Kalman filter . . . 85
3.14 Longitudinal view of the calorimeters . . . 87
3.15 Jet reconstruction . . . 87

4.1 Serial correlations of pseudo-random numbers . . . 95
4.2 Definitions of the different regions in tuning techniques . . . 99
4.3 $\hat{p}_t$ distribution from CMS PYTHIA 8 sample . . . 101

5.1 First measurement of Υ meson . . . 106
5.2 Υ peaks at CESR . . . 109

5.3 Event with b jet in pp collisions . . . 112
5.4 Production modes of b quarks . . . 115
5.5 Angular separation of two B hadrons . . . 116
5.6 b jet production at TEVATRON and CCFM evolution . . . 117
5.7 Illustration of flavour democracy . . . 117
5.8 Higgs branching ratios . . . 119

6.1 Contributions to the inclusive jet cross section . . . 128
6.2 Contribution to the inclusive b jet cross section . . . 129
6.3 PDFs in LHC conditions . . . 131
6.4 Content of the inclusive b jet signal . . . 132
6.5 ATLAS and CMS b jet cross section . . . 133
6.6 Ratio of ATLAS measurement with NLO predictions . . . 134
6.7 Ratio and fraction of CMS measurement with predictions . . . 135
6.8 Contributions to the p_T spectrum in slices . . . 138

7.1 Cross section of the simulation and data at detector level . . . 144
7.2 Ratio of the simulation to data at detector level . . . 145
7.3 Trigger efficiencies . . . 150
7.4 Check of the efficiency of the trigger . . . 151
7.5 Run-by-run cross section . . . 152
7.6 Run-by-run fraction of `CSVv2`-tagged jets . . . 153
7.7 Effect of the removal of the overweighted pile-up jet outliers . . . 158
7.8 Effect of the reweighting of the pile-up profile . . . 159
7.9 JEC levels . . . 161
7.10 Effect of JEC and JER on the p_T spectrum . . . 163
7.11 Resolution . . . 165
7.12 Sketches of discriminants . . . 167
7.13 Correlations between taggers . . . 168
7.14 Sketch of the b calibration . . . 169
7.15 b-tagging scale factors for tight selection . . . 172
7.16 Effect of the b calibration on inclusive b jet spectrum . . . 173
7.17 Performance of `CSVv2T` in simulation . . . 174
7.18 MET control plots . . . 176
7.19 Effect of MET fraction cut-off on the inclusive jet spectrum . . . 178
7.20 Effect of MET fraction cut-off on the inclusive b jet spectrum . . . 179
7.21 Effect of MET fraction cut-off on different processes . . . 180
7.22 Contributions of the different triggers to the total p_T spectrum . . . 182
7.23 Run-by-run average trigger pre-scales . . . 185
7.24 Jet $\eta - \phi$ spectrum in data (left) and simulation (right), for inclusive jet (top) and inclusive b jet (bottom). The bins corresponds to the count of jets (not to a cross section) with arbitrary normalisation. . . . 191
7.25 Additional resolution profiles . . . 192
7.26 Performance of `CSVv2` in simulation . . . 193
7.27 b-tagging scale factors for medium selection . . . 194
7.28 b-tagging scale factors for loose selection . . . 195
7.29 Effect of the b calibration for different working points . . . 196
7.30 Effect of the b calibration for different taggers . . . 197
7.31 Performance of b tagging for different working points . . . 198
7.32 Performance of b tagging for different taggers . . . 199

7.33 Jet constituent fractions . . . 200
7.34 Jet constituent multiplicities . . . 201

8.1 Sketch of correction on purity on discriminant . . . 205
8.2 Sketch of the JP templates . . . 206
8.3 Performance of b-tagging after additional correction . . . 207
8.4 Response matrix from PYTHIA 8 . . . 210
8.5 Migrations seen from detector level . . . 212
8.6 Closure test . . . 217
8.7 L-curve scan . . . 219
8.8 Ratio of data with simulation before unfolding . . . 220
8.9 Ratio of data and simulation after unfolding . . . 221
8.10 Comparison of the unfolding in different scenarios . . . 222
8.11 Description of the uncertainties . . . 223
8.12 D'Agostini unfolding . . . 225
8.13 Double differential JP discriminant for the different flavour templates in CSVv2-tagged sample . . . 228
8.14 Double differential JP discriminant for the different flavour templates in non-CSVv2-tagged sample . . . 229
8.15 Purity in simulation with PYTHIA 8 . . . 230
8.16 Purity in simulation with MADGRAPH . . . 231
8.17 Agreement of the JP discriminant before fit . . . 232
8.18 Agreement of the JP discriminant after fit . . . 233
8.19 Purity in simulation in the non-CSVv2-tagged sample . . . 234
8.20 Purity in simulation with MADGRAPH . . . 235
8.21 Agreement of the JP discriminant before fit . . . 236
8.22 Agreement of the JP discriminant after fit . . . 237
8.23 Regularisation matrix . . . 240
8.24 Tikhonov regularisation . . . 241
8.25 Input covariance matrix . . . 242
8.26 Output covariance matrix . . . 243
8.27 Backfolding . . . 244
8.28 χ^2s of agreement and of change . . . 245
8.29 Bottom Line Test . . . 246

9.1 Absolute cross section with LO predictions for inclusive jet production . . . 252
9.2 Comparison to LO predictions for inclusive jet production . . . 253
9.3 Absolute cross section with LO predictions . . . 254
9.4 Comparison to LO predictions . . . 255
9.5 Fraction of b jets with LO predictions . . . 256
9.6 Fraction ratio of b jets to LO predictions . . . 257
9.7 Theory uncertainties of the inclusive jet production . . . 258
9.8 Theory uncertainties of the inclusive b jet production . . . 259
9.9 Absolute cross section with NLO predictions . . . 262
9.10 Comparison to NLO predictions . . . 263
9.11 Absolute cross section with NLO predictions . . . 264
9.12 Comparison to NLO predictions . . . 265
9.13 Fraction of b jets with NLO predictions . . . 266
9.14 Fraction ratio of b jets to NLO predictions . . . 267
9.15 Contributions in POWHEG + CUETP8M1 for inclusive jet production . . . 268

9.16 Contributions in POWHEG + CUETP8M1 for inclusive b jet production 269

10.1 Tagging with hits 275

A.1 Track reconstruction in the context of alignment 280
A.2 Structure of the tracker 281
A.3 Ladders and blades in the pixel tracker 281
A.4 MILLEPEDE-II logo 282
A.5 Weak modes 284
A.6 Lorentz drift 286
A.7 Geometry comparison plots for 2016 Legacy Alignment 287
A.8 Distributions of the medians of the residuals in pixel tracker 289
A.9 Distributions of the medians of the residuals in strip tracker 290
A.10 Correction of Lorentz drift for 2016 Legacy Alignment 291
A.11 Picture of primary-vertex validation 292
A.12 Primary-vertex validation for 2016 Legacy Alignment 293
A.13 $Z \to \mu\mu$ validation for the 2016 Legacy Alignment 295

B.1 Logo of the DIS 2016 conference 297

C.1 Logo of the LHCP 2017 conference 303

List of Tables

1.1 Scales, units and objects . . . 14

2.1 Fields in the Standard Model . . . 24
2.2 Parameters of the SM . . . 26
2.3 Contributions to the cross section in α_S . . . 34
2.4 Number of diagrams for final state in e^+e^- collisions . . . 35
2.5 Decay lengths of particles . . . 52

3.1 CMS key figures . . . 78
3.2 Pseudorapidity coverage at CMS . . . 80
3.3 Pixel and strip modules . . . 82
3.4 Geometry of the calorimeters . . . 86

4.1 Examples of tuned parameters . . . 98

5.1 B mesons . . . 108
5.2 Main experiments and contributions to b physics . . . 110

6.1 Phase space and selection . . . 134
6.2 Luminosities per era . . . 135
6.3 Simulation samples to perform the analysis . . . 136
6.4 Sub-samples in PYTHIA 8 and MADGRAPH . . . 139

7.1 Trigger turn-ons, luminosities and number of entries per trigger . . . 148
7.2 Binning of the training of the `CSVv2` tagger. . . . 166
7.3 Turn-on points for trigger efficiency . . . 181
7.4 Number of jets per trigger and per era . . . 183
7.5 Jet ID definition . . . 188

9.1 Tune variations . . . 260
9.2 Tune parameters . . . 270

A.1 Structure of the tracker . . . 280
A.2 Statistics and balance for 2016 Legacy Alignment with MILLEPEDE-II . . . 288

Acronyms

Geant4 GEometry ANd Tracking. 100

NLOJet++ NLO Jet in C++. 97

Pluto after the magnet of the detector[1]. 107, 110

Professor PROcedure For EStimating Systematic errORs. 98

Zeus in reference to the relation of Zeus and Hera in the mythology. 15, 37, 110, 111

CSVv2 Combined-Secondary-Vertex. 4, 5, 133, 149, 166, 167, 170, 174, 187, 193, 196, 198, 199, 204–206, 218, 222, 226, 227, 230–237, 244, 310, 311, 313

CUETHppS1 CMS UE Tune HERWIG++ Set 1. 99, 136, 249, 252–257

CUETP8M1 CMS UE Tune PYTHIA 8 Monash 1. 99, 132, 136, 249, 251–260, 262–270, 311, 312

JP Jet Probability. 166–168, 187, 204, 206, 226, 232, 233, 236, 237, 311

NaN Not a Number. 226

PXB PiXel Barrel. 280, 281

PXF PiXel Forward. 280, 281, 292

TEC Tracker End-Caps. 280, 281

TIB Tracker Inner Barrel. 280, 281

TID Tracker Inner Disk. 280, 281

TOB Tracker Outer Barrel. 280, 281

cMVAv2 combined-Multi-Variate-Analysis. 4, 166, 167, 170, 218, 226

ABM after the three physicists ALEKHIN, BLÜMLEIN and MOCH. 37

ALICE A Large Ion Collider Experiment. 35, 71, 75, 110

ARGUS A Russian-German-United States-Swedish collaboration[2]. 110

ATLAS A Toroidal LHC ApparatuS. 4, 41, 71, 73, 75, 77, 110, 130, 133, 140, 249, 310

AVF Adaptive Vertex Fitter. 84

AVR Adaptive Vertex Reconstruction. 84, 166, 168

BaBar BB. 110

BBR Beam-Beam Remnants. 42, 115

BEH after the three physicists BROUT, ENGLERT and HIGGS. 23, 24, 67

Belle French word for beauty. 109, 110

1. Aachen-DESY-Hamburg-Siegen-Wuppertal-München collaboration
2. Though joined later by other countries.

BFKL after the four physicists Balitsky, Fadin, Kuraev and Lipatov. 2, 52, 53

BLT Bottom Line Test. 5, 208, 245, 246

Booster Proton Synchrotron Booster. 68, 73

BSM searches Beyond the SM. 17, 70, 97, 118

c.m.s. centre-of-mass. 60

CASTOR Centauro And STrange Object Research in nucleus-nucleus collisions at the LHC. 75

CCFM after the four physicists Catani, Ciafaloni, Fiorani and Marchesini. 2, 53, 98, 116

CDF Collider Detector at Fermilab. 109, 110

CERN European Organisation for Nuclear Research[3]. vi, 67, 68, 71–73, 90, 106, 110

CESR Cornell Electron-positron Storage Ring, pronounced "Caesar". 108, 110

CKM after the three physicists Cabibbo, Kobayashi and Maskawa. 25, 26, 108

CLEO *Cleopatra* (for her/its proximity with Caesar). 108–110

CMS Compact Muon Solenoid. v, vi, 4, 17, 29, 36, 39, 41, 50, 58, 67, 71, 74–80, 82, 84–86, 88–90, 96, 100, 110–112, 115, 116, 119, 120, 127, 130, 133–135, 140, 155, 156, 160–162, 166, 168, 175, 177, 203, 249, 260, 273–275, 279, 280, 282, 283, 285, 286, 294, 310

CSC Cathode Strip Chamber. 88, 89

CT Closure Test. 208, 216, 217

CTEQ Coordinated Theoretical-Experimental Project on QCD. 37

CUSB Columbia University-Stony Brook. 108–110

DASP2 Double Arm SPectrometer[4]. 107, 110

DESY *Deutsches Electroknen-Synchrotron*. v, vi, 14, 83, 107, 110, 279, 282, 283

DGLAP after the five physicists Dokshitzer, Gribov, Lipatov, Altarelli and Parisi. 37, 42, 44, 46, 51–53

DHHM DESY-Hamburg-Heidelberg-München collaboration. 107, 110

DIS Deeply Inelastic Scattering. 36, 37, 60

DORIS *DOppel-RIng Speicher*. 107, 108, 110

DPS Double-Parton Scattering. 46

DT Drift Tube. 88, 89

DY Drell-Yan. v, 36, 177

DØ after the location of the detector. 109, 110, 117

e.m. electromagnetic. 188, 321, 323

EB ECAL Barrel. 86

ECAL Electromagnetic CALorimeter. 77, 80, 85–90, 160

EE ECAL End-caps. 86

3. Originally *Conseil Européen pour la Recherche Nucléaire*, kept for the proximity with the Germanic root *kern*, meaning "nucleus".

4. DESY-Dortmund-Heidelberg-Lund

EOM Equations of Motion. 27, 57, 82
EW Electroweak. 27, 28, 97

FCR Flavour Creation. 113, 115, 127, 129
FEX Flavour Excitation. 113, 115, 116, 127, 129
FF Fragmentation Function. 44, 50, 251
FO fixed order. 96
FSR Final-State Radiation. 33, 42, 44, 46, 97, 98, 260

GPMC General-Purpose MC Event Generator. 42, 97, 98
GSP Gluon Splitting. 114, 115, 127, 129, 170

H1 HERA-1. 15, 37, 110, 111
HB HCAL Barrel. 86
HCAL Hadronic CALorimeter. 77, 80, 86–90
HE HCAL End-caps. 86
HEP High-Energy Physics. 11–13, 16–18, 21, 25, 29, 54, 58, 75, 93, 96, 98, 100
HERA *Hadron-Elektron-RingAnlage*. 15, 37, 38, 69, 110, 309
HERA-B HERA Beauty. 109, 110
HERAPDF HERA PDF sets. 37
HF Heavy Flavour. 26, 44, 50, 98, 108, 111, 112, 155, 160, 226, 274
HF HCAL Forward. 86, 87
HLS High-Level Structure. 285, 286
HLT High-Level Trigger. 90, 146, 148
HO HCAL Outer. 86

IC-PR Iterative Cone algorithm with Progressive Removal procedure. 40
IC-SM Iterative Cone algorithm with Split Merge procedure. 40
IP interaction point. 73–75, 79, 81, 112, 133
IRC infrared and collinear safe. 40
ISOLDE Isotope Separator On Line DEvice. 73
ISR Intersecting Storage Rings. 106, 110
ISR Initial-State Radiation. 33, 41, 44–46, 97, 98
IVF Inclusive Vertex Fitter. 84, 166, 168

JEC Jet Energy Correction. 136, 146, 155, 160, 161, 163, 164, 170, 192, 218
JER Jet Energy Resolution. 4, 161, 163, 164, 170, 192

KEK Koo Energy Ken. 110
KEKB KEK Beauty. 110

L1 Level 1. 90
LEIR The Low Energy Ion Ring. 68

LEP Large Electron Positron Collider. 67, 99, 110, 117, 167, 260

LHC Large Hadron Collider. v, vi, 2, 3, 22, 29, 35, 36, 38, 41, 54, 67–77, 89, 93, 99, 105, 109–112, 119, 120, 127, 130, 131, 154, 184, 260, 273–275, 310

LHCb LHC beauty. 71, 75, 110

LHCf LHC forward. 75

LINAC 2 Linear Accelerator 2. 68, 71–73

LINAC 3 Linear Accelerator 3. 68, 73

LLN Law of Large Numbers. 94, 95

LO Leading Order. 5, 34, 42, 44, 96, 97, 136, 140, 249, 252–257, 273, 311

LS Lumi Section. 75, 155

Mark (unknown origin). 110

MB Minimum Bias. v, 75, 98, 99, 147, 155, 157, 159, 283, 284, 288, 291

MC Monte Carlo. v, 17, 35, 43, 93, 96, 97, 100, 143, 149, 154–156, 160–162, 168–170, 175–179, 208, 209, 216, 218, 219, 224, 225, 238, 239, 250, 252–257, 262–267

ME Matrix Element. 21, 22, 29, 31–34, 37, 41, 42, 44, 48, 59, 96–98, 127–129, 136, 156, 274

MET Missing Transverse Energy. 4, 77, 89, 143, 175–180, 310

MMHT after the names of the four physicists Harland-Lang, Martin and Motylinski, Thorne. 37

MoEDAL Monopole and Exotics Detector At the LHC. 75

MPI Multi-Parton Interaction. 42, 46, 48, 49, 97–99, 136, 140, 211, 251, 260, 273, 274

MSTW after the names of the four physicists Martin, Stirling, Thorne and Watt. 37

MVA Multi-Variate Analysis. 111, 167

n.d.f. number of degrees of freedom. 232, 233, 236, 237

NLO Next to Leading Order. 5, 34, 37, 38, 42, 44, 96, 97, 117, 130, 140, 249, 251, 262–267, 273, 310, 311

NNLO Next to Next to Leading Order. 34, 37, 38, 44, 96

NNPDF Neural-Network PDF. 37

p.d.f. probability density function[5]. 93–96, 100, 101

PDF Parton Distribution Function. 37–39, 42–46, 49, 50, 52, 53, 96, 114, 120, 131, 132, 136, 140, 251, 260, 274, 309, 310, 318

PEP Positron-Electron Project. 108–110

PF Particle-Flow. 89, 146, 148, 154, 188

PM Probability Matrix. 208, 209, 213, 214, 239

pQCD perturbative QCD. 35

PS Proton Synchrotron. 73

PS Parton Shower. 2, 41, 44–48, 50, 51, 68, 96–99, 101, 114–116, 127–129, 136, 140, 251, 260, 273–275, 309

5. Not to be confused with Parton Distribution Function.

PU pile-up. 40

PV Primary Vertex. 82, 84, 112, 133, 134, 166, 167, 288, 291–293

QCD Quantum Chromodynamics. 2, 21, 24–26, 28–31, 33, 35, 39, 44, 46, 50, 53, 57, 96, 97, 111, 115, 116, 127, 130, 136, 146, 155, 161, 273, 274

QED Quantum Electrodynamics. 24, 25, 27–29, 31, 32

QFT Quantum Field Theory. 25, 28

RM Response Matrix. 137, 140, 208–211, 213, 214, 216, 238

RPC Resistive Plate Chamber. 88

SF Scale Factor. 162, 164, 168–170, 187, 194–197, 204

SIScone Seedless and Infrared Safe Cone. 40

SLAC Stanford National Accelerator Laboratory. 14, 57, 110

SLC SLAC Linear Collider. 110

SLD SLAC Large Detector. 110, 260

SM Standard Model. 2, 12, 17, 21–28, 34, 35, 57, 70, 109, 111, 118, 120, 130, 177, 180, 273, 313

Sp$\bar{\text{p}}$S Super Proton – Antiproton Synchrotron. 110

SPS Super Proton Synchrotron. 68, 71, 73

SV Secondary Vertex. 82, 84, 112, 130, 133, 166, 167, 279

SVD Singular Value Decomposition. 214

TMD Transverse-Momentum-Dependent PDF. 53

TOTEM TOTal cross section, Elastic scattering and diffraction dissociation Measurement at the LHC. 75

UA1 Underground Area 1, named after the location of the detector. 110

UE Underlying Event. 40, 48, 97–99, 117, 136, 157, 270, 274

WM Weak Mode. 283–285

WP Working Point. 167–171, 187, 196–198

Index

B factory, 109
B meson, 108, 109, 111, 112, 115, 118, 167
D meson, 111
L-curve scan, 216, 218
$W^{\pm}$ boson, 27, 33, 105, 118
Z^0 boson, 27, 33, 105
Υ meson, 105, 107–109, 112
b jet, 111, 116, 118, 119, 130, 168, 273
b quark, 111, 113, 115, 118, 127
b tagging, 77, 78, 80, 87, 111, 133, 137, 166, 187, 188, 275
b quark, 15
c jet, 111, 168, 206, 273, 274
k_T algorithm, 40
s-channel, 33
t channel, 33
z rescaling, 284
HERWIG, 97
MADGRAPH, 96
POWHEG, 96
PYTHIA, 97
SHERPA, 97
`RANLUX`, 95

absolute energy scale, 160
acceleration, 71
action, 27
alignment, 85
angular ordering, 46, 53
anti-k_T algorithm, 41
anti-colour, 25
anti-quark, 25, 28
asymptotic freedom, 29, 35, 37
atomism, 14
Avogadro number, 14
azimuthal correlations, 274

backfolding, 239, 245
background, 177, 208, 273
bad jet, 156
baryonic asymmetry, 17
baryons, 38
Bayesian unfolding, 219
beam remnants, 42, 97
bin width, 137, 209
boost, 23
bottomonium, 107
Bowler factor, 50, 251
Bremsstrahlung, 79
bunch splitting, 72, 73

calculations, 22, 29, 33, 35, 54
calorimeter, 80, 85
Cambridge-Aachen algorithm, 41
Cascade, 98
causality, 23
celerity, 11
charge conservation, 23, 25
charged e.m. energy fraction, 188
charged leptons, 27
charged-hadron energy, 188
charged-hadron multiplicity, 188
charmonium, 106, 107
circular collider, 70
cluster model, 50
clustering, 72
CMS, 17, 273
collider, 69
collinear factorisation, 37, 51
collinear radiation, 33
collision rate, 89
colour, 25, 28, 50
colour reconnection, 48, 98
composite particle, 13
Compton effect, 79, 87
cone algorithm, 40
confinement, 35, 38
congruential generator, 95
contamination, 170, 203, 204, 206

Cornell potential, 107
cosmic ray, 16
cosmic rays, 69
coupling, 25, 28
covariance matrix, 214, 238, 245
CP violation, 26, 75, 106, 108
cross section, 16, 21
cross-section area, 74
curvature, 80
cyclotron, 72

D'Agostini regularisation, 219, 224, 239
dark energy, 17, 18
dark matter, 17, 18
decay, 100
decay length, 16, 21
deep learning, 274
detector level, 100, 132, 143, 208
detector response, 100
diffractive dissociation, 36
Dirac spinor, 23
down-type quarks, 27

effective luminosity, 149
efficiency, 168–170, 203
elastic cross section, 74
elastic scattering, 36
electric charge, 11, 29
electromagnetism, 12, 27
electron, 14
electron interactions, 79
electron multiplicity, 188
electron-Volt, 11
electroweak interaction, 27, 130
era, 135, 180
error function, 147
event display, 39
event reconstruction, 79
evolution, 21, 116
exclusive cross section, 33
exclusive method, 146
expectation value, 94

factorisation, 36, 37, 42, 44, 46
factorisation scale, 37, 42
fake jet, 211
fake jets, 211
Feynman parton model, 15
field, 23
fine structure constant, 28, 29
fixed-order calculations, 96, 98
flavour, 15
flavour blindness, 28
flavour democracy, 28, 116, 274
flavour scheme, 44, 113
focusing, 72
fraction, 170
fraction ratio, 170, 187
fragmentation, 170
fundamental interactions, 12
fundamental parameters, 17, 26, 28, 34
fundamental particle, 13

Gauge invariance, 22, 23
Gauge symmetry, 25
Gauge transformation, 28
Gell'mann quark model, 15
generator, 96, 136
GIM mechanism, 106
gluon, 25, 28
golden-foil experiment, 14
Golden Channel, 119
good event, 133
good primary vertex, 133
gravitation, 12, 17

hadron, 29, 38, 50, 70
hadron decay, 42, 97
hadronisation, 42, 44, 49, 50, 97, 117, 118, 170, 260
hidden beauty, 106
Higgs boson, 26, 28, 69, 118, 119
Higgs physics, 118, 119
Higgs sector, 28
hit-or-miss integration, 101
hybrid method, 162

impact parameter, 112, 166, 167
importance sampling, 95
inclusive cross section, 33
inelastic non-diffractive scattering, 36
injector, 72
instantaneous luminosity, 74, 75, 155, 184
integrated luminosity, 74, 135
interaction, 11, 21
interaction Hamiltonian, 21, 31
interaction term, 27, 28
interference terms, 34
ionisation, 79

jet, 36, 39, 160, 186

jet clustering algorithm, 39, 130
jet ID, 133, 188
joined interactions, 48

Kalman filter, 82
kinematic term, 27

Lagrangian, 27–29, 31
lattice QCD, 35
lead, 70
lepton, 27
LHC Run-II, 41, 70
light jet, 111, 168, 273
line, 30
linear collider, 70
logistic function, 147
loop diagram, 33, 34
loose muon, 89
luminosity, 74, 137, 149, 154

Mandelstam variables, 59
mass, 23, 25
mass term, 27
matching, 209
matter, 11
Maxwell equations, 70
MC calibration, 160
mesons, 38
migrations, 211
misidentification, 168–170, 203, 206
miss jet, 211
missing energy, 79, 175
mistag, 168, 170, 203, 204, 206
momentum fraction, 37, 42, 46, 50, 53, 60
momentum sum rule, 39
momentum transfer, 59
Mueller-Navelet jets, 53
multi-count observable, 140, 169, 214
multiple scattering, 79, 84, 100
muon chambers, 80, 88, 133
muon energy fraction, 188
muon interactions, 79
muon multiplicity, 188

neural network, 166
neutral e.m. energy fraction, 188
neutral-hadron energy, 188
neutral-hadron multiplicity, 188
neutrinos, 25, 27
non-perturbative regime, 29, 35, 50
nuclear interactions, 79
nucleon, 14

observable, 16, 21
off-shell, 34, 98
offset correction, 160
on-shell, 34
open beauty, 106

pair production, 79
particle, 12, 23
particle level, 100, 132, 203, 208
parton, 15
perturbative regime, 28, 29, 42, 44, 46, 50
phenomenology, 36
photoelectric effect, 79
photon, 24, 27
photon interactions, 79
photon multiplicity, 188
pile-up, 74, 131, 133, 136, 149, 154, 160, 170, 275
Planck constant, 11
pomeron, 36
pre-scale, 146, 149, 180
precision measurement, 69, 131
precision measurements, 275
Principle of Least Action, 27
propagator, 30
pseudorapidity, 58
purity, 170, 203, 211, 226

QFT, 23
quantum field, 23
quantum field theory, 22, 23, 28
quark, 15, 25, 27, 28
quark-gluon plasma, 35, 70, 75

radioactive element, 16
radiofrequency cavity, 71
random number, 93
random variable, 93
rapidity, 58, 137
real, 30
recombination algorithm, 40
reference trigger method, 147
regularisation matrix, 214, 215, 238
regularisation parameter, 214, 216, 218
relative energy scale, 160
renormalisation, 22, 25, 29
renormalisation scale, 25
replica, 251
residual corrections, 160

resolution, 84, 87, 161, 208
resonance cavity, 71
run, 75, 135
Rutherford atom, 14

sagitta, 284
scale invariance, 29
scale variation, 170
scale violation, 29
scaling method, 162
scattering experiment, 16
scintillator, 86, 87
sea, 39
search, 69, 119
searches, 118, 119
sectors, 27
self-interactions, 25, 28
sigmoid function, 147
smearing scale factors, 162
soft activity, 99
soft lepton, 112, 130, 166, 167
soft muon, 89
soft photon radiation, 42, 97
spectroscopy, 108, 109
speed of light, 11
spin, 23
spinor, 23
splitting functions, 42
Standard Model, 17
steering, 72
step hunting, 149
stochastic method, 162
stratified sampling, 100
string model, 50
strong coupling, 29, 34, 116, 132
strong nuclear interaction, 12, 28, 38, 130
strong ordering, 46, 53
subtraction method, 101
Sudakov factor, 45, 46, 96, 101
synchrotron, 72

Tag & Probe method, 147
tagging, 111
Tag & Probe method, 147
tauon, 106
telescope, 284
template fit, 204, 226
Thomson atom, 14
Tier 0, 90
tight muon, 89
Tikhonov, 214
Tikhonov regularisation, 208, 224, 238, 239
time stability, 149, 160
top physics, 118
top quark, 118
total cross section, 74
tracker, 79, 133, 275
tracking, 82
transverse momentum, 137
tree level, 33–35, 59, 113
trigger, 89, 137, 146
trigger emulation method, 147
tune, 98, 99, 117, 251, 260
tuning, 98, 132
turn-on point, 146, 149, 180
twist, 37, 284

ultraviolet divergence, 25
uncertainties, 75, 137, 156, 157, 160, 164, 169, 206, 213, 216, 218, 224, 238, 251, 260
unfolding, 132, 136, 137, 154, 186, 203, 208, 213, 238
up-type quarks, 27

valence quark, 16, 38
variance, 94
vertex, 30, 31, 82, 112, 133, 154, 274
virtual, 30, 33, 34, 42
Voigtian function, 292

weak mode, 283
weak nuclear interaction, 12, 27
Weyl spinor, 23

Yukawa sector, 28

Acknowledgements

First of all, my gratitude goes to Hannes Jung, who supervised my work during three years and a half. Hannes, this time was really grand: always reachable, always supporting, always positive. Besides physics, you also taught me how to behave in a group (science, as I mentioned in the preamble of this work, is also made of social interactions). This work together will for sure have a great impact in my whole life.

Many thanks to Radek, Paolo and Benoît who, as post-docs in the group, largely contributed to my work; now it is my turn to take over. Ola, Armando, Daniela, Juan, Jindrich, Samantha, Svenja, Engin: thanks for all the good time spent together! (And by the way: thanks for the hat!)

But DESY is not limited to the QCD group. First, I spent also a lot of time in the alignment group: Matthias, Rainer, Nazar, Valeria, Gregor, Rostyslav, Chayanit, Claus, all of them have also contributed actively to my education. The DESY staff members, also, have always kindly helped me even on short notice: Achim, Stefan Schmidt, Olaf, Katerina, Christophe, Matthias Kasemann; for their time and help, I want to thank them too.

Merci à James, mon scientifique irlandais favori, qui a relu mon anglais dans quelques chapitres de cette thèse.

Grazie a Matteo, il mio personale esperto del b tagging, chi a avuto la buon'idea di lavorare nello stesso ufficio accanto a me. Thanks also to Kirill Skovpen, Ivan Marchesini and Caroline Collard, for their precious help in b tagging.

Au CERN, merci à Olivier Bondu et Gaëlle Boudoul, en particulier, qui m'ont bien aidé lors de mes premiers pas dans le monde du Tracker DQM.

In Helsinki, thanks to Mikko Voutilainen and Tapio Lampén, who helped me in their respective fields.

Hartelijk dank, Hans: elke keer is het weer tof, samen te diskuteren.

Even though from ATLAS, I would like to thank Nataliia Zakharchuk and Jihyun Jeong for their friendship, as well as So Young Shim and Shruti Patel from the theory group.

Ich bedanke mich auch bei Birgit & Gabriele, die immer so hilfbereit gewesen sind, und bei den Pförtnern, die immer sehr nett waren, als ich den Schlüßel des Flügels abholen kam.

Thanks to Samantha, who, first, helped me in the quest of a flat in Hamburg. Merci Marisa, pour l'ex-future colocation! Tu auras été mon premier contact à Hambourg en dehors de DESY. Danke, Martin, für die tolle WG in Eimsbüttel. Danke, Nils, Karin, Nora, Ruth, Marc-Ol für die WG in Altona. Je remercie la famille Bolliger, à commencer par

Aline qui m'a obtenu une chambre chez ses parents à Genève pour trois mois. Und danke, Annette & David für die WG in Bahrenfeld.

Ohne die Musik wären die Arbeit und das Leben nicht so schön: Deswegen bedanke ich mich auch bei meinem Chor *hamburgVOKAL* und bei meinen Musiklehrern Rainer und Michèle. Danke, auch, an Xaver, der mir umsonst Musiktheorie unterrichtet hat.

Außer der Musik habe ich auch viele Energie beim ETV Judo verwendet: Danke Mana, Stefan, Klara, Jawid und Eki!

Je ne pourrais pas me permettre d'oublier mon promoteur de mémoire, Laurent Favart, sans qui il n'y aurait eu ni Hambourg ni DESY pour moi. À l'IIHE, merci à David, Hugo, Isabelle, Hugues, Nastja, ... De l'*Université libre de Bruxelles*, deux professeurs en particulier ont contribué au dévelopement de mon intérêt pour la physique fondamentale: Thomas Hambye & Pierre Gaspard.

Remontant encore plus loin dans le temps, merci à Martin Casier, qui a déclenché la première étincelle d'intérêt pour la physique. De tout coeur, merci à Antoine Hoeffelman ainsi qu'à tous les membres de l'*Astroguindaine*. Hervé, quel plaisir de t'avoir vu à Hambourg pour ma défense.

Je pense aussi à mes professeurs du *Lycée Émile Jacqmain*, en particulier à Paul Van Langenhoven & Rony Vindelynckx qui, plus que tout autre, m'ont appris à structurer un texte. Leur trace dans ce volume est indéniable.

Bien entendu, mes derniers remerciements, les plus sincères également, vont à ma famille, surtout mes parents. Et merci à toi, ma Qun, du plus profond de mon coeur, pour avoir rejoint mes jours terrestres. 我爱你，王群！