

**First measurements of the  $t\bar{t}$  cross section in LHC pp  
and pPb collisions at 5.02 and 8.16 TeV and  
determination of the absolute luminosity in the  
CMS experiment**

*Paving the way for detailed investigations of top quark production in  
nuclear interactions*

---

Doctoral dissertation presented by

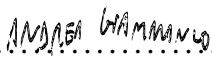
**GEORGIOS KONSTANTINOS KRINTIRAS**

in fulfillment of the requirements for the degree of Doctor in Sciences

**Doctoral advisor:** Andrea Giammanco  
UCLouvain

Approved by the committee composed of the members

(Signature)

.....  .....

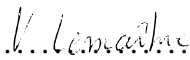
Andrea Giammanco  
UCLouvain

(Signature)

.....  .....

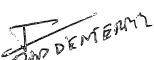
Christophe Delaere  
UCLouvain

(Signature)

.....  .....

Vincent Lemaître  
UCLouvain

(Signature)

.....  .....

David Enterría  
CERN

(Signature)

.....  .....

Gabriella Pásztor  
Eötvös Loránd University

Louvain-la-Neuve, January 2019






Université catholique de Louvain  
Secteur des Sciences et Technologies  
Institut de Recherche en Mathématique et Physique

Copyright © - All rights reserved. Georgios Konstantinos Krintiras, 2019.

No part of the material protected by this copyright notice may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage and retrieval system, without citing the primary source. No prior written permission or verbal consent from the author is required.

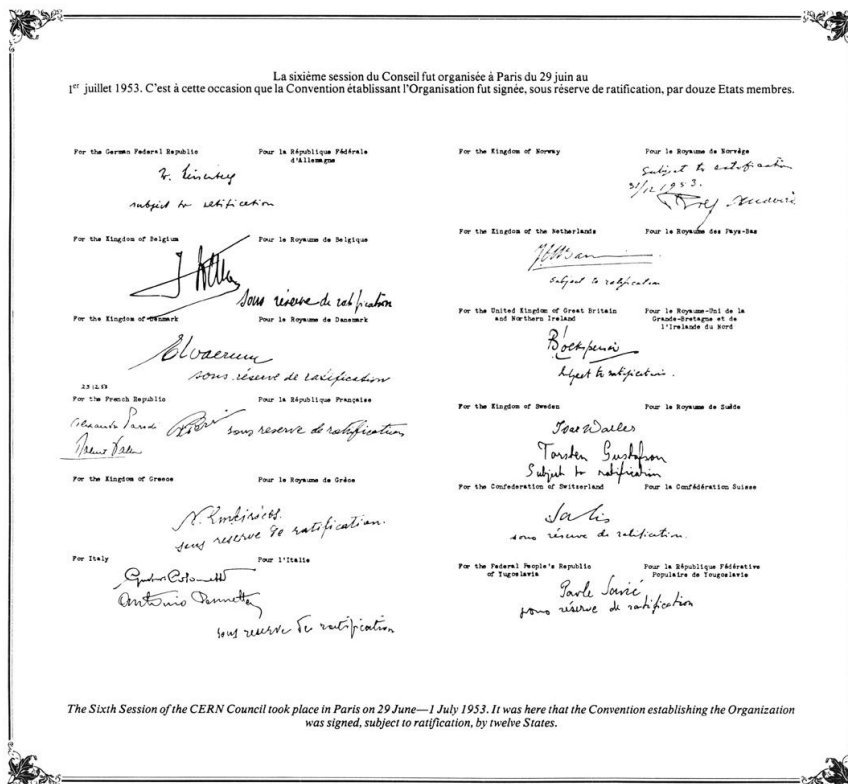
Printed in the Ciaco imprimerie. 

**Available Online:** <https://inspirehep.net/author/profile/G.Krintiras.1>

The research was initially supported by a grant from the Research Council of the University of Louvain (FSR). It was carried out within the Doctoral Commissions of Domains (CDD) that have drawn up special provisions approved by the Doctoral Commission of the University of Louvain (CODOL) to ensure consistency with the doctoral regulations. The special provisions apply exclusively to doctoral students in the Sector of Sciences and Technologies (SST).

*(Signature)*

*G.K. Krintiras*  
.....  
Georgios Konstantinos  
Krintiras



*The sixth session of the CERN Council took place in Paris, 29 June-1 July 1953. The monumental edition of the Convention establishing the European Organization for Nuclear Research was signed by 12 States, including Belgium and Greece.*



The Large Hadron Collider (LHC) at CERN is a 26.7 km two-ring, synchrotron accelerator and collider that initially approved for construction as a “missing magnet machine” in two stages and routinely started its operation in November 2009. By colliding high- and lower-intensity proton and lead—recently also xenon—beams with momentum of up to 6.5 and 6.5  $Z$  TeV, LHC has set new world records at the beginning of 2015 and the end of 2016, respectively. The achieved particle momenta correspond to unrivaled regimes in terms of the stored beam energy in both the proton and heavy ion programs, reaching values of more than 310 MJ. To bend and focus such rigid beams LHC is equipped with over one thousand superconducting magnets, most of which are operated at temperatures as low as 1.9 K, which can quench, if tiny fractions of the stored beam energy are deposited inside their coils. This hence puts high demands on the collimation system, which so far provided excellent cleaning with proton beams, whereas the cleaning performance has been sufficient for the heavy ion operation, characterized by high production yield of effectively off-momentum ion fragments.

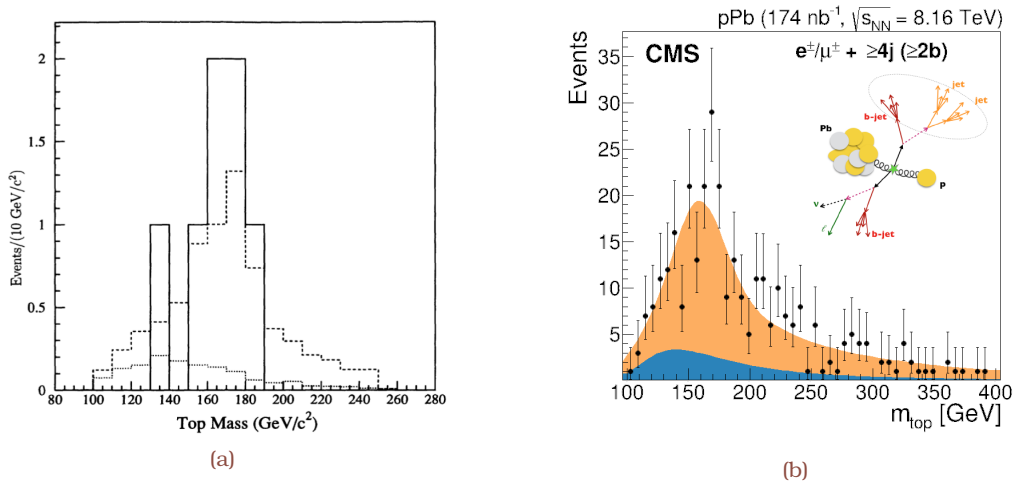
In most years the LHC apparatus is reconfigured for a month-long heavy ion run. However, asymmetric collisions were not included in the LHC design, and hence the physics case was based on a luminosity of  $1.15 \times 10^{29} \text{ cm}^{-2}\text{s}^{-1}$  at a beam energy of 7  $Z$  TeV (“design” parameters). Following up on a feasibility test and pilot physics Fill in October 2011 and September 2012, respectively, the first one month-long run took place in January 2013, meaning asymmetric proton-nucleus collisions remain a novel mode of operation at LHC. The 2015 operational period with heavy ions started with a “reference” proton run at 2.51 TeV to obtain the same center-of-mass energy as in the proton-nucleus run of 2013. For the same reason, the ensuing PbPb operation in November–December 2015 was carried out at an energy of 6.37  $Z$  TeV. The second proton-nucleus collision run in November–December 2016 offered the tremendous opportunity to answer a range of crucial physics questions, yet opening up the possibility to measure, for the first time in heavy ion collisions, various large-mass elementary particles, like the top quark. Despite the complex strategy for repeated recommissioning and operation of LHC, the peak luminosity surpassed the design value by a factor 7.8, and the amount of integrated luminosity substantially exceeded the requests of the majority of the LHC experiments.

The Compact Muon Solenoid (CMS) is one of the seven experiments at LHC, featuring a superconducting solenoid of 3.8 T. Within the solenoid volume are a silicon pixel and strip tracker, a lead tungstate crystal electromagnetic calorimeter, and a brass and scintillator hadron calorimeter, while forward calorimeters extend the pseudorapidity coverage up to  $\pm 5.2$ . Muons are detected in gas-ionization chambers embedded in the steel flux-return yoke outside the solenoid. CMS is therefore particularly suited for a global event description that aims to reconstruct and identify most of the produced particle types (photons, electrons, muons, charged and neutral hadrons), with an optimized combination of information from the various subdetectors. Events of interest are selected in real time using a two-tiered trigger system which reduces the event rate from the bunch crossing frequency

to around 1 kHz before data storage.

For almost all measurements performed at LHC, one crucial ingredient is the precise knowledge about the integrated luminosity. Despite being a key parameter in any particle collider, the task of calibrating its absolute scale has been proven particularly challenging at hadron colliders. The determination and precision of the integrated luminosity have direct implications on cross section measurements, and its instantaneous measurement gives essential feedback on the conditions at the experimental insertions and the accelerator performance. To determine the absolute luminosity dedicated beam-separation techniques are used, the so-called “van der Meer scans.” With the exception of the reference proton and heavy ion runs, these scans are not typically performed during normal physics operation, but under carefully controlled conditions and with beam parameters tailored to achieve the desired precision of  $O(2-4\%)$ . More recently, the advent of vertex-based “beam-imaging” techniques opened up interesting perspectives.

Soon after the discovery of the bottom quark [1], the quest for the top quark had begun. The search carried out for nearly 20 years because the mass of the top quark turned out to be unexpectedly large, around 40 times the mass of the bottom quark. The first study from the CDF Collaboration [2, 3] included the results from the 1992-1993 run at the Fermilab Tevatron Collider, firmly established the existence of the top quark (Fig. 1a). Simultaneous reports from CDF [4] and DØ [5] Collaborations later provided sufficient statistical significance to definitively claim its observation from the 1994-1995 run. The discovery of the top quark in nuclear collisions had to further wait another 20 years for the 2016 proton-nucleus run at the CERN LHC, as illustrated in Fig. 1b and described in the following. Of the heavy particles expected in the standard model, only the  $\tau$  lepton and the Higgs boson have not yet been detected in nuclear interactions.

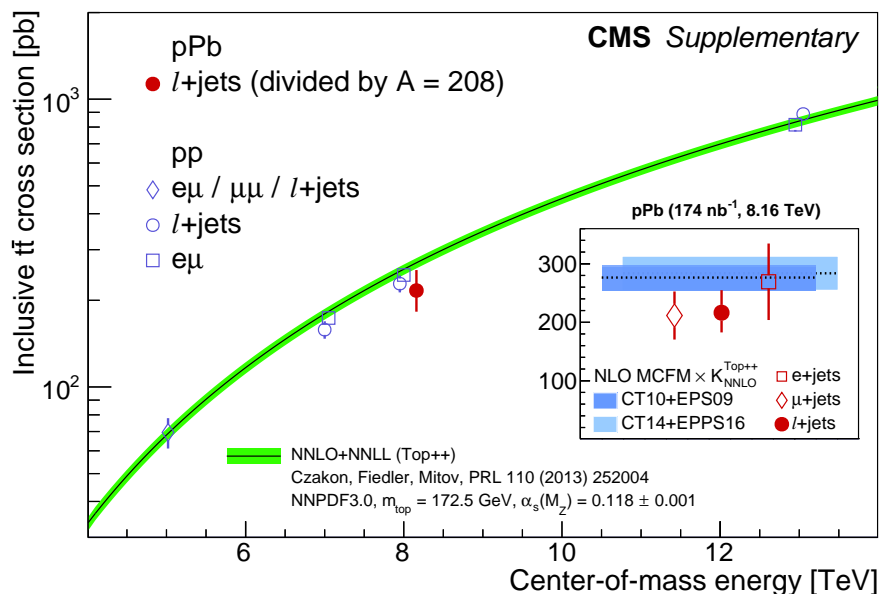


**Figure 1:** The top quark mass as first reconstructed in samples of (a) proton-antiproton [2, 3] and (b) proton-nucleus [TH1] collisions at 1.8 and 8.16 TeV, corresponding to integrated luminosities of  $19.3 \text{ pb}^{-1}$  and  $174 \text{ nb}^{-1}$  [TH2] (equivalent to  $36 \text{ pb}^{-1}$  of nucleon-nucleon collision data), respectively.

At hadron colliders the higher the center-of-mass energy the more top quarks are produced in pairs ( $t\bar{t}$ ). The large top quark mass uniquely provides a hard scale for the

associated cross section ( $\sigma_{t\bar{t}}$ ), a quantum chromodynamics (QCD) process determined with high accuracy based on perturbation expansions. Using the data sample of  $27.4 \pm 0.6 \text{ pb}^{-1}$  [TH3] collected by the CMS experiment during the proton-proton run at 5.02 TeV in 2015, the first measurement of the inclusive  $t\bar{t}$  cross section is presented for events with one or two high- $p_T$  leptons (electrons or muons), and at least two jets. The measurement is separately performed in four final states, i.e., using  $l$ +jets ( $l = e, \mu$ ) and dilepton ( $e^\pm\mu^\mp$  and  $\mu^\pm\mu^\mp$ ) events, and is then obtained from the combination of the individual measurements. The result is  $\sigma_{t\bar{t}} = 69.5 \pm 6.1 \text{ (stat)} \pm 5.6 \text{ (syst)} \pm 1.6 \text{ (lumi)} \text{ pb}$ , with a total relative uncertainty of 12% [TH4], which is consistent with the standard model prediction (Fig. 2).

Measurements of  $\sigma_{t\bar{t}}$  at various center-of-mass energies ( $\sqrt{s}$ ) probe different values of  $x$ , the fractional momentum of the proton carried by the partons, and thus can provide complementary information on the parton distribution functions (PDFs). The impact of the measured cross sections in the determination of the proton PDFs is studied in a QCD analysis at next-to-next-to-leading order. A moderate decrease of the uncertainty in the gluon distribution is observed in the less-explored kinematic range of  $x \gtrsim 0.1$ , consistent with the expectation of the high- $x$  region being probed at  $\sqrt{s} = 5.02 \text{ TeV}$ . Future measurements of  $\sigma_{t\bar{t}}$  in nucleus-nucleus collisions at the same nucleon-nucleon center-of-mass energy ( $\sqrt{s_{NN}}$ ) would profit from the availability of such a reference measurement, without the need to extrapolate from measurements at different  $\sqrt{s_{NN}}$ .



**Figure 2:** Top quark pair production cross section in proton-nucleus [TH1] and proton-proton [TH4] collisions as function of the center-of-mass energy; the measurements from the CMS Collaboration in the dilepton and  $l$ +jets final states are compared to the NNLO+NNLL theory predictions employing state-of-the-art proton and nuclear PDFs. The total experimental error bars (theoretical error bands) include statistical and systematic (PDF and scale) uncertainties added in quadrature.

The feasibility of top quark measurements in nuclear collisions is demonstrated with the first observation of the  $t\bar{t}$  process [TH1], using  $174 \pm 6 \text{ nb}^{-1}$  [TH2] of proton-nucleus

collisions at the higher  $\sqrt{s_{\text{NN}}} = 8.16 \text{ TeV}$ . The measurement is performed by analyzing events with exactly one isolated electron or muon and at least four jets. The inclusive cross section that is simultaneously measured in the two final states is  $\sigma_{t\bar{t}} = 45 \pm 8 \text{ (total) nb}$ , consistent with perturbative QCD calculations—employing state-of-the-art nuclear PDFs—as well as the expectations from scaled proton-proton data (Fig. 2). The statistical significance of the  $t\bar{t}$  signal against the background-only hypothesis is above five standard deviations. This first measurement paves the way for further detailed investigations of top quark production in nuclear interactions, providing, in particular, a new tool for studies of the strongly interacting matter created in nucleus-nucleus collisions.

With a total of only about eight weeks of operational experience, the luminosity with asymmetric proton-nucleus collisions surpassed the design value, and the threshold could have been pushed even further if the luminosity were not restrained for machine protection reasons. The long-term integrated luminosity goal of  $100 \text{ nb}$  was clearly surpassed in some experiments, albeit it represented the last proton-nucleus run for several years.

When proton and lead collide with each other again, one can reasonably expect to manage the luminosity debris better, to further increase the proton bunch intensity, and to deliver a few times more integrated luminosity than in 2016 within a similar time frame. At the long interval before the next run, the number and intensity of lead bunches meant for nucleus-nucleus collisions should increase substantially. Meanwhile, the excellent performance achieved in the 2018 PbPb run brought LHC one step closer to its high-luminosity era [6]. A series of improvements both in LHC and its injector chain, including an increase in the average colliding bunch intensity and a decrease in the nominal bunch spacing, resulted in reaching about six times higher the instantaneous luminosity than the design value of  $1 \times 10^{27} \text{ cm}^{-2}\text{s}^{-1}$ .



*The problem is that what we think is not interesting today might be fascinating tomorrow  
Or Hen, MIT, CLAS Collaboration*



## Acknowledgements

This work is a result of fruitful collaboration as well as guidance that I received during the past four years. In the following, I would like to express my sincere gratitude to people who made a substantial (direct or indirect) contribution towards the completion of this research. However, it is not possible to name all people involved because of either space restriction or an unfortunate oversight. I must apologize for failing to acknowledge anyone whose contribution has been crucial.

I am deeply grateful to my doctoral advisor, A. Giammanco, for his scientific guidance and ceaseless availability to help. This search has profited immensely from his constructive suggestions and broad expertise. His friendly and open-minded attitude formed the solid basis for our close collaboration. In particular, there was no moment that I felt my academic freedom to be retained. On the contrary, there was a constant motivation to pursue my interests and ideas about top quark physics and calibration of the physics objects. Thank you very much also for encouraging me to participate actively in international conferences and workshops, ending up to a total of about twice the credits required for graduation.

I thank for accepting to serve as additional members of the “Comité d’accompagnement” and dissertation committee, for their input and suggestions that have contributed to shaping the final version of the thesis, C. Delaere, D. d’Enterria, V. Lemaître, and G. Pásztor.

Moreover, I would like to thank my colleagues H. Bakhshiansohi, P. David (mon voisin de bureau), A. Jafari, M. Komm, and A. Popov (mon ex-voisin de bureau), i.e., the “The Offbeat Physicists @ CP3” team—a term coined by Andrea and Matthias—with whom I had much fun discussing about physics during the past four years.

While conducting the doctoral studies I collaborated with many people, but I would like to give special thanks to É. Chapon, K. Lipka, P. Silva, and M. Verweij. Apart from his support throughout the last four years, Pedro’s fascinating way for attacking and treating physics problems has been a unique lesson to me. I owe the personal inclination towards parton distribution functions to Katerina. Émilien and Marta make an admiring endeavor to lead the heavy ion (HIN) community within the CMS experiment and I had the distinct honor and privilege of participating in the HIN activities. As Marta remarked in a recent article in the *symmetry* magazine: “We have four weeks to collect all the data we will use for the next couple of years.”

The paper on the top quark observation in nuclear collisions would not be possible without the direct contribution from É. Chapon, D. d’Enterria, A. Giammanco, M. Mulders, P. Silva, and M. Verweij, and the hard work of the HIN calibration groups, including K.E. Jung, O. Kukral, C. McGinn, C. Mironov, M. Nguyen, J. Park, A. Ståhl *et al.*. Producing results that the world has not seen before entails an exhaustive journey from early analysis to publication. I would like therefore to thank i) the Analysis Review Committee (ARC) consisted of M. d’Alfonso, O. Kodolova, A.B. Meyer, and D. Noonan; ii) signatories to the paper for commenting on text and the physics described in the document during the collaboration-wide review (CWR); iii) the members of the Publication Committee (Pub-Comm) W. Busza, O. Evdokimov, V. Greene, D. Krofcheck, G.M. Roland, S.S. Padula, S.J.

Sanders, G.S.F. Stephans, and J. Velkovska, who are responsible for the final stage of the paper before it leaves the Collaboration. I would also like to thank G. Alverson and C. Lourenco for preparing the paper format for submission and comments on the draft prepared for the CERN Courier Journal, respectively.

For the successful collaboration on the top quark measurement at  $\sqrt{s} = 5.02 \text{ TeV}$  it was a pleasure to work with E. Eren, A. Giammanco, J.R. González, K. Lipka, E. Palencia, P. Silva, and M. Verweij. Thanks to the corresponding Physics Objects Groups for their support and comments. Special thanks to the ARC (J.R. Komaragiri, D. Horvath, E. Robutti, and H. Stadie) and PubComm members (T. Ferguson, F. Matorras, S. Qian, and K. Stenson), and everyone vividly contributed during the CWR process.

The topic of the very first research within CMS, i.e., the measurement of  $t$ -channel single top quark production at  $\sqrt{s} = 13 \text{ TeV}$  was initially proposed to me by Andrea. I feel honor bound to acknowledge the work of T. Chwalek, N. Faltermann, O. Iorio, S. Mitra, T. Müller, S. Paktinat, M. Zeinali, and “The Offbeat Physicists @ CP3” team; it was an invaluable experience of sharing my enthusiasm for the very first raw data from Run 2 with you.

I want to give my thanks for the great collaboration and fun we had while developing further the tools for the luminosity measurement at CMS to: A. Babaev, J. Benitez, F. Brivio, M. Casarsa, A. Dabrowski, S. Fiorendi, M. Guthoff, S.L. Higginbotham, O. Karacheban, J. Knolle, J.L. Leonard, P. Lujan, R. Manzoni, D. Marlow, A.B. Meyer, J. Salfeld-Nebgen, C. Palmer, G. Pásztor, R. Sosa, D. Stickland, D. Takaki, and O. Turkot.

I would also like to express my gratitude to the conveners and sub conveners of the CMS “TOP” and “HIN” Physics Analysis Groups M. Aldaya, J. Andrea, A. Giammanco, O. Iorio, A. Jafari, J.M. Keaveney, A.B. Meyer, P. Silva, L. Skinnari, and R. Suarez, and É. Chapon, W. Li, C. Mironov, and M. Verweij. It was a pleasure to contribute to the top quark and heavy ion efforts in the friendly and welcoming atmosphere you all created. I would like to thank the CMS Management Board, and especially the involved Physics Coordination (J. Alcaraz, T. Bose, J. Olsen, S. Rahatlou) for keeping track of the tremendous physics results.

Even without being directly involved in my projects, some people ended up being closer collaborators to me: T. Arndt, T. Cheng, P. Gunnellini, S. Harper, J. Kieseler, Q. Li, T. McCauley, K. Mondal, L. Pernie, L. Perozzi, T. Sakuma, S. Sekmen, K. Skovpen, L. Vanelderen, X. Wang, E. Yazgan, S. Zenz *et al.*

M. Drewes, A. Giammanco, J. Hajer, M. Lucente, and O. Mattelaer pursued “A Heavy Metal Path to New Physics”; their passion amazed me, while their effort provided me with mental stimulus for rethinking the future of heavy ion program at LHC and post-LHC. Thanks for organizing the “Heavy Ions and Hidden Sectors” workshop, a unique chance to bring together members of the involved communities.

I would also like to sincerely thank C. Mertens and G. Tabordon for their prompt assistance in administrative procedures, and I appreciate the help of P. Demin, J. de Favereau, O. Mattelaer, and A. Tanasijczuk in computing infrastructure.

Last but not least, I would like to acknowledge the work of a few hundreds of engineers, accelerator physicists, and experimentalists who designed and constructed LHC and the

CMS detector. Likewise, the excellent physics performance of the CMS experiment has been only achieved thanks to the hard work of people developing and maintaining ever-better reconstruction algorithms. The constant effort put into monitoring of detector and accelerator operation, both regarding crew control rooms and remotely, should also be emphasized.

I am deeply grateful to Université catholique de Louvain (UCLouvain), Institut de Recherche en Mathématique et Physique (IRMP), and the Centre for Cosmology, Particle Physics and Phenomenology (CP3) in Belgium for unceasingly supporting my work.

Louvain-la-Neuve, January 2019

*Georgios Konstantinos Krintiras*



<b>Preface</b>	<b>1</b>
<b>Acknowledgements</b>	<b>7</b>
<b>1 Introduction</b>	<b>15</b>
1.1 Setting the scene . . . . .	16
1.1.1 A special role in the EWSB mechanism? . . . . .	16
1.1.2 Theory predates experiment; experiment keeps a tight rein on theory	18
1.1.3 That could well be a top quark! . . . . .	21
1.2 The synchrotron legacy: a luminosity story . . . . .	23
1.2.1 “A great accomplishment is not the end of the road, just the starting point for the next leap forward” . . . . .	26
<b>I Part A</b>	<b>31</b>
<b>2 Accelerator facilities and experimental apparatus</b>	<b>33</b>
2.1 The Large Hadron Collider complex . . . . .	34
2.1.1 Performance and achievements of synchrotron accelerators . . . . .	34
2.1.2 Delivering collisions to the seven LHC experiments . . . . .	35
2.1.3 Fundamental principles of synchrotron accelerators . . . . .	38
2.1.4 LHC layout and global structure . . . . .	43
2.1.5 Proton-nucleus collisions at LHC . . . . .	47
2.2 The Compact Muon Solenoid detector . . . . .	51
2.2.1 The superconducting solenoid magnet . . . . .	54
2.2.2 The inner silicon pixel and the larger silicon strip tracker . . . . .	54
2.2.3 The electromagnetic calorimeter . . . . .	56
2.2.4 The hadron calorimeter . . . . .	58
2.2.5 The muon detectors . . . . .	59
2.2.6 Luminosity detectors at CMS . . . . .	61
<b>3 Global event and physics object reconstruction</b>	<b>65</b>
3.1 The online event filter . . . . .	67
3.1.1 L1 trigger overview . . . . .	67
3.1.2 HLT overview . . . . .	69
3.1.3 Electron and photon triggers . . . . .	69
3.1.4 Muon triggers . . . . .	71
3.1.5 Beam position timing trigger . . . . .	72
3.2 Hit, track, and primary vertex reconstruction . . . . .	73
3.2.1 Hit reconstruction performance . . . . .	74
3.2.2 Track reconstruction performance . . . . .	75
3.2.3 Primary vertex reconstruction performance . . . . .	77

3.2.4	Reconstruction of the LHC luminous region . . . . .	79
3.3	Physics object reconstruction . . . . .	81
3.3.1	Pileup mitigation and treatment of the underlying event . . . . .	82
3.3.2	Muon reconstruction and identification . . . . .	84
3.3.3	Electron reconstruction and identification . . . . .	87
3.3.4	Jet reconstruction and noise-jet rejection . . . . .	93
3.3.5	Identification of heavy-flavor jets . . . . .	99
3.3.6	Reconstruction and calibration of $\vec{p}_T^{\text{miss}}$ . . . . .	101
<b>II</b>	<b>Part B</b>	<b>107</b>
<b>4</b>	<b>Absolute luminosity calibration in the CMS experiment</b>	<b>109</b>
4.1	The concept of luminosity . . . . .	111
4.1.1	Interaction rate determination . . . . .	111
4.1.2	Formalism of absolute luminosity from separating the beams . . . . .	113
4.2	Offline luminosity measurement in CMS . . . . .	116
4.2.1	Experimental setup . . . . .	116
4.2.2	The vdM scan protocols . . . . .	117
4.2.3	The pixel-cluster counting method . . . . .	120
4.2.4	Extraction of the visible cross section . . . . .	121
4.3	Calibration corrections . . . . .	122
4.3.1	Concept and formalism of constant separation LSC scan . . . . .	122
4.3.2	Orbit drift . . . . .	127
4.3.3	Impact of nonfactorizable beam shape . . . . .	128
4.3.4	Beam-beam effects . . . . .	130
4.3.5	Bunch current corrections and normalization . . . . .	131
4.4	Cross-detector stability . . . . .	133
4.5	Conclusions . . . . .	135
<b>III</b>	<b>Part C</b>	<b>139</b>
<b>5</b>	<b>Measurement of the <math>t\bar{t}</math> cross section in LHC pp collisions at 5.02 TeV</b>	<b>141</b>
5.1	Filling the gap between Tevatron and LHC energies . . . . .	142
5.2	Data, simulated samples and theoretical cross section . . . . .	143
5.3	Event selection . . . . .	144
5.4	Background estimation . . . . .	147
5.4.1	The $\ell$ +jets final state . . . . .	147
5.4.2	The dilepton final state . . . . .	153
5.5	Systematic uncertainty . . . . .	157
5.5.1	Experimental uncertainty . . . . .	157
5.5.2	Theory uncertainty . . . . .	158
5.6	Measurement of the $t\bar{t}$ cross section . . . . .	161
5.6.1	The $\ell$ +jets final state . . . . .	161



5.6.2	The dilepton final state . . . . .	165
5.6.3	Combination of the $\ell$ +jets and dilepton final states . . . . .	168
5.7	QCD analysis . . . . .	171
<b>6</b>	<b>Observation of top quark production in proton-nucleus collisions</b>	<b>175</b>
6.1	Breaking the barrier of low $\sqrt{s_{\text{NN}}}$ and event count . . . . .	176
6.2	Event sample description . . . . .	176
6.3	Event sample selection and physics objects reconstruction . . . . .	177
6.4	Signal and background parameterization . . . . .	186
6.4.1	Signal parameterization . . . . .	187
6.4.2	Background parameterization . . . . .	188
6.5	Systematic uncertainty . . . . .	193
6.5.1	Experimental uncertainty . . . . .	194
6.5.2	Theory uncertainty . . . . .	195
6.6	Measurement of the inclusive $t\bar{t}$ cross section . . . . .	196
6.6.1	Fit model . . . . .	196
6.6.2	Extraction of the inclusive $t\bar{t}$ cross section . . . . .	196
6.6.3	Impact of the statistical and systematic uncertainties . . . . .	198
6.6.4	Statistical significance of the signal . . . . .	200
<b>IV</b>	<b>Part D</b>	<b>207</b>
<b>7</b>	<b>Projected improvements in the precision of the inclusive <math>\sigma_{t\bar{t}}</math> measurement</b>	<b>209</b>
7.1	The largest reference proton-proton data set at 5.02 TeV . . . . .	210
7.2	Incremental improvements in luminosity determination . . . . .	210
7.2.1	Refined impact of nonfactorizable beam shape . . . . .	210
7.2.2	Concept and formalism of variable separation scan . . . . .	212
7.3	Cross section measurements . . . . .	214
7.3.1	Statistical and systematic uncertainty treatment . . . . .	214
7.3.2	PDF constraints including the inclusive $\sigma_{t\bar{t}}$ . . . . .	214
	<b>Summary and outlook</b>	<b>217</b>
	<b>Appendices</b>	<b>219</b>
<b>A</b>	<b>Summary of the impacts and pulls of the most significant nuisance parameters in the <math>\ell</math>+jets final state at 5.02 TeV</b>	<b>221</b>
<b>B</b>	<b>Measurement and optimization of the lepton efficiency</b>	<b>223</b>
B.0.1	The tag-and-probe technique . . . . .	223
B.0.2	Electron efficiency . . . . .	225
B.0.3	Muon efficiency . . . . .	231

<b>C QCD multijet model from data and isolation of charged leptons</b>	<b>237</b>
C.0.1 Pileup mitigation techniques for the isolation of charged leptons . . .	238
<b>D Synopsis: Top quark in nuclear collisions</b>	<b>245</b>
<b>Scientific output</b>	<b>247</b>
<b>Internal notes</b>	<b>249</b>
<b>Bibliography</b>	<b>251</b>

## Contents

---

<b>1.1 Setting the scene</b> . . . . .	<b>16</b>
1.1.1 A special role in the EWSB mechanism? . . . . .	16
1.1.2 Theory predates experiment; experiment keeps a tight rein on theory . . . . .	18
1.1.3 That could well be a top quark! . . . . .	21
<b>1.2 The synchrotron legacy: a luminosity story</b> . . . . .	<b>23</b>
1.2.1 “A great accomplishment is not the end of the road, just the starting point for the next leap forward” . . . . .	26

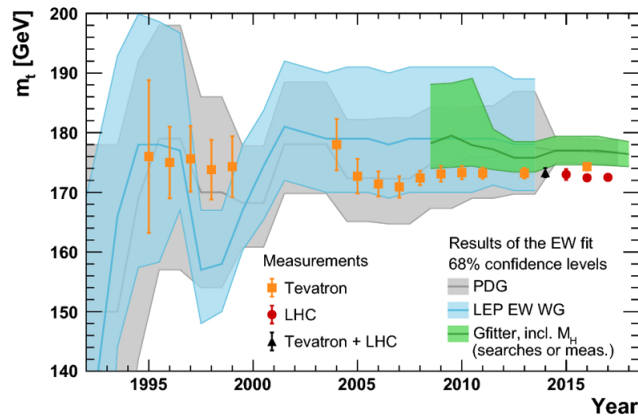
---

After the discovery of the top quark more than 20 years ago, top quark production cross sections, increasingly dominated by gluon fusion processes with higher center-of-mass energies, have been meticulously studied using hadron-hadron collisions. The rich variety of results from the LHC experiments are complemented with increasingly accurate theoretical predictions of heavy quark production and decay. Measurements of the top quark production provide a benchmark test of the standard model at energy scales much larger than those typically involved for the other quarks. The large top quark mass ( $m_{\text{top}}$ ), which lies remarkably close to the vacuum expectation value of the Higgs field, sets a scale for higher order perturbative calculations in quantum chromodynamics (QCD).

The proton structure can be extensively probed combining measurements of top quark production at various center-of-mass energies since they exhibit sensitivity on different values of Bjorken  $x \gtrsim 2m_{\text{top}}/\sqrt{s}$ , and hence can provide complementary information on the gluon distribution, as well as light quarks. The energies and large data samples available at LHC have already allowed to measure various large-mass elementary particles, for the first time, in nuclear collisions. Compared to hadron-hadron collisions, and based on parametrizations of nuclear modifications to the free nucleon densities, the inclusive top pair production is expected to be mildly enhanced because of an overall net gluon antishadowing, with different regions of their differential distributions being depleted as a result of shadowing or EMC-effect corrections. Novel physics opportunities are, in turn, opened up such as constraints of nuclear parton densities in less explored kinematic ranges, and studies of the dynamics of heavy-quark energy loss in the quark-gluon plasma, a deconfined QCD state of matter.

## 1.1 Setting the scene

With a mass close to the one of a gold atom, yet a size less than  $10^{-3}$  fm, the top quark is the heaviest elementary particle known. When the top quark was independently discovered by the CDF [4] and DØ [5] Collaborations in 1995 at the Tevatron proton-antiproton ( $p\bar{p}$ ) collider, a decades-long search for one of the last missing pieces of the standard model (SM) had come to an end. Despite an early blunder, that had resulted in claiming a “clear signal” compatible with a W boson decaying into a 40 GeV top quark in 1984 [7], indirect hints of a large top quark mass ( $\gg 50$  GeV) later came from significant flavor oscillation frequencies in  $B_d^0$  meson mass eigenstates and electroweak (EW) precision data from LEP Collaborations and SLD experiment [8]. In Fig. 1.1, the limits on the top quark mass,  $m_{\text{top}}$ , as a function of time are compared to direct measurements at the Tevatron, the CERN LHC, and their combination. The indirect determination within the framework of the SM adheres to a remarkably good agreement with the direct measurements. It is thus interesting to note that the top quark influences predictions regarding the stability of the Higgs field and its effects on the evolution of the Universe as well as the sensitivity of SM consistency checks.



**Figure 1.1:** Prediction of the top quark mass as a function of time as obtained by various analysis groups using electroweak precision data. The bands indicate the 68% confidence level. The direct measurements after the top quark discovery are displayed by the data points [9]. The good agreement between the top quark mass measured directly and the predicted mass determined indirectly from measurements at the Z pole is a convincing illustration of the validity of radiative corrections.

### 1.1.1 A special role in the EWSB mechanism?

The SM describes the fundamental elementary fields, that propagate in flat spacetime, and their interactions. The internal symmetry of the SM is

$$SU(3)_C \times SU(2)_L \times U(1)_Y \xrightarrow{\text{EWSB}} SU(3)_C \times U(1)_Q,$$

where  $U(1)_Y$  is a  $U(1)$  symmetry related to the hypercharge,  $Y$ , the  $SU(2)_L$  symmetry acts only on doublets of left handed Dirac spinors (fermions), and  $SU(3)_C$  is a non-Abelian local

symmetry with additional self-interacting (“coloured”) terms from the gluon field strength. The  $SU(2)_L \times U(1)_Y$  Lagrangian also includes a complex scalar doublet field, the Higgs doublet, whose symmetry is broken when the latter acquires a vacuum expectation value (VEV). The VEV should be electrically neutral, and perturbations around the ground state,  $v$ , can be described by the unitary gauge that gives a convenient physical interpretation: the Higgs field is the real scalar field [10–15]. Interaction terms between the Higgs field and the gauge bosons break the symmetry via their mass terms ( $W^\pm, Z$ ) that are proportional to  $v^2$ , while the new unbroken  $U(1)_Q$  symmetry describes the massless gauge boson ( $\gamma$ ). The SM “lives” at the energy scale of  $v = 246 \text{ GeV}$  [16] that triggers the EW symmetry breaking (EWSB) of the Higgs doublet. On the one hand, the discovery of a Higgs boson at the CERN LHC [17–19] and measurements of its properties, e.g., Ref. [20], consolidate the SM. On the other hand, producing particle collisions with energies above this scale is bound to probe the mechanism of EWSB, i.e., whether it is given by an elementary Higgs boson or by some alternative dynamics.

The corresponding SM Lagrangian,  $\mathcal{L}_{\text{SM}}$ , is invariant under the Poincaré group, and can be divided into four parts:

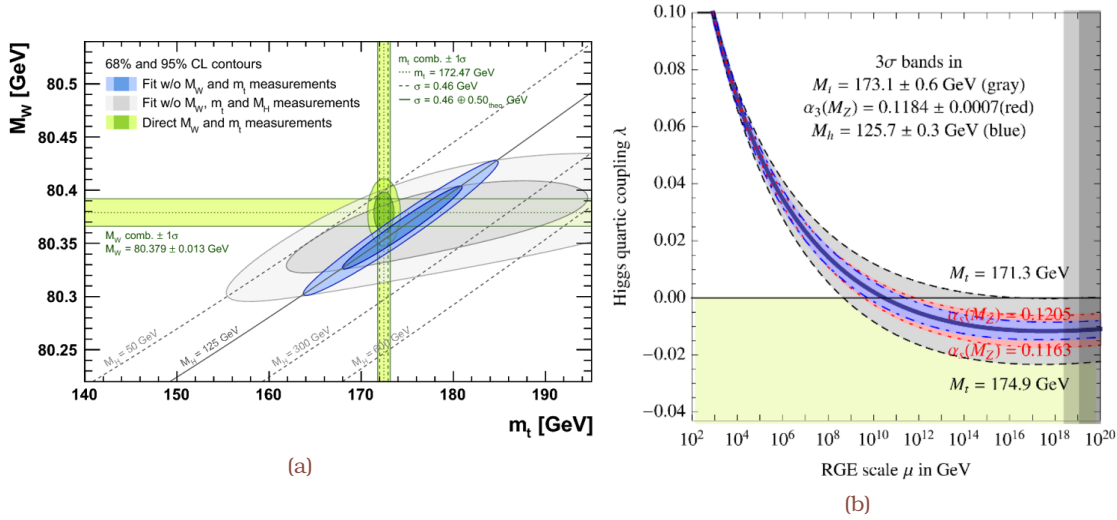
$$\mathcal{L}_{\text{SM}} = \mathcal{L}_{\text{Gauge}} + \mathcal{L}_{\text{Matter}} + \mathcal{L}_{\text{Higgs}} + \mathcal{L}_{\text{Yukawa}}.$$

The  $\mathcal{L}_{\text{Gauge}}$  term refers to the kinetic terms and interactions among the gauge bosons. The  $\mathcal{L}_{\text{Matter}}$  term includes the interactions of the fermionic particles with the gauge bosons through covariant derivatives. The  $\mathcal{L}_{\text{Higgs}}$  term includes the Higgs self-interaction terms, the interactions between the Higgs and the gauge bosons and the EW gauge boson mass terms. Finally, the  $\mathcal{L}_{\text{Yukawa}}$  term describes the interactions between the Higgs field and the fermions as well as the fermion mass terms. The Lagrangian  $\mathcal{L}_{\text{SM}}$  contains 19 independent parameters in total: nine fermion (excluding the neutrino) masses, the mass of the Higgs and the VEV, three gauge couplings, three mixing angles and one CP-violating phase in the Cabibbo–Kobayashi–Maskawa (CKM) matrix, the latter accounting for the fact that the flavor and mass eigenstates are different for quarks. It is thus evident that the Yukawa sector accounts for nearly two-third of all free SM parameters.

In the SM, the quark and lepton masses have to be thus “inserted” through adjustable parameters that describe how feebly or strongly each type interacts with the Higgs boson. For a top quark the Yukawa coupling,  $y = \sqrt{2} \frac{m}{v}$ , is almost exactly unity, unlike for any other fermion. This relatively strong coupling with the Higgs boson suggests that the top quark may play a special role in the mechanism of EWSB. Quadratically divergent radiative corrections appear in the computation of the Higgs boson mass and are dominated by top quark loop contributions. Repeating the exercise of indirect determination of  $m_{\text{top}}$  from global SM fits to EW precision data is thus essential for testing the overall SM consistency. A scan [9] of the confidence level profile of W boson mass versus  $m_{\text{top}}$  is shown in Fig. 1.2a for the scenarios where the direct Higgs boson measurements [21, 22] are included (blue) or not (grey) in the fit. Both contours agree with the direct measurements (green bands and ellipse).

In addition, owing to its large value, the top quark mass has a direct impact on

extrapolations of the EWSB- to high-energy scales. Radiative corrections can drive the Higgs boson self-coupling ( $\hat{\lambda}$ ) towards zero or even negative values, potentially leading to an unstable vacuum. The determination of the energy scale ( $\mu$ ) where the vacuum becomes unstable, either requiring or not beyond the SM (BSM) physics at lower energies like the EWSB scale, is strongly influenced by the precision of the measurements in the top quark sector and their unambiguous interpretation in a clear theoretical framework: In Fig. 1.2b  $\mu$  varies by several orders of magnitude under a seemingly insignificant variation of  $\pm 2$  GeV in  $m_{\text{top}}$ . Last but not least, the indication of a flat Higgs potential ( $\hat{\lambda} = 0$ ) at very high  $\mu$  values could imply the possibility that the Higgs boson plays the role of the scalar field in the early universe, an appealing phenomenological consequence, e.g., Ref. [23].

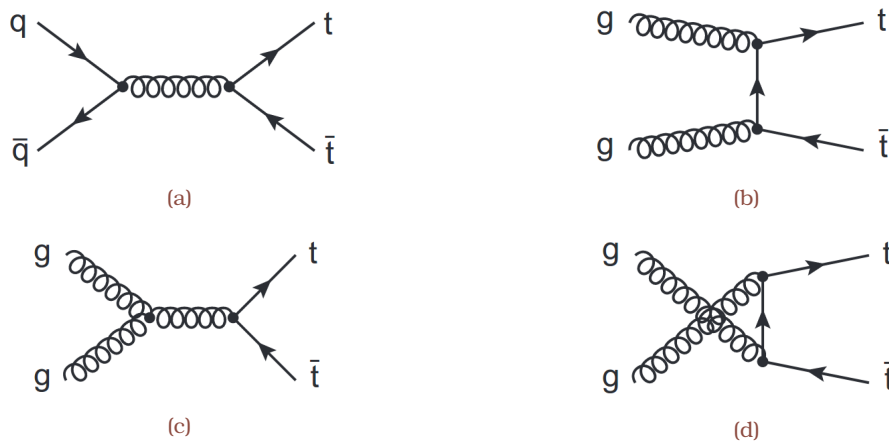


**Figure 1.2:** (a) Contours at 68% and 95% confidence level obtained from scans of fits [9] with the W boson and top quark mass fixed. The corresponding direct measurements are excluded from the fit. The narrower blue and larger grey allowed regions are the results of the fit including and excluding the Higgs boson mass measurement [24], respectively. The horizontal bands indicate the  $\pm 1\sigma$  regions of the world averages used for the W boson [16] and top quark mass [25]. A theoretical uncertainty of 0.5 GeV is added to the direct top mass measurement to take into account nonperturbative effects that are deemed to be of  $\mathcal{O}(\Lambda_{\text{QCD}})$ . (b) Renormalization-group evolution of the Higgs boson self-coupling varying the top mass, strong  $\text{SU}(3)_C$  coupling, and Higgs boson mass within a range of  $\pm 3\sigma$  [26]. Residual theoretical uncertainty [27] is safely smaller than the experimental uncertainty, dominated by the uncertainty in the determination of the top quark mass.

### 1.1.2 Theory predates experiment; experiment keeps a tight rein on theory

The basic tool to study top quark is a high energy particle collider. For instance, the dominant production of top quark pairs ( $t\bar{t}$ ), which is a quantum chromodynamics (QCD) process, is accessible both at lepton and hadron (Fig 1.3) colliders, provided the  $\sqrt{s}$  value of the collisions is above the production threshold of twice  $m_{\text{top}}$ . The SM predicts the EW production of single top quarks in addition to the dominant  $t\bar{t}$  process [TH5], while the associated production of top quarks with the EW gauge bosons or the Higgs boson is also

possible [TH6]. A natural question can be then addressed: how theory and experiment have coped with the broad range of studies in top quark physics, from inclusive production cross sections and precise measurements of the top quark mass and of its couplings to a variety of searches for BSM physics with top quarks [TH7].



**Figure 1.3:** The Feynman diagram approach is based on perturbation theory assuming that the coupling constants that appear in the vertices are parameters smaller than one. In this case, representative diagrams [28] are given for  $t\bar{t}$  production at the lowest (“leading”) order in the strong coupling constant:  $q\bar{q}$  annihilation (a) and  $gg$  fusion in the  $t$ - (b),  $s$ - (c), and  $u$ -channel (d) configuration. The inclusion of diagrams with real and virtual corrections defines the order of a calculable observable in  $a_s$ .

The probability (cross section),  $\sigma$ , for any process can be expressed as a weighted product of a high-energy (“hard”) parton-parton scattering multiplied with parton distribution functions (PDFs), integrated over all parton momenta and summed over all parton types (“phase space”):

$$\sigma = \sum_{j,k}^{\text{partons}} \int_0^1 dx_j dx_k f_k(x_k, \mu_F^2) \hat{\sigma}(sx_j x_k, \mu_F, a_s) .$$

While the hard scattering cross section,  $\hat{\sigma}$ , is process specific and can be computed in perturbative QCD at different levels of accuracy, PDFs,  $f_i(x_i, \mu_F^2)$ , are deemed universal functions that describe the probability to find a parton  $i$  with a given longitudinal momentum fraction  $x_i$  (Fig. 1.4), when the hadron is probed at a momentum transfer scale of  $Q^2 = \mu_F$ . The PDF parameterization absorbs all long-distance effects in the initial state, and ultraviolet divergences in  $\hat{\sigma}$  are treated with a renormalization procedure of the strong coupling constant,  $a_s$ , introducing an additional energy scale ( $\mu_R$ ). The “default” (but not unique) choice of energy scales are process specific, e.g., for  $t\bar{t}$  production a typical choice is  $m_{\text{top}}$ .

Inclusive cross sections and distributions are already known to the impressive level of a few percent in uncertainty. Top pair and single top production can be computed fully-differential at next-to-next-to leading order (NNLO) in perturbative QCD for stable tops [29] (and References therein), and recently NNLO corrections can also describe the decay

products differentially in both production and decay subprocesses [30]. The fixed-order predictions are often supplemented with various resummations—typically at next-to-next-to-leading logarithm (NNLL) accuracy—which stabilize the predictions against the effects of large logarithms. Since the expected hierarchy between the strong and electroweak interactions may not be respected at such large scales, electroweak corrections are thus also studied at this level of accuracy, including even photon-induced contributions from the hadron content [31].

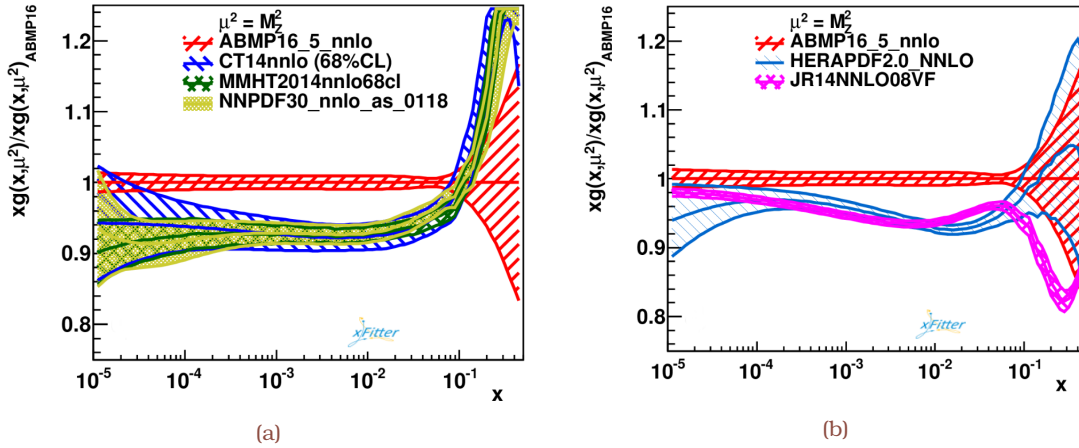


Figure 1.4: The proton gluon PDF as a function of the longitudinal momentum  $x$  as determined from the (a) ABMP16 [32], CT14 [33], MMHT14 [34], NNPDF3.0 [35], (b) HERAPDF2.0 [36], and JR14 [37] sets with their  $\pm 1\sigma$  uncertainties at a factorization scale equal to the  $Z$  boson mass. The distributions are normalized to the central values from ABMP16 for comparison [32].

QCD color confinement requires the final state particles to be color-neutral, but the process of turning partons into hadrons cannot be treated perturbatively, and it thus relies on phenomenological models. To compare calculations to data software tools are used based on the Monte Carlo (MC) method that simulates hadron-hadron collisions. “General-purpose” MC event generators, e.g., as described in Refs. [38–41], can automatically compute the full set of contributions to the hard process at NLO given the Feynman rules of the underlying theory (both SM and BSM). For most of the processes, the normalization of the MC event generation is corrected to match the most precise calculation, ignoring the effect that higher-order corrections may cause on the shape of kinematic observables.

The final process of turning partons into hadrons is separated into two steps; a probabilistic method to model the fragmentation of partons that effectively resums soft and collinear radiation (parton shower), and hadron formation (hadronization). To avoid double-counting parton emissions due to higher order processes, specialized matching and merging techniques (e.g., Refs. [42–44]) are available to consistently interface NLO MC event generators to the parton shower, typically to leading-logarithmic (LL) precision. Frameworks have been recently developed to consistently match additional sets of NLO calculations to parton showers, including the treatment of the top quark propagator in the narrow-width approximation as well as keeping the top quarks off their mass shell



(e.g., Ref. [45]). Hadronization is described with phenomenological approaches that are subjected to scrutiny, e.g., Refs. [46–48], and with the most popular being the Lund string [49–51] and the cluster [52] models, albeit their unique implementation has been lately questioned in view of multiparticle production effects, with two typical examples being the enhanced strangeness production [53] and the formation of a “ridge.” [54–56]

Throughout this thesis, data are compared to the predictions of different generator settings for the  $t\bar{t}$  (“signal”) and other SM (“background”) processes. Table 1.1 summarizes the main characteristics of the setups and abbreviations used.

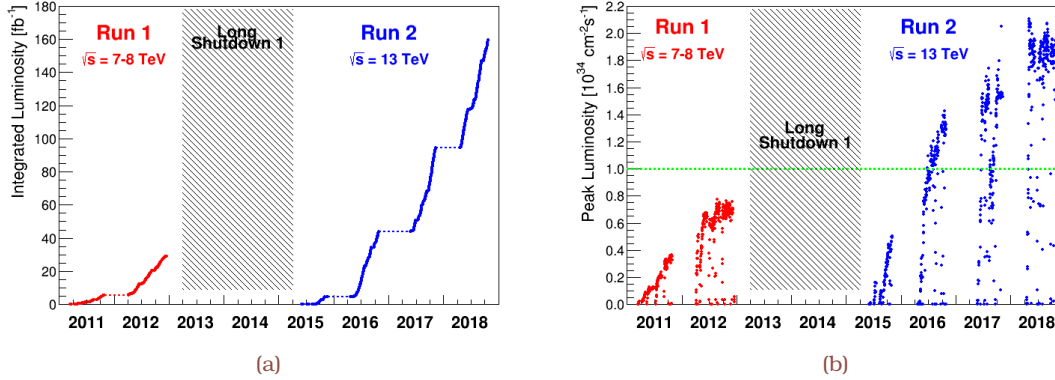
**Table 1.1:** MC simulation settings used for the comparisons with the inclusive and differential  $\sigma_{t\bar{t}}$  measurements. The table lists the main characteristics and values used for the most relevant parameters of the generators.

Event generator	POWHEG (v2) [39, 40]	MADGRAPH5_aMC@NLO (v2.2.2) [38]
<i>Matrix element (ME) characteristics</i>		
Mode	hVq [57]	FxFx merging [43]
Scales ( $\mu_R, \mu_F$ )	$\sqrt{m_{\text{top}}^2 + p_{T,\text{top}}^2}$	$\sum_{t,\bar{t}} \sqrt{m_{\text{top}}^2 + p_{T,\text{top}}^2} / 2$
$\alpha_s(M_Z)$	0.118 [16]	0.118
PDF	NNPDF3.0 NLO [35]	NNPDF3.0 NLO
Accuracy	$t\bar{t}$ (NLO)	$t\bar{t} + 0, 1, 2$ jets (NLO)
<i>Parton shower</i>		
Setup designation		
Parton shower		PYTHIA (v8.205) [49, 51]
Tune		CUETP8M1 [58, 59]
PDF		NNPDF2.3 LO [60, 61]
$(\alpha_s^{\text{ISR}}(M_Z), \alpha_s^{\text{FSR}}(M_Z))$		(0.1108, 0.1365)
ME corrections		on
<i>Parton shower alternative</i>		
Setup designation		
PS	HERWIG++ (v2.7.1) [52]	<b>X</b>
Tune	EE5C [62]	
PDF	CTEQ6 (vL1) [63]	
$(\alpha_s^{\text{ISR}}(M_Z), \alpha_s^{\text{FSR}}(M_Z))$	(0.1262, 0.1262)	
ME corrections	off	

### 1.1.3 That could well be a top quark!

Until the end of its data taking in 2011, the Tevatron collider had delivered more than  $10 \text{ fb}^{-1}$  of integrated luminosity [64]. The successful attempt from LHC at  $\sqrt{s} = 7 \text{ TeV}$  in 2010 (“Run 1”) has marked the beginning of a new era in top quark physics. Enhanced cross sections for top quark production, on the one hand, and ever increasing data sets culminating at  $25 \text{ fb}^{-1}$  [65] by the end of 2012 at  $\sqrt{s} = 8 \text{ TeV}$  (Fig. 1.5a), on the other hand, rendered LHC the first “top quark factory.” After a two-year shutdown for maintenance of the LHC machine and experiments, the LHC has restarted in early 2015 (“Run 2”). The  $\sqrt{s}$  value has further increased to  $13 \text{ TeV}$ , boosting typical top quark cross sections by

about a factor of three compared to Run 1 (Fig. 1.5a). With an availability for luminosity production of about 50% [66] an integrated luminosity of almost 50 [TH8] and  $70 \text{ fb}^{-1}$  has been further delivered in 2017 and 2018, respectively, producing more than  $185 \text{ fb}^{-1}$  of proton-proton (pp) data for Run 2 and comfortably surpassing the target of  $150 \text{ fb}^{-1}$ .



**Figure 1.5:** (a) Cumulative (“integrated”) delivered and (b) instantaneous (Eq. (1.2)) luminosity versus time for pp collisions during LHC Run 1 and 2 periods at  $\sqrt{s} = 7$  or 8 and 13 TeV, respectively [67].

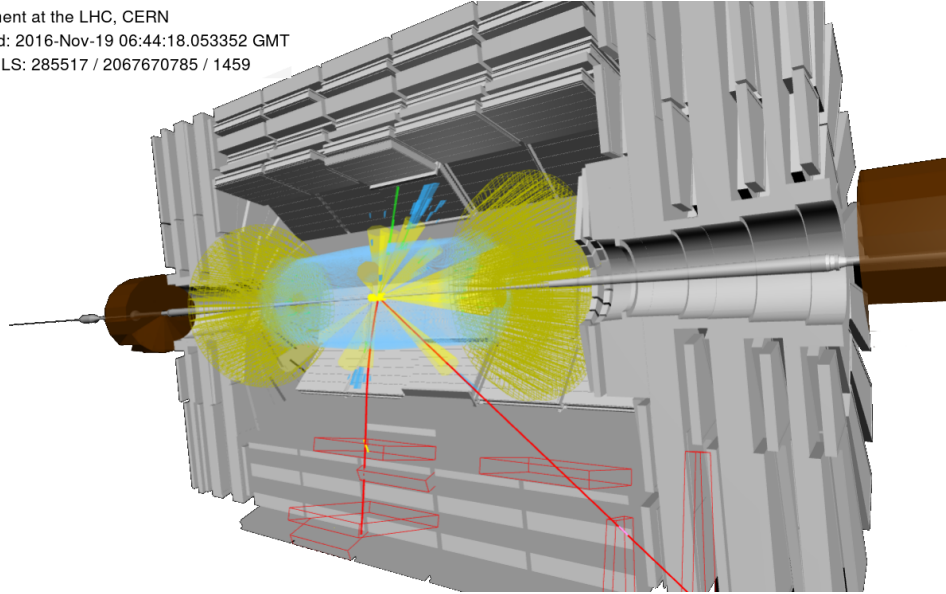
The top quark has been thus the subject of numerous detailed studies based on data samples with the integrated luminosity ranging from few  $\text{pb}^{-1}$  to several  $\text{fb}^{-1}$  both in  $p\bar{p}$  and pp collisions [TH9]. Until recently, top quark remained elusive in nuclear collisions because of the small amount of events recorded during the first heavy ion runs at LHC, and the low nucleon-nucleon center-of-mass energies available at RHIC. This situation changed [TH10, TH11] when the 2016 LHC proton-lead run at  $\sqrt{s_{\text{NN}}} = 8.16 \text{ TeV}$  produced a data set corresponding to an integrated luminosity of  $174 \text{ nb}^{-1}$  [TH2] (equivalent to  $36 \text{ pb}^{-1}$  of nucleon-nucleon collision data), releasing the very first  $t\bar{t}$ -like event signatures (Fig. 1.6).

Unlike the up and down quarks, which are stable, the top quark has a mean lifetime of only about  $10^{-24} \text{ s}$ . In the SM, the top quark decays almost instantly and nearly all the times into a W boson and a bottom (b) quark [16]. Neither the W boson nor the b quarks can be directly observed. When a quark emerges from a collision, it gets “dressed up” by a cloud of quarks and antiquarks. What is then experimentally observed is a jet (Fig. 1.7), a directed spray of particles that in the vacuum have roughly the same energy and direction as the original quark. The W boson decays with a probability of almost 70% [16] into a quark and an antiquark of the same generation (“hadronically”). In this case, the quark and antiquark show up as two jets correlated in phase space. But the W boson also decays, though with a decreased probability of around 10% [16], “leptonically,” i.e., into a charged and a neutral lepton of the same generation, such as a muon and a neutrino. The neutrino traverses a detector completely unobserved. Its presence can be indirectly deduced because of a significant amount of missing momentum, assumed to have been carried away by the neutrino.

Among the most striking features of a top quark signature are therefore the jets containing b quarks. General-purpose detectors, like CMS [72], put great emphasis on



CMS Experiment at the LHC, CERN  
 Data recorded: 2016-Nov-19 06:44:18.053352 GMT  
 Run / Event / LS: 285517 / 2067670785 / 1459



**Figure 1.6:** A pPb collision featuring a  $t\bar{t} \rightarrow e^\pm \mu^\mp + X$  candidate event in the  $\sqrt{s_{NN}} = 8.16$  TeV data sample recorded by CMS. Calorimetric energy deposits in ECAL (HCAL) are depicted as green (blue) bars with a height proportional to the magnitude of the deposit. Tracks from hits using the combined tracker and muon systems are represented by red lines. Apart from the beam pipe, the highlighted CMS detectors are the pixel and ECAL barrel, ECAL and HCAL endcaps, HF calorimeter, and part of the muon system along with the matching chambers, CSC and DT segments. The measured kinematics of the reconstructed electron, muons (from  $t\bar{t}$  and B-hadron candidates, respectively), jets (yellow cones) and  $p_T^{\text{miss}}$  (not highlighted) can be found in Ref. [68]. The superimposed model of the CMS detector [69] is imported in the browser-based event display of Ref. [70].

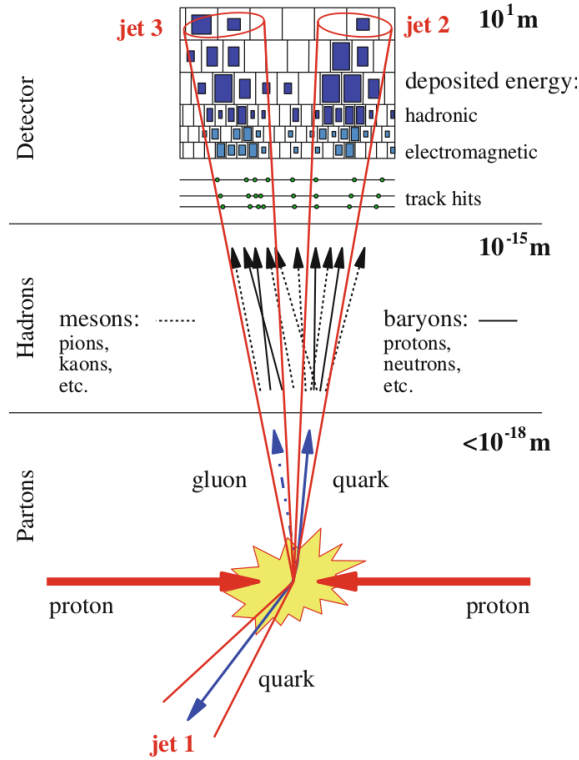
the ability to accurately track the paths of individual particles in a magnetic field, and in parallel, rely on extremely precise segmented calorimeters. The b quark lies within a jet as part of a hadron, then the latter decays roughly half a millimeter from where it was created. The charged particles in jets can be very precisely tracked using a silicon vertex detector with an accuracy of few  $\mu\text{m}$ . By identifying most of the tracks in a jet, and extrapolating them backwards towards the primary interaction vertex, the point where the bottom quark decayed could thereby be determined and the jet identified (“tagged”) as a b jet, based on its displacement relative to the primary interaction vertex.

## 1.2 The synchrotron legacy: a luminosity story

Assuming an arbitrary process  $X$ , the luminosity  $\mathcal{L}(t)$  is the process-independent proportionality factor between the event rate  $R(t)$  and its production cross section  $\sigma$ :

$$R(t) = \mathcal{L}(t)\sigma \quad 1.1$$

The luminosity of a particle collider is thus determined by the rate of particle collisions it can produce, making it an important measure for the performance of a collider. At circular colliders, it can be expressed in terms of the collision geometry and of the density



**Figure 1.7:** Illustration of a jet to which bundles of partons, hadrons, or detector measurements are grouped together [71]. The parton and particle levels, the latter containing color-neutral particles with mean decay lengths of at least  $c\tau > 10$  mm, are also denoted as the “truth level.” While traversing the detector, these particles interact with the detector material at the “detector level,” producing characteristic signatures, i.e., tracks and/or energy deposits.

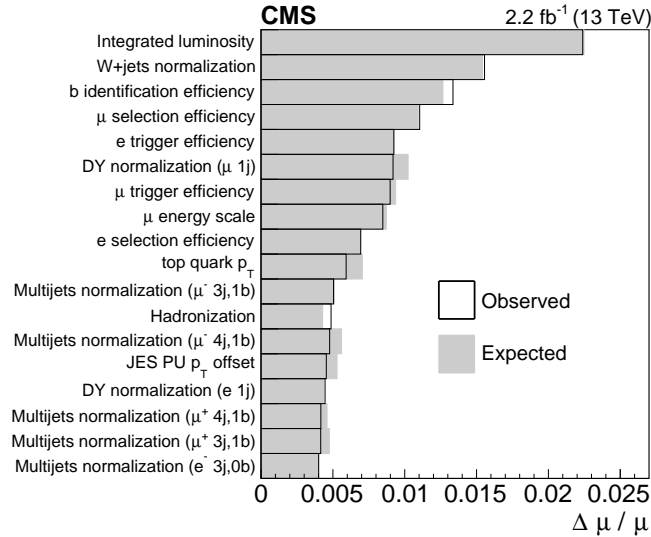
distribution of the counter-rotating beams. For colliders operating in “bunched mode” it is hence proportional to the number of bunches ( $n_b$ ), the product of the number of particles per bunch ( $N_1 N_2$ ), the revolution frequency in the apparatus ( $f_r$ ), yet inversely proportional to values related to the collision geometry, meaning [73]

$$\mathcal{L}(t) = \frac{n_b N_1 N_2 f_r}{A_{\text{eff}}}, \quad 1.2$$

where  $A_{\text{eff}} = 4\pi\sigma_x\sigma_y$  is the effective beam overlap cross section at the collision point;  $\sigma_x$  and  $\sigma_y$  are the horizontal and vertical beam sizes, respectively, that depend on the beam emittance ( $\epsilon$ ) and the demagnification value ( $\beta^*$ ) of the relevant optical function. Identical  $\epsilon$  and  $\beta^*$  in horizontal and vertical planes for both beams are assumed as well as geometric (zero crossing angle) and separation (beams with complete overlap) factors that equal unity.

The definition in Eq. (1.2) is referred to as the “instantaneous luminosity,” and is usually expressed in units of  $\text{cm}^{-2}\text{s}^{-1}$ . As running conditions vary with time, the luminosity of a collider also has a time dependency. The integral over time is called the “integrated luminosity” that is commonly denoted by  $\mathcal{L} = \int \mathcal{L}(t)dt$ , and is conveniently measured in units of inverse barns ( $1\text{b}^{-1} \equiv 10^{-24}\text{cm}^{-2}$ ). One further distinguishes into “delivered”

integrated luminosity, which refers to the integrated luminosity the machine delivered to an experiment or series of experiments, and into “recorded” integrated luminosity, which refers to the amount of collisions actually written to tape by the experiments. Maximizing the potential of colliders for delivering statistically significant data samples as well as the precise knowledge of the determination of the luminosity are of utmost importance for rare events, new phenomena, and many cross section measurements, e.g., Ref. [74] or as illustrated in Fig. 1.8.

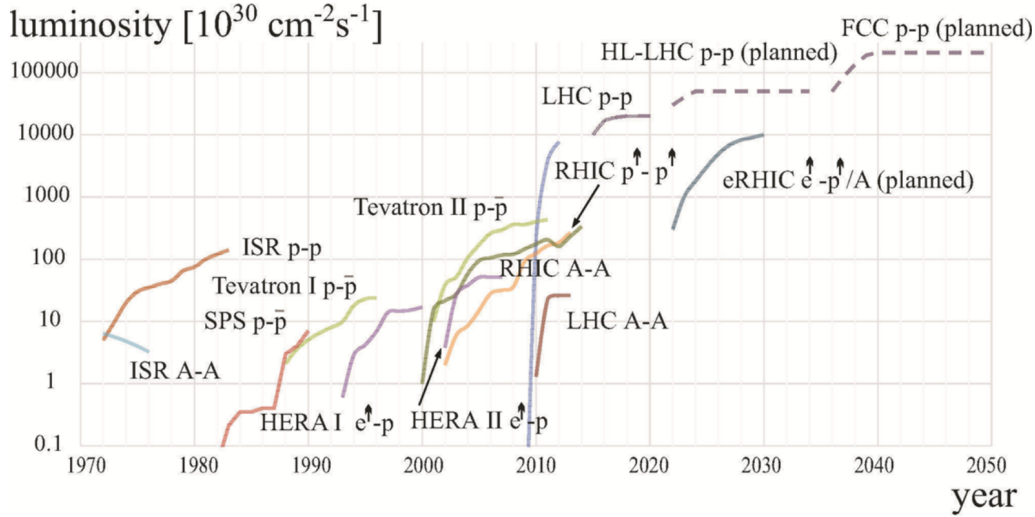


**Figure 1.8:** Estimated impact—change in the expected and observed signal strength—resulting from the listed sources of experimental and theoretical uncertainty in the  $\sigma_{\bar{t}t}$  analysis of Ref. [75]. The various contributions are shown from the largest to the smallest observed impact. The dominant source of uncertainty is related to the integrated luminosity, which is expected and observed to impact the signal strength at the level of 2.2%.

Limitations on the energy available in the center of mass system and luminosity [76]—two critical parameters coupled to the potential for discovery—are typically surpassed following the path of a somewhat lower energy that can be compensated for by a higher luminosity. The FNAL Tevatron was the first collider based on superconducting magnets with the colliding-beam operation having started in 1987. Tevatron luminosity was limited by  $\bar{p}$  intensity, beam-beam interaction, including long-range effects, luminosity lifetime, number of (simultaneous) events per crossing, and intrabeam scattering. RHIC at BNL—the first heavy ion collider at relativistic energies—delivers luminosity since 2000. The main limiting factor is intrabeam scattering. Other factors again are beam-beam interaction, luminosity lifetime, and the number of events per crossing.

LHC has been designed [77] to provide experiments with pp and PbPb luminosity of up to  $10^{34}$  and  $10^{27} \text{ cm}^{-2}\text{s}^{-1}$  (both already surpassed), respectively. It is the first machine where radiation damping is stronger than intrabeam scattering, while the scarcity of antiprotons is no longer a problem. Similarly to Tevatron, limits are the beam-beam interaction, luminosity lifetime, and events per crossing. In addition, the electron cloud produced by photoemission or beam-induced multipacting, and local magnet quenches

induced by the collision products could be important limiting factors. The main “path” to very high luminosity at LHC is therefore to use a high number of bunches and to reduce the transverse beam size at the experimental areas by manipulations of the magnetic focusing system at top energies. This can be combined with beams beyond the design brightness that the injector chain can produce, i.e., higher bunch intensities and normalized transverse emittance.



**Figure 1.9:** Evolution of the peak luminosity at hadron-hadron, lepton-hadron, and nucleus-nucleus colliders since 1970 (image courtesy of W. Fischer, BNL). The CERN ISR has been put into operation in 1970, reaching a peak luminosity of  $2.2 \times 10^{32} \text{ cm}^{-2} \text{ s}^{-1}$ , followed by the second hadron collider, i.e., the CERN SPS which operates since 1981 at ten times higher energy than achieved at ISR. Future projects like the High-Luminosity LHC [78], eRHIC [79] and a possible 100 km circumference collider (FCC) [80] are also indicated.

### 1.2.1 “A great accomplishment is not the end of the road, just the starting point for the next leap forward”<sup>1</sup>

At hadron collider experiments, protons, as well as heavy nuclei, are brought into collision, and the results are interpreted to obtain better knowledge about the dynamics of the fundamental interactions at high energies. The strong force plays a central role for the study of pp and heavy ion collisions. In the case of pp collisions, the general purpose MC event generators interfaced to PDF parameterizations succeeded in describing the dynamics of strong and EW processes from very high momentum transfer scales, where perturbation theory is applicable, down to scales around  $\Lambda_{\text{QCD}}$ . This has resulted in a remarkably precise description of the majority of observations in pp collisions, e.g., as shown in Fig. 1.10, so that further experimental and theoretical developments can rely upon. Since PDFs provide the essential link between the measurable hadronic and the perturbatively calculable cross sections, their precise determination is an extremely active area of research, where several groups perform global analyses including wide variety of experimental data.

<sup>1</sup>Harvey Mackay

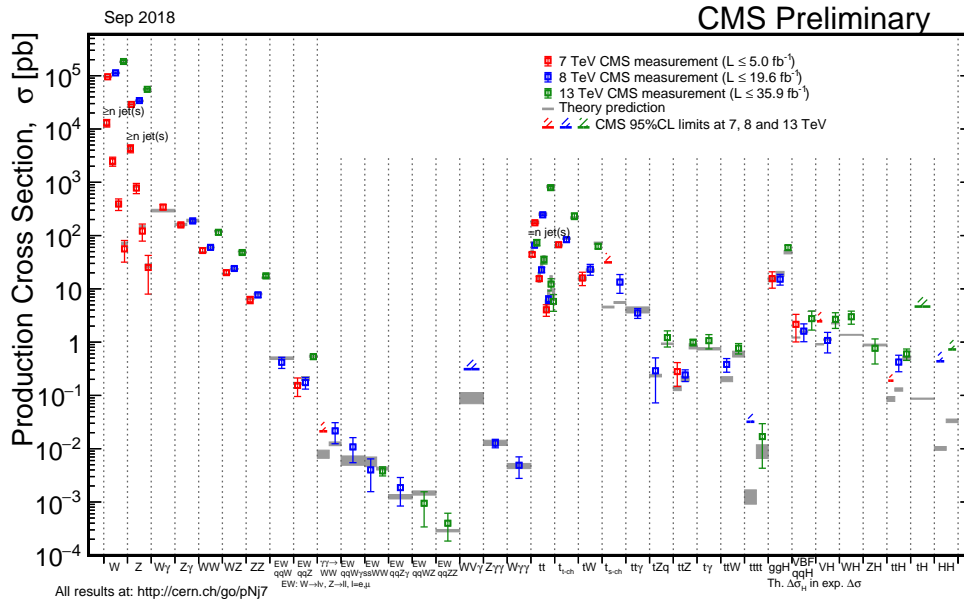


Figure 1.10: Compilation of CMS measurements and SM predictions in pp collisions at  $\sqrt{s} = 7, 8,$  and  $13$  TeV for a variety of processes as indicated on the horizontal axis [81].

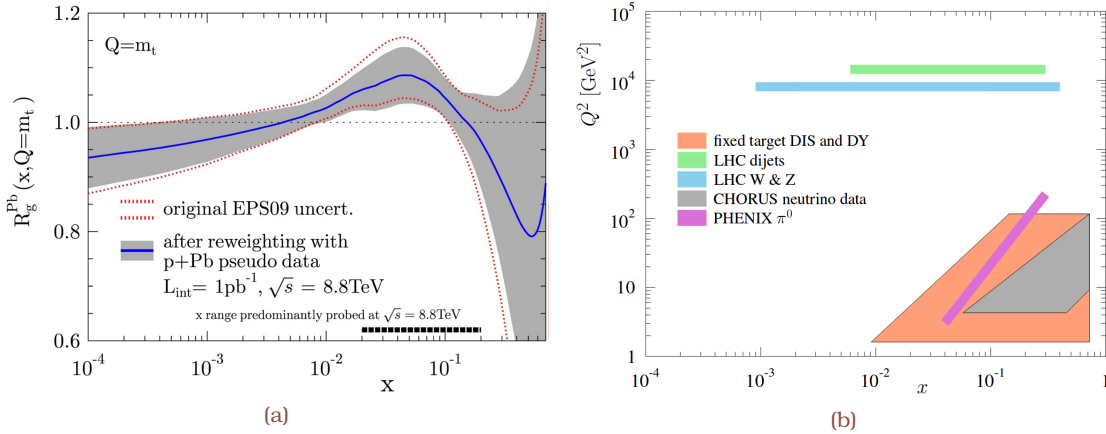
For high energy heavy ion collisions, efforts are often directed towards signals for the formation of the quark-gluon plasma (QGP), and studies of its properties, e.g., Ref. [82]. The existence of such a phase can be demonstrated in lattice QCD calculations [83], and it is conjectured to be present in the interior of compact, dense stellar objects, and also presumed to have existed in the early phase of the evolution of the Universe [84]. In this area event generators also exist, but are usually more “special” than event simulations of general purpose, each investigating different particle production mechanisms. The situation is thus more challenging, but not less interesting, for PDFs of nucleons inside nuclei (nPDFs) with nuclear data being significantly more complicated to collect, and with two additional degrees of freedom, i.e., the number of protons ( $Z$ ) and nucleons ( $A$ ) in the studied nuclei.

Since the early 1980s, it is known that the nuclei do not behave as a simple collection of free nucleons, and nPDFs are not equal to the sum of nucleon PDFs. The binding energies of nucleons in the nucleus are several orders of magnitude smaller than the momentum transfers of deep-inelastic scattering (DIS). Therefore, such a ratio naively should be close to unity, except for small corrections for the Fermi motion of nucleons in the nucleus at large  $x \gtrsim 0.8$ . Contrary to expectations, three regions of nuclear modification are well established for quarks:

- a depletion in the  $0.3 \lesssim x \lesssim 0.8$  (EMC),
- an enhancement in the  $0.1 \lesssim x \lesssim 0.3$  (“antishadowing”), and
- a dip in the  $x \lesssim 0.1$  (“shadowing”) regions.

Different physics mechanisms were proposed to explain this behavior at medium and large  $x$ , but a fully conclusive picture has not yet emerged after the discovery from the European Muon Collaboration experiment of a declining slope to the modification ratio [85], a result later confirmed with high-precision electron- and muon-scattering DIS data. At

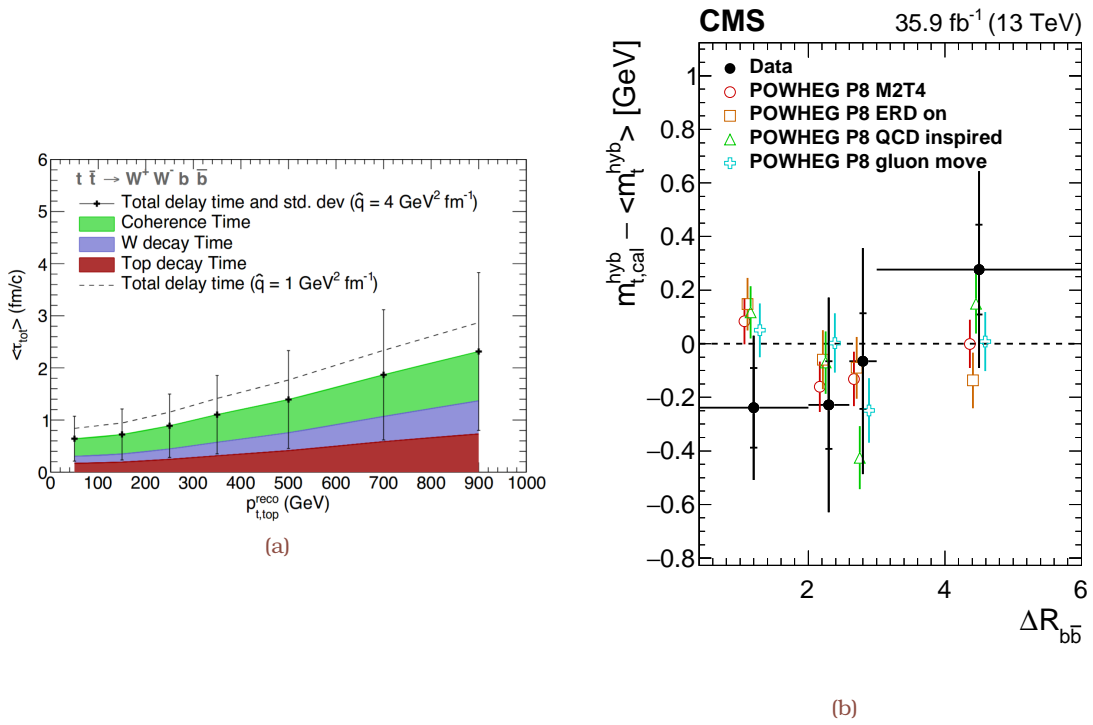
small  $x$ , coherent scatterings inside the nucleus explain the observed suppression, while antishadowing is even less understood. The conclusions from the combined experimental evidence demonstrates the consistency in modifying the free nucleon PDFs,  $f_i^p(x, Q^2)$ , at low  $Q^2$  and letting the Dokshitzer–Gribov–Lipatov–Altarelli–Parisi evolution to take care of the momentum transfer dependence, i.e.,  $f_i^{p/A}(x, Q^2) = R_i^A(x, Q^2)f_i^p(x, Q^2)$ , where  $R_i^A(x, Q^2)$  is the scale-dependent nuclear modifications encoded in nPDFs (Fig. 1.11a).



**Figure 1.11:** (a) Nuclear modification factor (bound-over-free parton densities ratio) for gluons as computed with EPS09 [86] at  $Q^2 = m_{\text{top}}^2$  (region enclosed by the red dotted lines) as a function of  $x$ . The ratio is also shown after a reweighting procedure using pseudo-experiments that correspond to an hypothesis about  $1 \text{ pb}^{-1}$  of pPb data at  $\sqrt{s_{\text{NN}}} = 8.8 \text{ TeV}$  [87]. At that  $\sqrt{s_{\text{NN}}}$  value the nuclear PDFs are predominantly probed in the range of  $0.02 \lesssim x \lesssim 0.3$ . (b) Approximate regions in the  $(x, Q^2)$  plane at which different data in the EPPS16 fit [88] probe the nuclear PDFs. For the first time, data constraints from LHC pPb collisions had been used.

Just like in the nucleon case, nPDFs are determined by performing global analyses of experimental data. Compared to the quark, the gluon content of the nuclei is even less known. To compensate this lack of constraints the most recent global NLO analyses of nPDFs, EPPS16 and nCTEQ15, use RHIC pion and LHC jet data (in case of EPPS16, Fig. 1.11b) to constrain the gluon densities down to about  $x = 10^{-3}$ . However, there had been no data that could exhibit enhanced sensitivity at  $x \lesssim 10^{-2}$  and EMC regions—at least till recently—and hence the gluon distribution still remains largely unconstrained. The nuclear PDF analyses assume that the bound proton PDFs have the same evolution equations and sum rules as the free proton PDFs, provided contributions from the region  $x > 1$  are neglected [89]. The lack of knowledge about the gluon nPDF is not reflected by the set of error PDFs, which are given together with the best-fit PDFs; it is accompanied by an increased flexibility of the initial nPDF parametrization that leads to increased uncertainty in this region, e.g., as evidenced by the EPPS16 set as opposed to its predecessor EPS09. Clearly, a determination of the small- and medium- $x$  gluon nPDFs and the reduction of their uncertainties is necessary for the heavy ion phenomenology. The relevance of experimental data on heavy-flavor [90] and dijet production [91, 92] has been lately demonstrated, clearly pointing to a shadowed and antishadowed gluon distribution at small- and medium- $x$  regions, respectively.



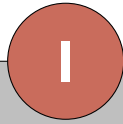


**Figure 1.12:** (a) Time delay between the moment of the primary collision and that when the W boson decay products start interacting with the QGP medium [93]. It is further split into the average contribution of separate components as shown from the colored stacked bands; vertical lines represent the dispersion in the sum of these three components. For comparison, the total delay time is shown for two different values of the transport coefficient of the medium,  $\hat{q}$ , the squared transverse momentum broadening per unit length. (b) A bias in the measurement of the top quark mass can be spotted by a failure of the color-reconnection modeling to reproduce measurements that offer the best sensitivity, e.g. the difference in the pseudorapidity and azimuthal angles (in radians) between the b jets assigned to the  $t\bar{t}$  decay [94].

Collisions of lead nuclei with each other (PbPb) and with protons (pPb) at LHC can guide performance projections either at High-Luminosity LHC, which is planned to be operated from 2026, or at a future hadron collider, in which LHC can be used as the final injector synchrotron. In heavy ion collisions, high transverse momentum ( $p_T$ ) hadronic processes are known to show strong medium-induced modifications, often referred to as “jet quenching,” up to at least the highest  $p_T > 100 \text{ GeV}$  explored at LHC so far [95]. The increase in energy and/or integrated luminosity at a future collider will provide much larger abundance of high- $p_T$  (“boosted”) processes.

Since the width of the top quark is much larger than the parton-to-hadron formation scale, the  $t/\bar{t}$  quarks are the only colored particles that decay before their hadronization, and their short lifetime implies they mostly decay within the lifetime of the strongly-interacting medium. A remarkable example is represented by the boosted  $t/\bar{t} \rightarrow W \rightarrow q\bar{q}'$  decay chain that is a promising color-singlet probe of the time evolution of the QGP density (Fig. 1.12a), and hence of the role of color coherence. Last but not least the reconstruction of the hadronic  $m_{t_{\text{top}}}$  in proton-nucleus interactions or in the QGP, assuming its feasibility

is not risked by the sizeable observed b quark energy loss [96], could provide insights into nonperturbative QCD effects on such a crucial parameter. The color flow through gluon exchanges and/or nonperturbative string overlaps between  $t$  and  $\bar{t}$ , the resulted b quarks, and the underlying event surrounding the initial hard scattering may lead to an uncertainty in the reconstructed top mass of a few hundred MeV (Fig. 1.12b). Indeed, the amount of top quark interactions with the color fields stretched among many partons involved in nuclear collisions will be enhanced compared to pp collisions.



Part A



## Contents

---

<b>2.1 The Large Hadron Collider complex</b> . . . . .	<b>34</b>
2.1.1 Performance and achievements of synchrotron accelerators . . . . .	34
2.1.2 Delivering collisions to the seven LHC experiments . . . . .	35
2.1.3 Fundamental principles of synchrotron accelerators . . . . .	38
2.1.4 LHC layout and global structure . . . . .	43
2.1.5 Proton-nucleus collisions at LHC . . . . .	47
<b>2.2 The Compact Muon Solenoid detector</b> . . . . .	<b>51</b>
2.2.1 The superconducting solenoid magnet . . . . .	54
2.2.2 The inner silicon pixel and the larger silicon strip tracker . . . . .	54
2.2.3 The electromagnetic calorimeter . . . . .	56
2.2.4 The hadron calorimeter . . . . .	58
2.2.5 The muon detectors . . . . .	59
2.2.6 Luminosity detectors at CMS . . . . .	61

---

The Large Hadron Collider (LHC) is a hadron accelerator with a circumference of 26.7 km designed to operate separately with proton and lead beams with energies of up to 7 and 7 Z TeV, respectively. It features four experimental insertions: two aiming at instantaneous luminosity of at least  $10^{34} \text{ cm}^{-2}\text{s}^{-1}$ , and the rest dedicated to “medium” ( $\sim 10^{32} \text{ cm}^{-2}\text{s}^{-1}$ ) and “low” ( $\sim 10^{29} \text{ cm}^{-2}\text{s}^{-1}$ ) interaction rates using pp collisions. The CERN heavy ion injectors and LHC have demonstrated the feasibility of proton-lead as well as xenon-xenon collisions. Bunches can be injected into the LHC in every 10th of the 35 640 RF buckets, a spacing conventionally referred to as “25 ns,” although other spacings exist and can be any multiple of 25 ns. The minimum spacing between heavy ion bunches, as dictated from the injectors, is nominally 100 ns. LHC has very tight tolerances on nearly all beam parameters, meaning their precise measurement is crucial for controlling and understanding the apparatus. With at least two orders of magnitude higher stored beam energy than previous colliders machine protection is also imperative.

The Compact Muon Solenoid (CMS) detector is one of the two general-purpose experiments operating at LHC. Its central feature is a superconducting solenoid of 6 m internal diameter. Within the solenoid volume are a silicon pixel and strip tracker, a lead tungstate crystal electromagnetic calorimeter, and a brass and scintillator hadron calorimeter, each composed of a barrel and two endcap sections. Muons are detected in gas-ionization chambers embedded in the steel flux-return yoke outside the solenoid. The forward hadron calorimeter uses steel as an absorber and quartz fibers as the sensitive material, and extend the coverage in the range  $3.0 < |\eta| < 5.2$ .

## 2.1 The Large Hadron Collider complex

### 2.1.1 Performance and achievements of synchrotron accelerators

Within a synchrotron accelerator, charged particles can gain energy from a high amplitude alternating  $\vec{E}$  field that is contained in metallic chambers, i.e., radiofrequency (RF) cavities, while they are steered and focused by magnets. At relativistic velocities  $v = \beta c$ , such an extra factor in the magnetic force makes a large difference in the scale of the needed magnetic compared to electric field: A magnetic field of 1 T would be equivalent to a gargantuan electric field of  $3 \times 10^8 \text{ Vm}^{-1}$  [97]. Producing such an electric field is far beyond current technical limits, meaning magnetic fields are copiously used to steer the beams. The physical fundamentals of beam steering and focusing are the so-called “beam optics.”

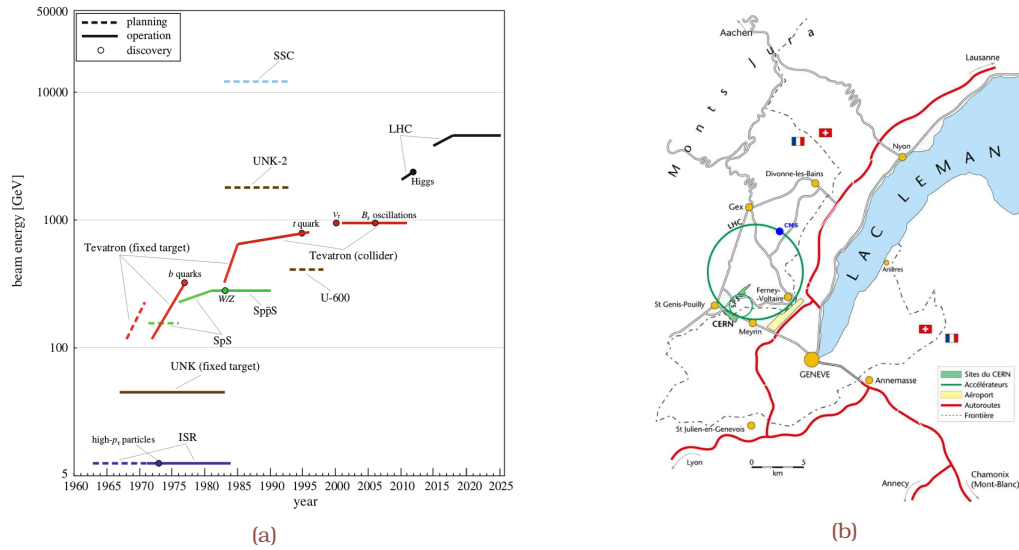
Although the advantage of directing two beams of approximately equal energy at each other with respect to fixed-target collisions (regarding usable energy) had been early realized, the low particle intensities obtained in accelerators in the late 1950s rendered colliders an impractical option. This changed in 1957 (Fig. 2.1a), when the idea of stacking particles into circular accelerators was first put forward with an alternating-gradient accelerator [98]. Collider storage rings can in principle be designed for a variety of particle species.

Two main categories of synchrotron accelerators can be distinguished in the field of collider physics, namely hadron and lepton colliders. For instance, the Large Electron-Positron collider (LEP) was the previous large-scale project at CERN, i.e., an electron–positron accelerator that was constructed between 1984 and 1989 and whose ring circumference was almost 27 km. The synchrotron radiation produced from bending the particle trajectory is highly dependent on the mass to charge ratio as

$$P \propto \frac{E^4 p}{\rho^2 m^4}, \quad 2.1$$

where  $P$  is the power radiated from the steered particle,  $E$  is the energy of the particle,  $\rho$  is the radius of curvature of the trajectory, and  $m$  is the rest mass of the particle. The energy loss due to synchrotron radiation was 3% per turn in the LEP apparatus; were the RF cavities turned off, the beam would be lost only in a fraction of a turn. To counteract the energy loss RF cavities were extensively installed, rendering LEP the largest circular lepton collider built so far. For the specific case of protons, the emitted radiation is approximately eight orders of magnitude lower, for the same beam energy and accelerator size, relative to electrons. After LEP was dismantled, the construction of a hadron collider in the same tunnel started in 2001 (Fig. 2.1b), although the main idea for a multi-TeV hadron collider to investigate the origin of mass and to search for signs of unification beyond the SM was initially conceived in the late 1970s and early 1980s.

The Large Hadron Collider (LHC) is a hadron accelerator with a circumference of 26.7 km designed to run at a nominal collision energy of 14 TeV, approximately 7 times higher than that of the Tevatron collider. The key objective of LHC is the exploration of the SM in the



**Figure 2.1:** (a) Overview of  $pp$  and  $p\bar{p}$  colliders, their beam energies and major achievements [99]. The nominal LHC design aims at proton beam energies of 7 TeV. (b) LHC is a two-ring superconducting hadron accelerator. The 26.7 km tunnel has eight straight sections and eight arcs, and lies between 45 and 170 m below the surface on a plane inclined at 1.4% sloping towards the Léman lake [100]. Approximately 90% of its length is immersed in molasse rock, and the remaining 10% lies in limestone under the Jura mountain. There are two transfer tunnels, each approximately 2.5 km long, linking LHC to the CERN accelerator complex that acts as injector.

TeV energy range and the search for potential new physics signatures. The LHC design parameters, as shown in Table 2.1, had meanwhile evolved significantly over the following years in a continuous comparison and competition with the Superconducting Super Collider (SSC) project—nicknamed the “Desertron”—in the United States leading to a tenfold increase of the LHC performance from the first design projections [101] to the nominal performance specification [102]. The maximum achieved energy, or more precisely, the beam momentum is given by the maximum bending field strength, and it follows a fairly clear exponential trend as a function of time, as also shown in Fig. 2.1a. Hadron accelerators can reach higher energies than their lepton counterparts, as Eq. (2.1) delineates. For that reason, they are deemed better suitable to explore new energy regimes, albeit the collisions are intrinsically more complex. The use of protons implies that the collisions occur actually between the proton constituents that carry only a fraction of the total proton beam energy. Aiming at having at least 10 TeV collision energy requires proton beam energies well above 1 TeV.

### 2.1.2 Delivering collisions to the seven LHC experiments

During the 1992 workshop “Towards the LHC Experimental Program” in Evian, proto-collaborations presented “expressions of interest” describing their detector plans [103]. The interest in contributing to the LHC experimental program was enormous: 12 proposals were made in total, out of which four intended for general-purpose experiments, three meant for b-flavor and heavy ion physics, respectively, while two additional proposals were

Parameter	LHC	LEP2	Tevatron
Colliding species	p,p	e <sup>+</sup> ,e <sup>-</sup>	p, $\bar{p}$
Dipole field at top energy (T)	8.33	0.11	4.4
Momentum at collisions (TeV)	7	0.1	0.98
Number of bunches per beam	2 808	4	36
Particles per bunch ( $\times 10^{11}$ )	1.15	4.2	2.9, 0.8
Typical beam size in the ring ( $\mu\text{m}$ )	200-300	1 800/140 (H/V)	500
Beam size at IP ( $\mu\text{m}$ )	16	200/3 (H/V)	24
Peak luminosity ( $\text{cm}^{-2}\text{s}^{-1}$ )	$10^{34}$	$10^{32}$	$4.3 \times 10^{32}$
Fraction of energy lost in synchr.rad. per turn	$10^{-9}$	3%	$10^{-11}$
Total current per beam (A)	0.58	0.003	0.08
Total energy stored in each beam (MJ)	362	0.03	0.9

**Table 2.1:** The main LHC parameters are compared to its predecessor during the final operational phase, i.e., LEP2, and Tevatron. The total number of bunches in the LHC ring was reduced from the initial 2 835 [77] to 2808 [102], translated into a modest reduction in luminosity. Synchrotron radiation from protons at LHC energies becomes noticeable but is not a limitation. The LHC is built with superconducting NbTi magnets that operate at superfluid  $^4_2\text{He}^+$  temperature of 1.9 K and allow fields up to 8.33 T. The notations “H” and “V,” that are used in the beam size, stand for the horizontal and vertical plane, respectively.

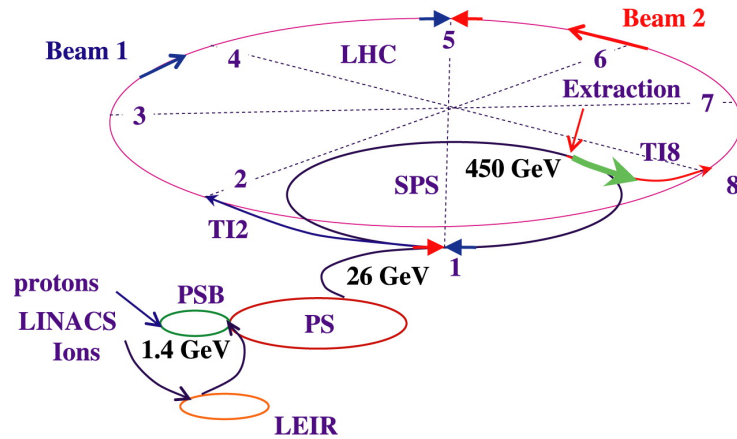
targeting to neutrino physics. LHC beams could in principle collide at eight points. Four of these coincided with the four big experiments at LEP, while the remaining four points had to be used for the “beam-cleaning” system to ensure high performance by reducing troublesome beam halo. Another was reserved for the beam dump where protons can be absorbed, once the circulating beams are no longer required.

It was clear from the beginning that only two new general-purpose experiments would be accepted at LHC, one of them potentially being a toroidal apparatus like the Apparatus with SuperCONducting Toroids (ASCOT) and the Experiment for Accurate Gamma, Lepton and Energy measurements (EAGLE). Therefore, these two proto-collaborations, in a voluntary move, merged to form A Toroidal LHC ApparatuS (ATLAS) in 1992. Among the remaining general-purpose experiments, ATLAS and Compact Muon Solenoid (CMS) were invited to provide detailed technical proposals by 1994. ATLAS and CMS were finally accepted in January 1996, and later approved for construction on 31st of January 1997 with an expenditure ceiling of 475 MCHF [104] (1995 currency rate). The resulted four big experiments, i.e., ALICE, ATLAS, CMS and LHCb, were followed by three smaller, more focused proposals for experiments: The TOTEM experiment (LoI 1997) is investigating the total pp cross section, elastic pp scattering and diffraction dissociation; MoEDAL (LoI 1998) is searching for magnetic monopoles and other exotic phenomena; LHCf (LoI 2003), finally, uses very forward particles created in pp collisions to simulate cosmic rays.

The acceleration of protons at the CERN complex is performed in stages as they pass through the different accelerators along the injector chain, shown Fig. 2.2. The protons are generated by a duoplasmatron source from which they are extracted with an energy of 100 keV and injected into LINAC2<sup>1</sup>, i.e., an 80 m long linear accelerator with an extraction

<sup>1</sup>The Linear accelerator 4 (Linac4) is designed is scheduled to become the source of proton beams for LHC





**Figure 2.2:** The accelerator complex at CERN is a succession of machines with increasingly higher energies [105]; two beams circulate in LHC following a clockwise and an anticlockwise direction. They are accelerated up to the record energy of 6.5 TeV. Most of the other accelerators in the chain have their own experimental halls, where their beams are used for experiments at lower energies. In addition to accelerating protons, the CERN complex can also accelerate lead ions by first passing them through LINAC3 and then LEIR before continuing on the same path as protons.

energy of 50 MeV. During the acceleration process in the LINAC the proton beam is also bunched using RF cavities. Once extracted from the LINAC the protons are injected into the Proton Synchrotron Booster (PS Booster), a 157 m circular accelerator complex capable of accelerating protons up to 1.4 GeV, which consists of a stack of four separate rings. From the PS Booster the particles are injected into the Proton Synchrotron (PS), a 628 m ring where they are accelerated up to an energy of 26 GeV. All these machines are installed at ground level. At this point the beams are sent in an underground machine, the Super Proton Synchrotron (SPS), which is a 6.9 km circular accelerator lying 50 m under the surface. They are then accelerated to 450 GeV, corresponding to the LHC injection energy. The injection of the LHC beams from the SPS is done in IR2 for beam 1, going clockwise, and IR8 for beam 2, going counter-clockwise. All these accelerators are linked through transfer lines which allow for the particles to travel from one machine to the other.

The LHC has also been designed to collide heavy ions that follow a slightly different path relative to protons. They are produced from a highly-purified lead sample—enriched with the isotope  $^{208}\text{Pb}$ —heated to a temperature of about 800°C where the lead vapor is ionized by an electron current. The  $^{208}\text{Pb}^+$  ion state is selected and accelerated to a energy of 4.2 MeV (per nucleon) before it is directed towards a carbon foil, the latter stripping most of them to  $^{208}\text{Pb}^{82+}$ . The formed  $^{208}\text{Pb}^{82+}$  beam is first accumulated, then accelerated to 72 MeV in the Low Energy Ion Ring (LEIR), which transfers it to the PS (5.9 GeV), and finally sent to the SPS (177 GeV) fully stripped to  $^{208}\text{Pb}^{82+}$ , and after passing through a second foil. The SPS finally injects the beams to the LHC, that accelerates them to the record beam energy of 6.5 ZTeV.

The original design of LHC did not foresee the operation with species other than protons or  $^{208}\text{Pb}^{82+}$  ions, let alone a mixed particle mode. However, the CERN heavy ion

after the long shutdown 2

injectors and LHC have demonstrated the feasibility of asymmetric (see Section 2.1.5) and xenon-xenon, i.e., a medium-mass nuclear species, collisions. In 2017, high-intensity xenon ( $^{129}_{54}\text{Xe}^+$ ) beams were produced in the injector chain, and brought into collision at a nucleon-nucleon centre-of-mass energy ( $\sqrt{s_{\text{NN}}}$ ) of 5.44 TeV [106]. During 6 h of stable XeXe collisions about  $3 \mu\text{b}^{-1}$  were delivered to ATLAS and CMS, while fractions of  $1 \mu\text{b}^{-1}$  were delivered to ALICE and LHCb because of the larger demagnification values. Such production outcome is comparable to the  $10 \mu\text{b}^{-1}$  of PbPb collisions delivered per experiment in the first one month-long heavy ion run in 2010 [107]. Since this was probably the last time—at least for several years—that a species other than Pb would be available from the injector complex valuable data were acquired on the beam-cleaning and collimation efficiency with lighter ions [108, 109]. In 2018, for the very first time, operators injected and accelerated not just atomic nuclei but partially stripped  $^{208}_{82}\text{Pb}^{81+}$  ions into LHC. This represents a proof-of-principle test for broadening the present CERN research program making use of such a novel concept of light source [110].

### 2.1.3 Fundamental principles of synchrotron accelerators

#### Transverse particle motion

If a particle is deflected in the presence of a magnetic field, the Lorentz and the centripetal force—directed perpendicular to the direction of motion—are always equal. Using a Taylor expansion, the magnetic field transverse components,  $B_x \hat{x}$  and  $B_y \hat{y}$  (Fig. 2.3a), can be expressed as a function of a dipolar and quadrupolar term:

$$B_x = B_x(0, 0) + \frac{\partial B_x}{\partial y} y + O(y^2) \quad 2.2$$

and

$$B_y = B_y(0, 0) + \frac{\partial B_y}{\partial x} x + O(x^2), \quad 2.3$$

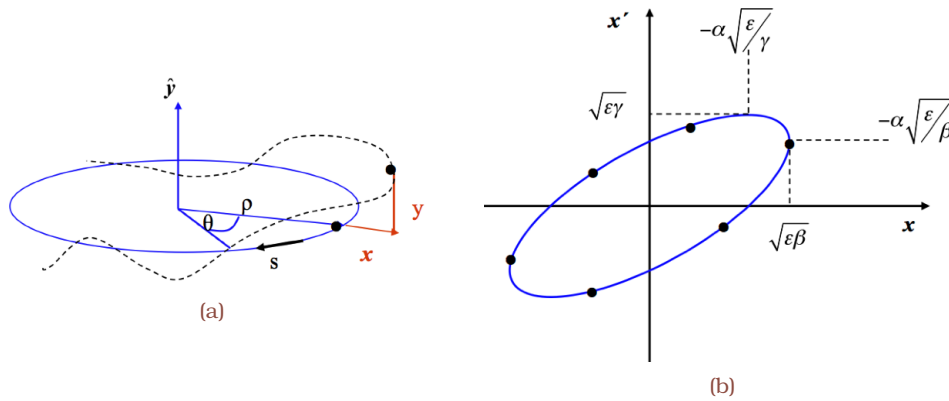
respectively. The total bending angle ( $\vartheta$ ) of a circular accelerator is  $2\pi$ , and the integrated dipole field is

$$\int_{s_1}^{s_2} B dl = \frac{2\pi p}{q} = 2\pi B\rho. \quad 2.4$$

This defines the curvature ( $\rho$ ) of a particle with charge  $q$  and momentum  $p$  in the magnetic field of strength  $B$ , from which the so-called “beam rigidity” ( $B\rho$ ) can be derived:

$$B[\text{T}]\rho[\text{m}] \approx \frac{p[\text{GeV}]}{0.3}. \quad 2.5$$

Since synchrotron radiation for hadrons is less of an issue than leptons, the maximum attainable energy in hadron colliders is limited by the magnetic field strengths available for bending and focusing the particles on their circular design trajectory. In the case of LHC, the magnetic bending radius is determined by the LHC tunnel and it is restricted to  $\rho = 2804$  m. To fulfill the designed proton beam momentum of 7 TeV the maximum required magnetic dipole field has to be at least 8.33 T (Eq. (2.5)). This relatively high field can only efficiently be provided by superconducting magnets. The accelerator lattice is designed



**Figure 2.3:** (a) The trajectory coordinates with respect to the Frenet–Serret frame [111]. An ideal particle is performing transverse oscillations, as indicated by the dotted line, around the design orbit. (b) The motion of a single particle, at an arbitrary longitudinal location  $s$ , defines an ellipse in phase space  $(x, x')$  characterized by the parameters  $a$ ,  $\beta$ , and  $\gamma$  [111]. The beam emittance, i.e., the area in the  $(x, x')$  space that contains 68.3% of an ensemble of particles, equals to the area  $\pi\varepsilon = a^2$  enclosed by the ellipse of a single particle; the beam width and divergence can be then proven to be  $\sqrt{\beta\varepsilon}$  and  $\sqrt{\gamma\varepsilon}$ , respectively.

to mainly provide bending and focusing fields for a reference particle of a defined particle species. For particles of the main beam species (monoisotopic case), dispersive effects arise only from momentum offsets. For particles of other species, additional dispersive offsets are caused by the different mass and charge with respect to the reference isotopes  ${}_{Z_0}^{A_0}X_0^+$  (see Section 2.1.5).

To describe the particle trajectories in a synchrotron, the task for solving the equation of motion relative to the design orbit is simplified using a comoving coordinate system, known as the “Frenet–Serret” coordinate system (Fig. 2.3a). Around the accelerator, the focusing properties ( $K$ ) of dipoles and quadrupoles are not constant but depend on  $s$ . However,  $K(s)$  is periodic with the lattice period ( $L$ ), e.g.,  $L$  can be the circumference of the accelerator, leading to a second order homogeneous differential equation for the transverse motion of a particle in the magnetic structure of an accelerator

$$x''(s) + K(s)x(s) = 0. \quad 2.6$$

This type of motion with non constant but periodic restoring force is described by the so-called “Hill equation,” and its general solution is a quasi-harmonic “betatron” oscillation:

$$x(s) = \sqrt{\varepsilon\beta(s)}\cos(\psi(s) + \varphi). \quad 2.7$$

The amplitude and phase of the oscillation depend on the exact position in the ring;  $\varepsilon$  and  $\varphi$  are integration constants that can be defined from an initial position  $x_0$  and angle  $x'_0$  at location  $s(0) = s_0$  and  $\psi(0) = 0$ . The so-called “beta function,”  $\beta(s)$ , is also a periodic function of  $L$ ,  $\beta(s + L) = \beta(s)$ , and is determined numerically by the focusing properties of the lattice; the value of the  $\beta$  function at the IP is colloquially known as  $\beta^*$ . The number

of betatron oscillations per turn, i.e., the machine “tune,”

$$Q = \frac{\psi(s)}{2\pi} = \frac{1}{2\pi} \oint_L \frac{ds}{\beta(s)}, \quad 2.8$$

is of great importance for beam stability, since it regulates the orbit response around the ring for dipole field (terms of  $\sin(\pi Q)$ ) and gradient (terms of  $\sin(2\pi Q)$ ) errors. With the trajectory at any point of the ring expressed in terms of position and angle,  $\varepsilon$  can be calculated at any point as

$$\varepsilon = \gamma(s)x(s)^2 + 2a(s)x'(s) + \beta(s)x'(s)^2, \quad 2.9$$

where two commonly used functions

$$a(s) = -\frac{1}{2} \frac{\partial \beta(s)}{\partial s}, \quad 2.10$$

and

$$\gamma(s) = \frac{1 + a(s)^2}{\beta(s)}, \quad 2.11$$

have been introduced. The quantities  $\beta(s)$ ,  $a(s)$  and  $\gamma(s)$  are the so-called “Twiss parameters,” also referred to as the optical functions, since they are defined by the magnetic lattice of the machine that transforms the beam equivalently to a lattice of lenses in classical optics. The evolution of an initial set of Twiss parameters is equivalent to the transformation of the particle coordinates described by transfer matrices for the individual beamline elements of length  $L$  and strength  $K$ , i.e., the transfer matrices of a drift space ( $M_D, K = 0$ ), a focusing ( $M_{Q,f}, K > 0$ ) and defocusing ( $M_{Q,d}, K < 0$ ) quadrupole.

The parametric representation in the  $(x, x')$  space (Fig. 2.3b) defines a constant of motion in the absence of nonconservative forces; the area of the ellipse  $\pi\varepsilon$  is a phase space invariant. This constant describes the amplitude of a single particle, and it can be further generalized to an intrinsic beam property. Assuming a Gaussian profile for the particle distribution in the accelerator, normalized to unity, the area in the  $(x, x')$  space that contains 68.3% of an ensemble of particles, i.e., the “beam emittance” (Fig. 2.4), is defined as

$$\varepsilon_x = \langle x^2 \rangle \langle x'^2 \rangle - \langle xx' \rangle^2 \quad 2.12$$

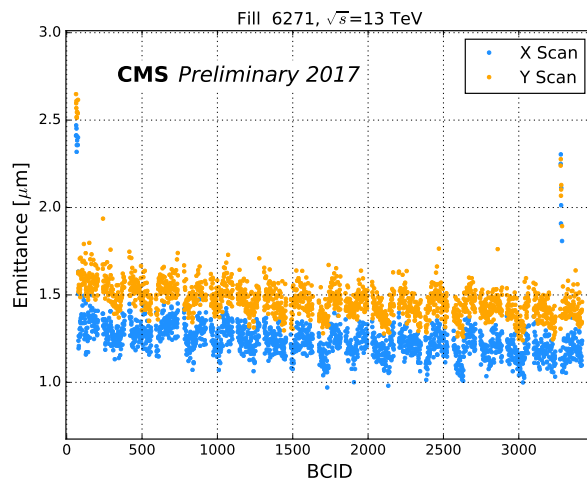
such that the standard deviation of the Gaussian distribution corresponds to

$$\sigma_x = \langle x^2 \rangle = \sqrt{\varepsilon \beta_x} \quad 2.13$$

and

$$\sigma_{x'} = \langle x'^2 \rangle = \sqrt{\varepsilon \beta_x \gamma_x} \quad 2.14$$

in  $x$  and  $x'$ , respectively. In Eq. (2.12) the symbols  $\langle \rangle$  denote the expectation value over the considered ensemble, while  $\sigma_x$  and  $\sigma_{x'}$  are referred to as the beam size and divergence, respectively.



**Figure 2.4:** The (normalized) beam emittance values in the horizontal and transverse plane per bunch crossing identification number calculated based on the model of Ref. [112], and using as inputs the convolved beam widths measured by CMS [113]. The bunch length values are the published entries in the logging database from LHC, whereas the nominal values of the crossing angle, the relativistic factor, and  $\beta^*$  are used.

### Longitudinal particle motion

Charged particles are traveling on two design orbits in opposite directions and are colliding at the interaction point(s) (IP(s)). The RF system of a synchrotron is required to operate in three different modes [114]. *It must capture and accumulate in stationary buckets (closed trajectories in longitudinal phase space) successive groups of bunches from its injector; accelerate these bunches in moving buckets up to the design energy; and finally store them, for several hours, while maintaining a minimum ratio of bunch-to-bucket area. In the storage mode, it provides the nominal energy gain per turn required to make up for synchrotron radiation losses and for the power loss due to voltages induced in the impedance presented to the beam by the vacuum chamber and accelerating structures.* More specifically, the angular radiofrequency,  $\omega_{\text{RF}}$ , needs to be synchronous to the angular revolution frequency,  $\omega_{\text{r}}$ . Synchronous particles need to repeatedly arrive at the cavity with the same phase. This implies that  $\omega_{\text{RF}}$  has to be an integer multiple of  $\omega_{\text{r}}$

$$\omega_{\text{RF}} = h\omega_{\text{r}},$$

2.15

where  $h$  is an integer and is called the “harmonic number.” As a consequence, the number of stable synchronous particle locations equals the harmonic number  $h$ ; they are equidistantly spaced around the circumference of the accelerator. All synchronous particles have the same nominal energy and follow the nominal trajectory.

For synchrotron accelerators, the metric that regulates the stability of the longitudinal motion is the scale of the so-called “transition energy.” This energy is a property of the transverse lattice and is linked to the stability of the revolution frequency,  $\omega_{\text{r}}$ , for particles with a small momentum deviation relative to the synchronous particle; the momentum-dependent change of velocity and path-length compensate each other at the transition

energy. Yet, the change of revolution frequency with momentum offsets is opposite below and above transition, featuring a confining range for stable longitudinal oscillations, as described in the following.

### Dispersion effects in longitudinal and transverse dynamics

If a particle is slightly shifted in momentum relative to a synchronous particle, it will have a different velocity and orbit, hence orbit length. The “momentum compaction” parameter quantifies the relative change in orbit length,  $\Delta L/L$ , with momentum, i.e.,

$$a_c = \frac{\Delta L/L}{\Delta p/p}. \quad 2.16$$

At the transition energy there is no change of the revolution frequency for particles with a small momentum deviation. Below or above this critical point the longitudinal oscillations are governed by the so-called “slip factor,”  $\eta$ , which is given by

$$\eta = \left( \frac{1}{\gamma^2} - a_c \right). \quad 2.17$$

From the definition of  $\eta$ , it is clear that an increase in momentum is linked to the opposing effects of

1. a higher revolution frequency below the transition energy ( $\eta > 0$ ) resulting to an increase of velocity
2. a lower revolution frequency above the transition energy ( $\eta < 0$ ) resulting to an increase of path length.

Since the change in revolution frequency with momentum is opposite below and above transition, this is reflected on the range for stable oscillations. For small phase deviations,  $\Delta\phi$ , with respect to the synchronous particles, the conditions for longitudinal stability can be derived assuming constant  $R_s, p_s, \eta, \omega_{r,s}$  parameters

$$\frac{d^2\phi}{dt^2} + \frac{\Omega_{r,s}^2}{\Delta\phi} = 0. \quad 2.18$$

Eq. (2.18) corresponds to an harmonic oscillator equation of motion with

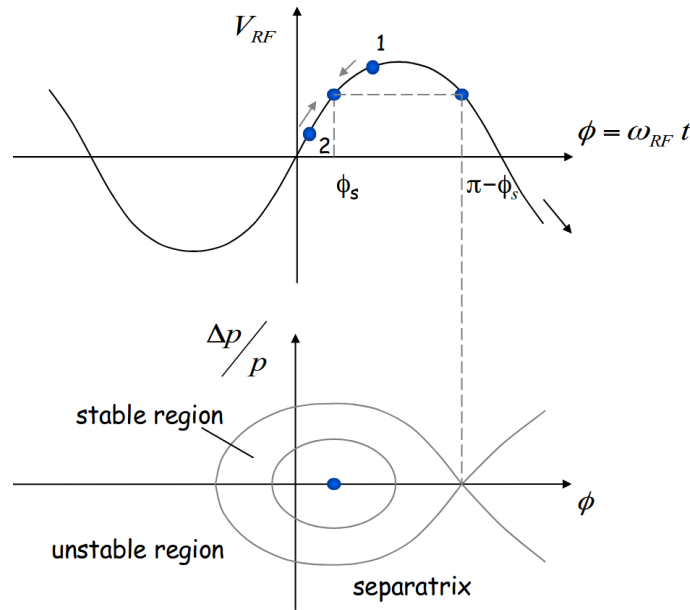
$$\Omega_{r,s} = \frac{h\eta\omega_{r,s}qV_0\cos\phi_s}{2\pi R_s p_s} \quad 2.19$$

the so-called “synchrotron angular frequency.” Stability is obtained when  $\Omega_{r,s}$  is a real number so that  $\Omega_{r,s}^2 > 0$ . Since most of the terms in Eq. (2.19) are positive, this finally reduces to

$$\eta\cos\phi_s > 0, \quad 2.20$$

rendering the stable region for the synchronous phase dependent on the energy with respect to the transition energy (Fig. 2.5).

In the case of LHC, the momentum compaction factor is of the order of  $10^{-4}$ , and the zero-crossing condition of Eq. (2.17) implies  $\gamma_{Tr} = \frac{1}{a_{c,Tr}} \sim 100$  meaning LHC operates



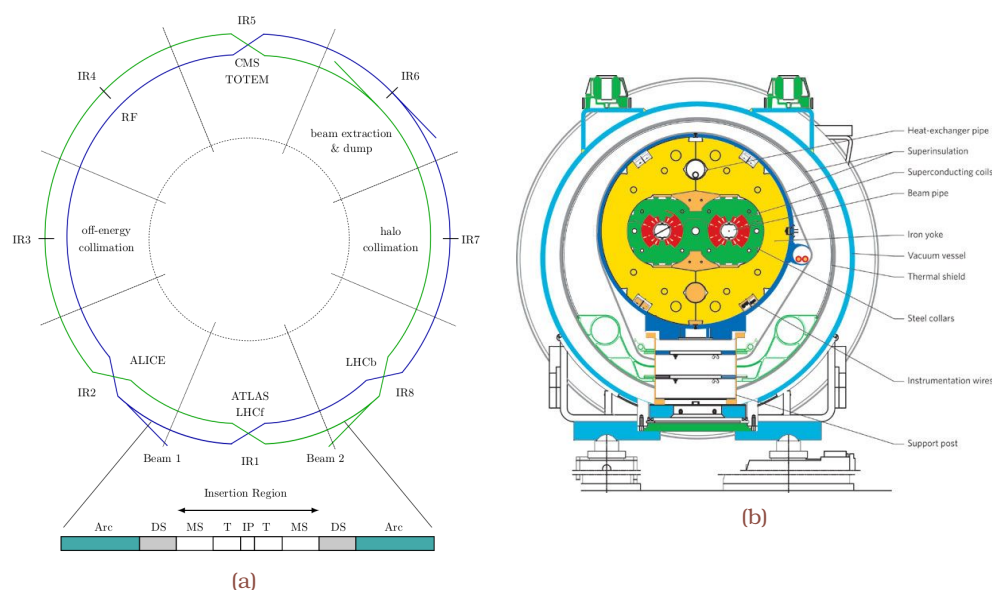
**Figure 2.5:** The applied longitudinal electric field as a function of the particle phase  $\phi$ . Particles that arrive later (1) than the reference particle receive a larger energy transfer, while those arriving before the reference particle (2) receive a smaller energy kick. The restoring force vanishes when  $\phi$  approaches  $\phi - \phi_s$ , and becomes nonrestoring either above or below the transition point (Eq. (2.17)). This phase space trajectory—the separatrix—separates the region of stable motion from the unstable region, contrary to the area within the separatrix—the RF bucket—which corresponds to the maximum acceptance in phase space for a stable motion [111].

always above transition. The cavities at LHC operate at  $f_{RF} = 400 \text{ MHz}$  that is a multiple of the revolution frequency  $f_r = 11\,245.5 \text{ Hz}$  [115] with an harmonic number  $h = 35\,640$  (Eq. (2.15)). The harmonic number defines the maximum number of longitudinally stable regions, referred to as “buckets,” that can be used to capture, store and accelerate particle species. The collectivity of those species inside a bucket is called “bunch.” At LHC only about every tenth bucket will nominally capture a bunch, which accordingly corresponds to a bunch spacing of about 25 ns.

#### 2.1.4 LHC layout and global structure

A circular collider project can follow two distinct design options: either the collider features collisions between particles and antiparticles with opposite charges, allowing an efficient accelerator design where both beams share the same vacuum chamber and magnetic elements, e.g., in LEP, Sp $\bar{p}$ S and Tevatron colliders, or the collider features separate vacuum and magnet systems for the two counter-rotating beams, e.g., in the ISR or RHIC collider, allowing collisions between a wider range of particle species and larger number of bunches resulting in higher luminosity. Hadron colliders relying on particle and antiparticle collisions are intrinsically limited by the rate at which antiparticles can be generated. The high-luminosity requirements at LHC, i.e., luminosity in excess of  $10^{34} \text{ cm}^{-2}\text{s}^{-1}$ , exclude the use of  $\bar{p}$ , and hence require a two-ring design with separate magnet and vacuum

systems for the two counter-rotating beams.



**Figure 2.6:** (a) Schematic layout of the LHC with its eight straight sections and two-ring design [105]. There are four experimental insertions of similar design (IR1, 2, 5 and 8) that allow particles of the same (proton or nucleus) or unequal (proton and nucleus) charge to be focused and eventually collide. The other four long straight sections are used for collimation (IR3 and 7), acceleration using RF cavities (IR4) and beam extraction (IR6). The two counter-rotating beams are conventionally called beam 1 and 2, which circulate in clockwise and counter-clockwise directions, respectively. (b) Schematic cross section of the LHC dipole magnet cryostat with its two separate vacuum chambers and magnet coils for the two counter-rotating beams, and the common infrastructure for the powering and cooling of the magnet [116].

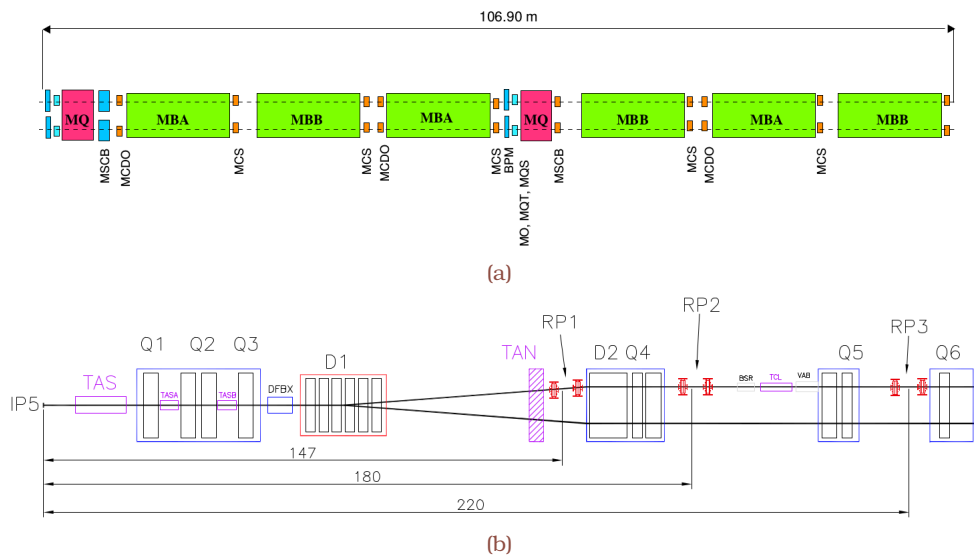
The LHC apparatus is designed as a two ring-like accelerator bearing an eightfold symmetry (Fig. 2.6a) with separate magnet fields and beam chambers, and with common straight sections intercepted by the experimental caverns, where the beams collide. The interaction regions (IRs) consist of separation dipole magnets (not highlighted in Fig. 2.6a) and main quadrupoles left and right of the interaction point (IP), i.e., the symmetry point of the IR. Three magnets on either side of the IP, the triplets (T), are situated only in the common region and are used for the final focusing. The triplets thus affect both beams, whereas the rest of the quadrupoles act separately on each beam. The remaining matching section (MS) and the dispersion suppressor (DS) consist of magnets with separate beam pipes for each ring. LHC consists of a total of 9 593 superconducting magnets out of which 1 232 are dipoles of about 15 m long equally shared over the eight arcs, and 392 main quadrupoles. Because of spatial restrictions imposed by the same tunnel that hosted LEP, it was thought beneficial LHC to use twin magnets, i.e., two sets of coils and beam channels sharing the mechanical structure and cryostat, as illustrated in Fig. 2.6b.

The IRs host the RF cavities (IR4), which accelerates the beams and keeps them bunched, the beam dump (IR6) system, used for extraction from the LHC, and cleaning devices (IR3 and 7), which are critical for the machine protection [117]. The two beams counter-rotate around the ring in two separate beam pipes; beam 1 is injected at IR2 and



circulates clockwise, while beam 2 is injected at IR8 and travels counter-clockwise. To provide collisions and data of high quality to the experiments with the desired rates the beam parameters have to be precisely controlled, e.g., Refs. [118, 119].

The arcs (Fig. 2.7a) are separated by IRs and extend over most of the length of each 3.3 km long sectors. They are built by periodically repeating 23 times a common lattice, the so-called “FODO” cell (107 m long), that is composed of a horizontally focusing quadrupole (MQ), three dipoles (MB), a vertically focusing quadrupole and another three dipoles [102]. The main bending magnets and quadrupoles are further equipped with sextupoles (MS), octupoles (MO), higher order (MC0) and orbit correctors (MCB) for adjustments of the various beam parameters around the ring. The magnets, equipped with their helium vessel and end covers, constitute the “cold masses,” which, in normal operation, contain superfluid helium at 1.9 K and 0.13 MPa, and are thermally insulated from the vacuum enclosure.



**Figure 2.7:** (a) Schematic layout of the LHC lattice structure. Each element installed in the LHC tunnel has its individual identification, constructed following a special convention as part of the LHC methodical accelerator design (MAD) sequence [120]. (b) The low values of the  $\beta$  function at the IPs of the experiments are generated with the help of a triplet quadrupole assembly (Q1 to Q3). The free space around the IPs that is reserved for the experiments is  $\pm 23$  m [105]. The two rings share the same vacuum chamber, the same low- $\beta$  triplet (superconducting) magnets, and the D1 separation dipole (warm) magnets. The remaining region is comprised of a second dipole (D2), a superconducting magnet operating at a cryogenic temperature due to lower radiation levels, and four matching quadrupole magnets (Q4 to Q7, the latter not highlighted).

The LHC has four IRs of very similar design but with different beam optics: two diametrically opposite insertions at IR1 (IR5) for “high-”luminosity collisions in ATLAS (CMS), and “medium-” and “low-”luminosity operation at IR8 (LHCb) and 2 (ALICE), respectively. The lower luminosity operation should be only considered as a relative term, e.g., the LHCb experiment nominally obtains about the same luminosity as the CDF and DØ detectors at Tevatron (Table 2.1). All four experimental IRs have similar designs and are equipped with dipoles and main quadrupoles symmetrically (asymmetrically) placed left and right of the IPs (IP8), as shown in Fig. 2.7b for IR5.

More specifically, the beam separation is steered by six dipole orbit corrector magnets per plane and beam. Two of them (D2 in Fig. 2.7b) are located further away from the IP and act on each beam separately. They are used for fine-tuning of the offsets. The other magnets are installed closer to the IP and affect both beams at the same time. By creating a local distortion in the beam orbit, which is called “closed orbit bump,” the D1 magnets are able to steer the beam separation—the cornerstone of the absolute luminosity calibration scans—and generate a crossing angle ( $\partial_C$ ) between the colliding bunches. The values of  $\partial_C$ , that lies in the vertical (horizontal) plane for IP1 and 2 (IP5 and 8) to minimize the impact of beam-beam effects (see Section 4.3.4), can be chosen large enough to minimize parasitic long-range encounters since the beams traverse a common vacuum chamber for about 120 m.

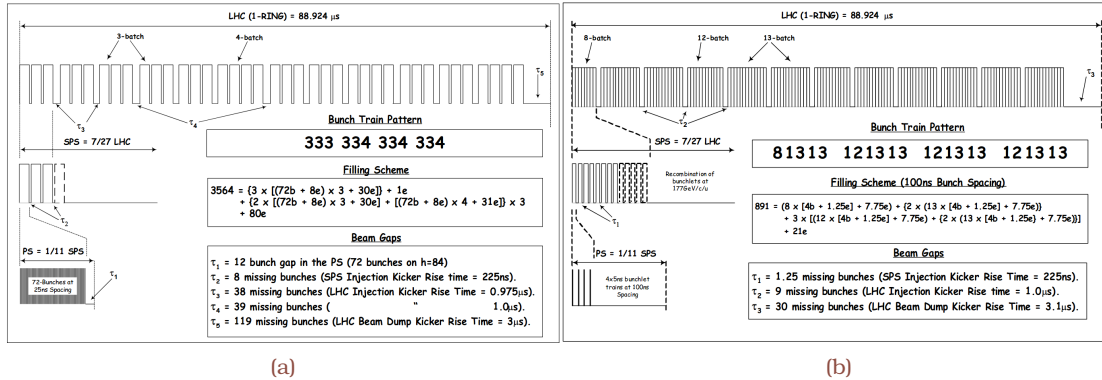
### LHC filling schemes and magnetic cycle

As discussed in Section 2.1.3, there is a chain of 35 640 RF buckets around LHC which could potentially be filled with bunches. In the nominal pp filling scheme, bunches are spaced by 25 ns, meaning 3 564 potential slots are available, each given a unique bunch crossing identifier (BCID). The injection from SPS has a bunch train structure, i.e., a specific number of equally spaced bunches. Between the trains, short gaps for the injection kicker magnets must be accounted for, and a  $3\ \mu\text{s}$  abort gap is kept free for a safe abort of the LHC beam. In general, not all bunches are “paired,” i.e., colliding in both beams in the same BCID, but different groups of BCIDs can be defined, e.g., “empty” BCIDs without a proton or ion bunch. By convention, the first BCID after the abort gap is numbered as 1.

A schematic view of the bunch distribution for the nominal pp filling scheme is given in Fig. 2.8a. The scheme involves 12 injections from SPS with each cycle consisting of either three or four batches from the pre-injectors. The order of SPS cycles is 333 334 334 334 meaning that the first SPS cycle injected into LHC contains  $3 \times 72$  batches from the pre-injectors, while the 6th, 9th and the last contain four batches. With each SPS cycle taking 21.6 s to complete, filling of each LHC ring takes around 4 min. In practice, LHC is flexible to operate with several different filling schemes meant for various purposes. In addition to the nominal 25 ns spacing for protons, the bunch-splitting in PS allows different bunch spacing: 50 ns (physics operation for Run 1 and part of Run 2 but with beams of reduced intensity and transverse emittance), 75 ns (initial period of physics operation) and greater than 75 ns (very first machine commissioning).

The nominal filling scheme for ions is optimized to cope with space charge problems in the injector chain. The baseline scheme has been based on 100 ns bunch spacing, although relatively few bunches are actually at this spacing. Figure 2.8b illustrates such a rather complex scheme; there are 891 possible bunch positions in LHC with a total of 592 filled bunches. In that case, bunches are fabricated four at a time in the pre-injectors. The SPS constructs trains by taking several set of bunches injected with a spacing of 225 ns. These are then accelerated and delivered to the LHC. The order of SPS cycles is: 8 13 13 12 13 13 12 13 13 12 13 13. Hence, the majority of SPS cycles involves 13 sets of

four bunches transferred from the pre-injectors to the LHC. As there are a large number of injections per SPS cycle, each cycle lasts about 54 s. With 12 SPS cycles it will take around 10 min to fill one LHC ring.

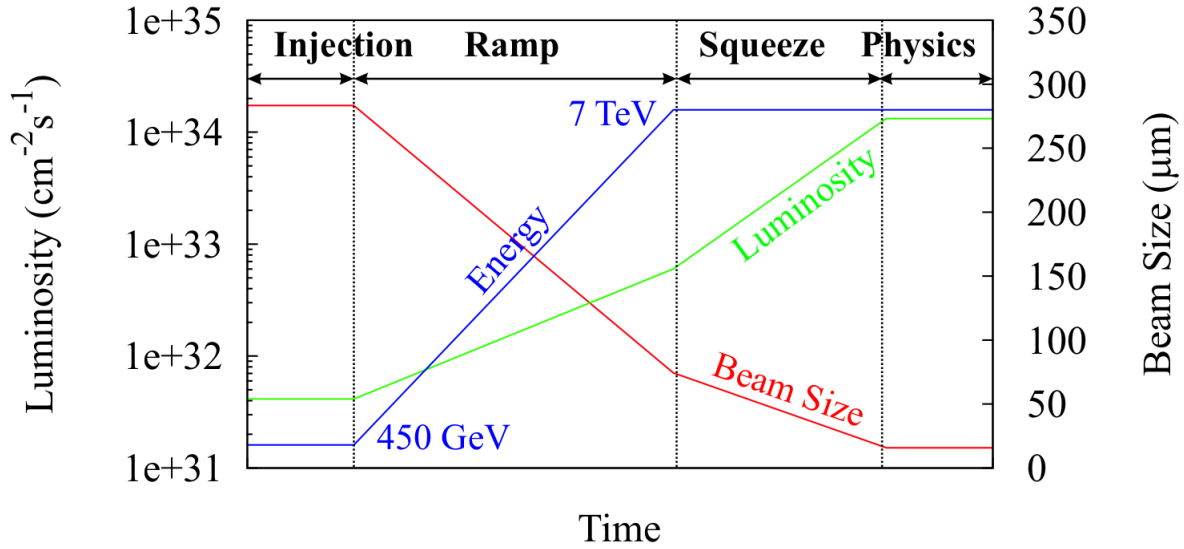


**Figure 2.8:** Schematic of the nominal bunch disposition around one LHC ring for the 25 ns proton (a) and 100 ns heavy ion (b) filling schemes; “b” indicates a position with beam and “e” indicates an empty bunch [121].

The evolution of the main beam parameters during a physics Fill is illustrated in Fig. 2.9. Note that the luminosity is just plotted for illustration since the beams are kept separated until the end of the “squeeze.” The beams are injected at large  $\beta^*$  values and then are “ramped” to high energy before squeezed at IP1 and 5 to a smaller  $\beta^*$ . During ramping the longitudinal momentum of the particles increases, while the transverse component is left unchanged, leading to a reduction of the transverse emittance. A smaller beam emittance, in turn, leads to a higher luminosity, while during the squeeze  $\beta^*$  gets decreased, increasing luminosity production further. Measurements indicated that the emittances are typically preserved during the LHC ramp and squeeze, whereas possible increasing and decreasing emittances during the same stages are mainly caused by a nonmonotonically changing  $\beta$  function. The knowledge of the latter is thus required in order to measure the beam emittance. Differences in time spent for each stage relative to Run 1 are fairly small [122], which is an indication that the long shutdown 1 did not affect the operational performance of LHC. The feasibility of combining the energy ramp with the betatron squeeze has been technically proved during a machine development study [123], while this method was first applied to the combined 2.51 TeV ramp with a squeeze to  $\beta^* = 4$  m in November 2015. Merging the two operations has resulted in a considerable gain (up to about 10 min) for each magnetic cycle, reducing the LHC turnaround, i.e., the time needed to establish physics (“stable beams”) conditions after a “beam dump” has occurred.

### 2.1.5 Proton-nucleus collisions at LHC

In the first stage (2010–2011 [115]) of its heavy ion program, LHC collided lead ( $^{208}_{82}\text{Pb}^+$ ) nuclei to study the properties of hot nuclear matter at extreme energy densities. The second stage of this program (2011–2013 [124]), following the pattern at the Relativistic



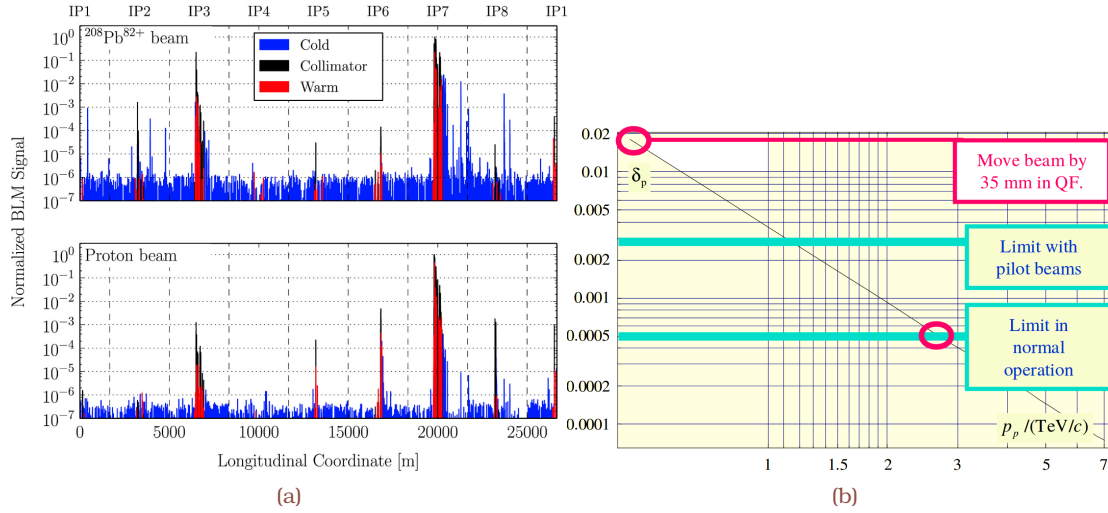
**Figure 2.9:** The LHC operation cycle is a defined protocol that ensures safe operation, and is characterized from the dipole magnetic field and dipole magnet current, that both receive their maximum values of 8.33 T and 11.7 kA, respectively, at physics conditions. The associated evolution of the nominal beam parameters (energy, luminosity, and beam size, the latter as defined in Eq. (2.13) assuming  $\epsilon_N = 3.75 \mu\text{m}$ ) is also shown for the different operation stages (here simplified), whose relative duration is indicated by the double-headed arrows. The luminosity represents the maximum achievable pp luminosity, in case the beams were to collide head-on at IP1 and 5.

Heavy Ion Collider (RHIC), was expected to be a crucial control experiment in which the nuclear matter is instead studied by colliding lead ions with protons; deuterons offer no particular advantage and are not readily available at LHC. At first sight, pPb collisions at LHC thus appear analogous to the successful dAu collisions at RHIC. For hybrid collisions to properly occur at IPs though, it is necessary to inject a proton beam with the same bunch pattern as a typical ion beam, meaning that the existing p and Pb injector chains have to operate in tandem efficiently; the two-in-one magnet design of LHC imposes different revolution frequencies for the two different mass-to-charge-ratio species; beam dynamics and the potential performance of LHC should be reevaluated.

The multistage collimation system [125] is designed to intercept protons at large amplitudes relative to primary collimators that scatter them into secondary and tertiary collimators, where they should eventually be absorbed. In heavy ion operation, although the stored beam energy is considerably smaller than for protons, the collimation is less efficient due to the yield of effectively off-momentum ion fragments in the primary collimators. These fragments with different magnetic rigidity can be further subject to fragmentation processes, and hence a large variety of different ion types can be produced through the interaction with the collimator material. Owing to the large cross sections for the involved fragmentation processes (few hundred Barns [126]), it is crucial to measure and understand ion loss patterns.

The cleaning performance is described by a local inefficiency (“loss map”) which is the ratio of the number of lost particles at any location of the ring in a given length over the

total number of lost particles. The required local cleaning inefficiency at 7 TeV is about  $10^{-5} \text{ m}^{-1}$  for the nominal intensity [127]. Figure 2.10a displays exemplary loss maps for  $^{208}\text{Pb}^{82+}$  [128] and p [129] beams at  $3.5Z \text{ TeV}$  having almost identical collimator settings and optics. Both distributions are dominated by losses in the betatron collimators at IR7, followed by the momentum collimators in IR3 and the dump protection devices in IR6. The different loss patterns in the IR2 region is due to the different optical configuration used in the two measurements.



**Figure 2.10:** (a) The representation of the local cleaning inefficiency as a function of the longitudinal position is referred to as a “loss map.” All Beam Loss Monitoring signals are cleaned from the background and normalized to the highest signal. The vertical dashed lines mark the LHC octants. (b) Proton momentum,  $p_p$ , shift required to equalize p and Pb revolution frequencies during the energy ramp. For  $p_p \lesssim 2.7 \text{ TeV}$ , the orbit displacement exceeds the normally accepted limits at LHC, imposing a lower limit on collision energy for pPb collisions. Injection and part of the ramp must be thus performed with unlocked RF systems and different revolution frequencies for the two beams [130].

Following the high integrated luminosity accumulated at  $\sqrt{s_{\text{NN}}} = 2.76 \text{ TeV}$  in the first two PbPb collision runs in 2010 [107] and 2011 [115], respectively, the LHC heavy ion physics community requested a first run with proton-nucleus collisions. This almost unprecedented mode of collider operation had not been foreseen in the baseline design of the LHC whose two-in-one magnet design (see Section 2.1.4) imposes equal rigidity, and hence unequal revolution frequencies per se for asymmetric colliding species.

Equal-rigidity configuration using deuteron-Gold collisions at RHIC has been considered in the initial injection setup, taking advantage of the two independent rings each comprised of single aperture magnets. However, in such condition, it has been later realized that unequal frequencies between the two beams modulate long-range beam-beam forces creating untunable beam losses during the acceleration ramp.

To equalize the RF frequencies at LHC, and hence allow the beams to encounter each other at the same position every turn, implies that the lead ion has to move to the inside of the ring to compensate for being slower, and equivalently, the proton beam has to move to the outside of the other ring to travel a larger distance to compensate for being faster

(Fig. 2.10b). The amount of orbit shift depends on the beam energy, and for instance, at injection energy, each beam should have been moved outside the reference orbit by about 70 mm, clearly out of reach given the dimensions of the beam pipe. This difficult exercise, called “RF frequency lock and cogging,” had therefore to be performed at flat top energies, and doubts were long extant as to whether LHC would ever deliver collisions with asymmetric species.

**Table 2.2:** Beam-related parameters in various LHC running periods. Since asymmetric collisions were not included in the LHC design, the physics case [131] was based on a luminosity  $\mathcal{L}_{\text{peak}} = 1.15 \times 10^{29} \text{ cm}^{-2}\text{s}^{-1}$  at a beam energy of 7 Z TeV. We can thus refer to these values as the “design” parameters, similarly to those in pp and PbPb collisions. Relevant quantities refer to IP1 and 5, while the values of  $\mathcal{L}_{\text{peak}}$  correspond to conditions of stable beams. The final performance values in the 2018 PbPb run are under evaluation.

Parameter		Design		PbPb		pPb		pp	
		PbPb	pp	2011	2015	2013	2016	2015	2018
$E$	(TeV)	7Z	7	3.5Z	6.37Z	4Z	6.5Z	6.5	6.5
$n_b$		592	2808	358	518	338	540	2244	2556
$N_{208\text{Pb}^{+},\text{p}}$	( $\times 10^8, \times 10^{11}$ )	0.7	1.15	1.2	1.9	1.4	2.1	1.1	1.4
$\varepsilon_N$	( $\mu\text{m}$ )	1.5	3.75	2.0	2.1	2.0	1.6	3.5	2.2
$\beta^*$	(m)	0.5	0.55	0.1	0.8	0.8	0.6	0.8	0.25
$E_S$	(MJ)	3.81	360	1.98	8.60	2.77	9.70	277	320
$\mathcal{L}_{\text{peak}}$	( $\times 10^{27}, \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$ )	1	1	0.5	3	116	850	0.5	2.0

After a successful pilot physics Fill in 2012 [132], LHC provided its four major experiments with approximately  $31 \text{ nb}^{-1}$  of pPb luminosity at an unprecedented nucleon-nucleon center-of-mass energy of 5.02 TeV in early 2013 [133], with several variations of the operating conditions. The gain of about a factor of 25 in collision energy relative to previous asymmetric systems has been one of the largest leaps in the history of particle accelerators. Together with a “reference” pp run at 2.76 TeV, they rendered the very last physics operation before the long shutdown 1 (Table 2.2).

For five of the LHC experiments, the second proton-nucleus collision run in 2016 offered the tremendous opportunity to answer a range of crucial physics questions [131] arising from unexpected discoveries, e.g., collective phenomena in small systems reminiscent to the creation of the QGP state of matter [54–56]. Unlike earlier runs, the requirements for the 2016 proton-nucleus run, and hence operating conditions, were dissimilar [134] among the different experiments, in terms of collision energy, luminosity, and pileup. Luminosity sharing deemed a critical issue to be solved since these requests appeared mutually incompatible within the available one month of operation. Nevertheless, a plan to satisfy most requirements was implemented successfully exploiting the different beam lifetimes at two nucleon-nucleon center-of-mass energies of 5.02 and 8.16 TeV, a variety of luminosity sharing and filling schemes, the later further complicated by the separate proton and heavy ion injection, and reversal of beam directions.

Despite the complex strategy for repeated recommissioning and operation of LHC the longest ever Fill (numbered 5510) was achieved with luminosity leveled for almost 38 h. The peak luminosity also surpassed the “design” value by a factor 7.8 (Table 2.2), and the amount of integrated luminosity substantially exceeded the requests of the majority

of the experiments. With the improvements in the performance of the injector complex and transmission efficiency into LHC [135], average Pb bunch intensities of ions and normalized transverse emittances of about  $2.1 \times 10^8$  and  $1.6 \mu\text{m}$ , respectively, were achieved at 6.5 Z TeV.

Table 2.3 summarizes all primary goals of the 2016 pPb run along with further useful data sets delivered parasitically, and Fig. 2.11 demonstrates the integrated luminosity performance achieved at all heavy ion periods at LHC so far.

$\sqrt{s_{\text{NN}}}$	Experiment	Planned luminosity	Delivered luminosity
5.02 TeV pPb	ALICE	$7 \times 10^8$ min.-bias events	$7.8 \times 10^8$ min.-bias events
	ATLAS, CMS	-	$> 0.4 \text{ nb}^{-1}$
	LHCb	-	SMOG [136] $p_2^4\text{He}^+$ , etc.
8.16 TeV Pbp	ATLAS, CMS	$100 \text{ nb}^{-1}$	$> 180 \text{ nb}^{-1}$
8.16 TeV pPb	ALICE, LHCb	$10 \text{ nb}^{-1}$	$14, 13 \text{ nb}^{-1}$
	LHCf	$(9 - 12 \text{ h}) \times 10^{28} \text{ cm}^{-2} \text{ s}^{-1}$	9.5 h
8.16 TeV Pbp	ALICE, LHCb	$10 \text{ nb}^{-1}$	$25, 19 \text{ nb}^{-1}$

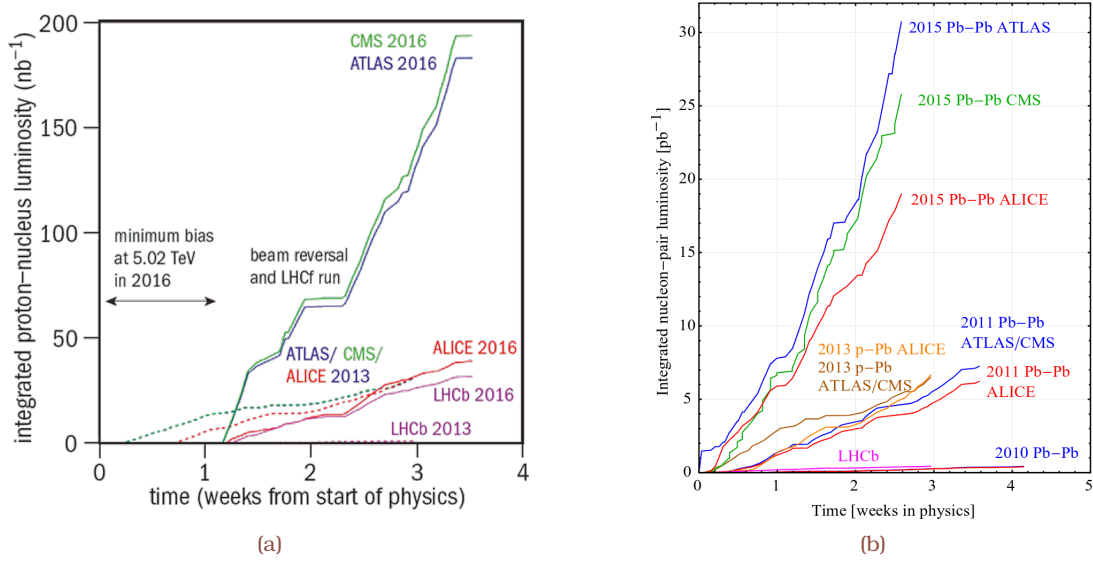
**Table 2.3:** Different activities including primary and additional goals were carried out during the various phases of the 2016 proton-nucleus run. More specifically, it involved two nucleon-nucleon center-of-mass energies, i.e., 5.02 and 8.16 TeV, and different particle species circulating in the two rings. Protons and  $^{208}\text{Pb}^+$  ions were injected in beam 1 and 2, respectively, for both energies, and beam species were switched once the required integrated luminosity was reached. All these changes in machine configuration took place in about one month, achieving a return to 5.02 TeV operation to fulfill additional requests. The peak luminosity exceeded the “design” value by a factor 7.8 [137], and the long-term integrated luminosity goal of  $100 \text{ nb}^{-1}$  [131] has been substantially surpassed in some LHC experiments.

## 2.2 The Compact Muon Solenoid detector

The CMS detector [72] is one of the two general-purpose detectors operating at LHC. It has an overall length of 22 m, a diameter of 15 m, and weighs 14 000 t. The CMS experiment makes use of a right-handed coordinate system, with the origin at the nominal interaction point, the  $x$  axis pointing to the center of the LHC ring, the  $y$  axis pointing up (perpendicular to the LHC plane), and the  $z$  axis along the counter-clockwise beam direction. The azimuthal angle  $\phi$  is measured in the  $x$ - $y$  plane, with  $\phi = 0$  along the positive  $x$  axis, and  $\phi = \pi/2$  along the positive  $y$  axis, and it is expressed in radians. The radial coordinate in this plane is denoted by  $r$ , while the polar angle  $\vartheta$  is defined in the  $r$ - $z$  plane with respect to the  $z$  axis, and the pseudorapidity is defined as  $\eta = -\ln(\tan(\vartheta/2))$ . The component of the momentum transverse to the  $z$  axis is denoted by  $p_{\text{T}}$ .

The CMS detector is well-suited to a global event description that is achieved correlating the basic elements from all detector layers. Most of the produced particle types in an event can be reconstructed and identified with an optimized combination of subdetector information, using [138]

- a large magnetic field (Section 2.2.1), to measure the momentum of charged particles



**Figure 2.11:** Integrated luminosity delivered to the four major experiments at LHC using (a) proton-nucleus and (b) nucleus-nucleus collisions during Run 1 and 2. Following up on a feasibility test and pilot physics Fill in October 2011 and September 2012, respectively, the first one month-long run took place in January 2013, meaning asymmetric proton-nucleus collisions remain a novel mode of operation at LHC. As in most years, the LHC apparatus was reconfigured for a month-long heavy ion run; 2016 was devoted to colliding beams of protons and  $^{208}\text{Pb}^+$  ions. Despite such novelty, the accelerator team succeeded in delivering enormous data sets to all major LHC experiments for the investigation of nuclear matter [137].

and to separate the calorimeter energy deposits of charged and neutral particles in jets;

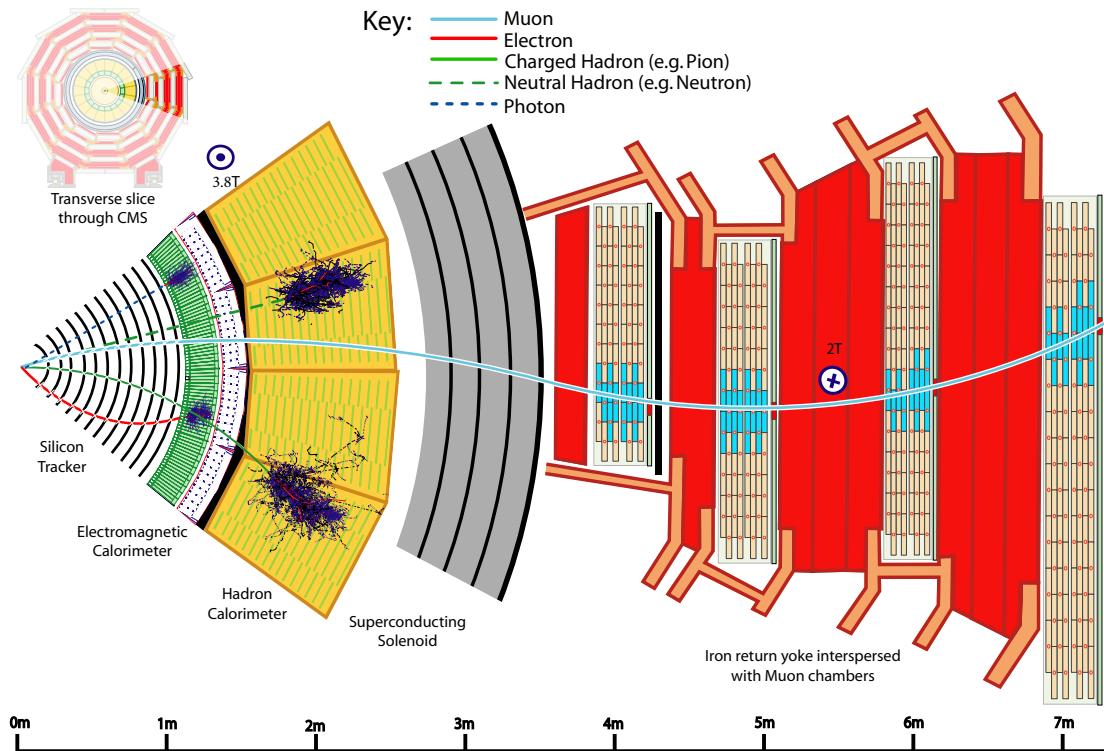
- a fine-grained tracker (Section 2.2.2), providing a pure and efficient charged-particle trajectory reconstruction in jets with  $p_T$  up to around 1 TeV, and therefore an excellent measurement of  $\sim 65\%$  of the jet energy;
- a highly-segmented electromagnetic calorimeter (ECAL) (Section 2.2.3), allowing energy deposits from particles in jets (charged hadrons, neutral hadrons, and photons) to be clearly separated from each other up to a jet  $p_T$  of the order of 1 TeV. The resulting efficient photon identification, combined with the high ECAL energy resolution, allows for an excellent measurement of another  $\sim 25\%$  of the jet energy;
- a hermetic hadron calorimeter (HCAL) (Section 2.2.4) with a coarse segmentation, but still sufficient to separate charged and neutral hadron energy deposits in jets up to  $p_T$  of 200–300 GeV, allowing the remaining 10% of the jet energy to be reconstructed, although with a modest resolution;
- an excellent muon tracking system (Section 2.2.5), delivering an efficient and pure muon identification, irrespective of the surrounding particles.

This simplified view is graphically summarized in Fig. 2.12, which displays a sketch of a transverse slice of the CMS detector.

The first “commissioning” operation in the 3.8 T magnetic field took place during October–November 2008, when a month-long data-taking exercise, henceforth known



as the “Cosmic Run at Four Tesla” (CRAFT) [139], was conducted, providing CMS Collaboration with invaluable experience in operating the experiment and understanding the performance of its subdetectors. At the start of each operating year, CMS also observes the muon halo from single circulating beams and receives several single shot “beam splash” events. In such an event, single circulating beams are steered onto closed collimators upstream of CMS, releasing muons that produce signals in most channels of the detector.



**Figure 2.12:** A sketch of the specific particle interactions in a transverse slice of the CMS detector, from the beam interaction region, on the left, to the muon detector, on the right. For this illustration, the muon and the charged pion are positively charged, and the electron is negatively charged [138].

Two subdetectors, the Centauro And STRange Object Research (CASTOR) and Zero Degree Calorimeter (ZDC), enhance the hermeticity of the CMS detector by extending the rapidity coverage in the forward region. The former is located 14.37 m from IP5—installed on the collar table between the HF shielding and the rotating shielding [140]—and extends the forward rapidity coverage to the region  $-6.6 < |\eta| < -5.2$  (no segmentation), while the later is installed 140 m away from IP5 in both the forward and backward directions with a full acceptance to measure neutral energy flow in the  $|\eta| > 8.3$  region (multifold segmentation). Both calorimeters are made of quartz fibers and plates embedded in tungsten absorbers, providing a fast collection of Cerenkov light; each is further divided into an electromagnetic and hadronic section of 20.12 (19) radiation and 9.5 (5.6) interaction lengths, respectively. The significance of the forward physics program is essential, e.g., it offers constraints on the modeling of the underlying event in both pp and nuclear collisions, revealing the proton and nucleus structure, and the parton evolution. A more

detailed summary of the forward physics program can be found in Ref. [141].

The characteristics of the magnet and the rest of CMS subdetectors are described in the following.

### 2.2.1 The superconducting solenoid magnet

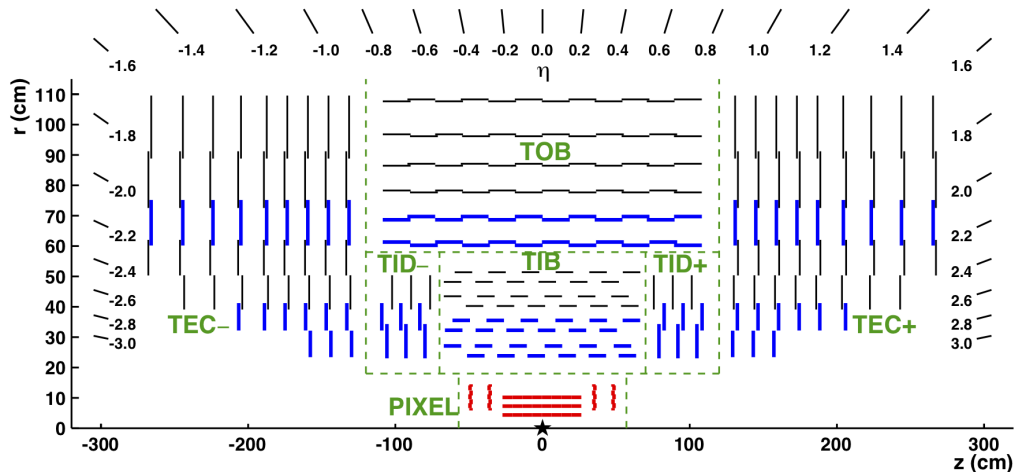
The central feature of the CMS design is a large superconducting solenoid magnet [142]. It delivers an axial and uniform magnetic field of 3.8T over a length of 12.5 m and a free bore radius of 3.15 m. This radius is large enough to accommodate the tracker and both the ECAL and HCAL, thereby minimizing the amount of material in front of the calorimeters. This feature is an advantage for a global event reconstruction, as it eliminates the energy losses before the calorimeters caused by particles showering in the coil material and facilitates the link between tracks and calorimeter clusters. At normal incidence, the bending power of 4.9 Tm to the inner surface of the calorimeter system provides strong separation (few cm) between charged- and neutral-particle energy deposits, i.e., large enough distance to resolve energy deposits of the former from that of the latter emitted in the same direction.

### 2.2.2 The inner silicon pixel and the larger silicon strip tracker

The full-silicon CMS tracking system [143, 144] is a cylinder-shaped detector consisting of two main detectors: the smaller inner pixel detector and the larger silicon strip tracker. The original (“phase-0”) pixel detector—used in the current thesis—had three barrel pixel (BPiX) layers and two endcap disks (FPiX) per side, covering the region from 4 to 15 cm in radius, and spanning 98 cm along the LHC beam axis. The pixel modules, shown by the red lines in Fig. 2.13, provide three-dimensional hits. The silicon strip tracker had ten barrel layers and twelve endcap disks per side, covering the region from 25 to 110 cm in radius, and spanning 560 cm along the LHC beam axis. Strip tracker modules that provide two-dimensional hits are illustrated with thin, black lines in Fig. 2.13, while pairs of modules mounted back-to-back with a slight tilt are shown by thick, blue lines. Within a given layer, each module is shifted slightly in  $r$  or  $z$  with respect to its adjacent modules, which allows them to overlap, thereby avoiding gaps in the acceptance. The latter extends up to a pseudorapidity of  $|\eta| = 2.5$ .

The silicon strip tracker has four subsystems. The innermost four barrel layers comprise the tracker inner barrel (TiB) detector, and the outer six barrel layers form the tracker outer barrel (ToB) detector. The three endcap disks to either side of the TiB detector form the tracker inner disks (TiD<sup>-</sup> and TiD<sup>+</sup>), and the nine endcap disks at each end constitute the tracker endcap (TEC<sup>-</sup> and TEC<sup>+</sup>). The 16 588 silicon sensor modules were finely segmented into 66 million  $100 \times 150 \mu\text{m}^2$  pixels (1 440 modules) and 9.6 million 80-to-180  $\mu\text{m}$ -wide strips (15 148 modules). This fine granularity offers separation of closely-spaced particle trajectories in jets. With about  $200 \text{ m}^2$  of active silicon area (Table 2.4) the CMS tracker is the largest silicon tracker ever built.

Tracker layers and the vital services (cables, support, cooling) represent though a substantial amount of material in front of the calorimeters, up to 0.5 interaction lengths or



**Figure 2.13:** Schematic view of the CMS tracker (phase-0) detector with labels identifying the silicon pixel detector (three barrel layers and two endcap disks per side) and the silicon strip tracker (ten barrel layers and twelve endcap disks per side), in total covering the region from 4 to 110 cm in radius, and spanning 560 cm along the LHC beam axis. In this view, the tracker is symmetric about the horizontal line at  $r = 0$ , so only the top half is shown [145]. The center of the tracker, corresponding to the approximate position of the collision point, is indicated by the star.

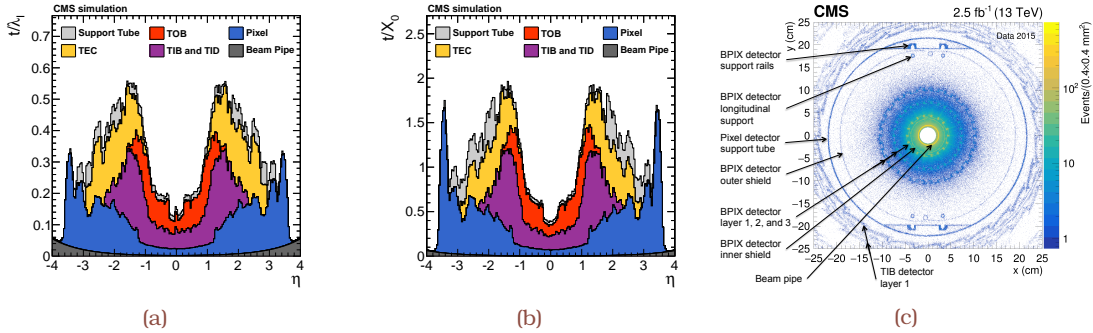
Tracker subsystem (modules)	Layers	Location (cm)	Pitch	Intrinsic $r\phi$ resolution ( $\mu\text{m}$ )
BPIX (768)	3 cylindrical	$4.4 < r < 10.2$	$100 \times 150 \mu\text{m}^2$	10
TIB (2 724)	4 cylindrical	$20 < r < 55$	$80\text{--}120 \mu\text{m}$	13–38
TOB (5 200)	6 cylindrical	$55 < r < 116$	$183\text{--}122 \mu\text{m}$	
FPIX (672)	2 disks	$34.5 <  z  < 46.5$	$100 \times 150 \mu\text{m}^2$	20–40
TID (816)	3 disks	$58 <  z  < 124$	$100\text{--}141 \mu\text{m}$	18–47
TEC (6 400)	9 disks	$124 <  z  < 282$	$97\text{--}184 \mu\text{m}$	

**Table 2.4:** Summary of the principal characteristics of the various tracker subsystems [145]. The number of disks corresponds to that in a single endcap. The location specifies the region in  $r$  ( $|z|$ ) occupied by each barrel (endcap) subsystem. The modules of the pixel detector use silicon of  $285 \mu\text{m}$  thickness, and achieve resolutions that are roughly the same in  $r\phi$  as in  $z$ , because of the chosen pixel cell size. The modules in the TIB, TID and inner four TEC rings use silicon that is  $320 \mu\text{m}$  thick, while those in the TOB and the outer three TEC rings use silicon of  $500 \mu\text{m}$ .

1.8 radiation lengths, as estimated from simulation and displayed in Fig. 2.14a and 2.14b, respectively. At  $|\eta| \approx 1.5$ , the probability for a photon to convert or for an electron to emit a bremsstrahlung photon by interacting with this material is about 85% due to the presence of the detector service cables. Similarly, a hadron (charged pion) has a 20% probability to experience a nuclear interaction before reaching the ECAL surface [145]. A large number of emerging secondary particles turned out to be a major source of complication for a global event reconstruction algorithm, and it required harnessing the full granularity and redundancy of the silicon tracker measurements.

Despite being undesirable events that degrade the quality of the reconstruction of charged and neutral hadrons, nuclear interactions can be alternatively used to produce a high-precision map of the material inside the tracker, as shown in Fig 2.14c for the

$x$ - $y$  plane in the barrel region ( $|z| < 25$  cm) using a data set of pp collisions at  $\sqrt{s} = 13$  TeV [146]. The positions of the secondary vertices can determine the locations of passive material with a precision of the order of  $100 \mu\text{m}$ , verifying the simulation of the tracker material budget with an accuracy better than 10%. The signatures of the beam pipe, the BPIX detector with its support, and the first layer of the TIB detector can be clearly observed above the background of misreconstructed vertices in Fig 2.14c.



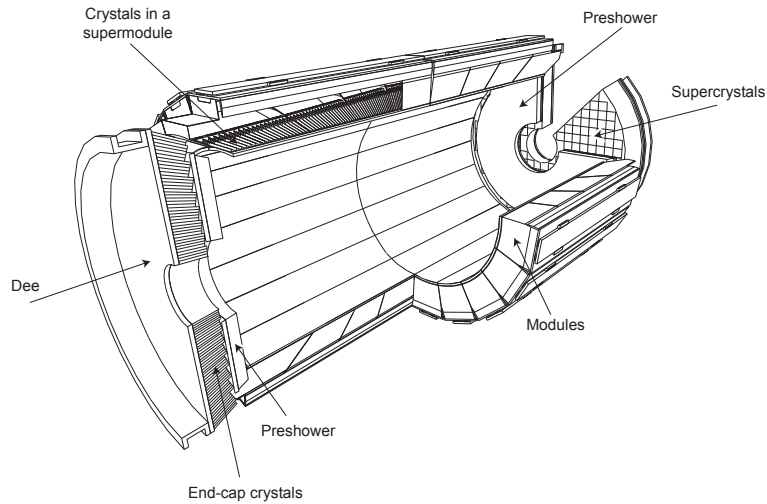
**Figure 2.14:** Total thickness  $t$  of the inner tracker material expressed in units of interaction (a) and radiation (b) lengths,  $\lambda_I$  and  $X_0$ , respectively, as a function of the pseudorapidity  $\eta$  [145]. (c) “Hadrography” is based on samples of secondary hadronic interactions and produces a high-precision map of the material within the tracking volume. The density of vertices is indicated by the color scale [146].

The tracker measures the  $p_T$  of charged hadrons at normal incidence with a resolution of 1% for  $p_T < 20$  GeV. The relative resolution then degrades with increasing  $p_T$  to reach the calorimeter energy resolution for track momenta of several hundred GeV. Because the fragmentation of high- $p_T$  partons typically produces many charged hadrons at a lower  $p_T$ , the tracker is expected to contribute significantly to the measurement of the momentum of jets with a  $p_T$  up to a few TeV [145].

### 2.2.3 The electromagnetic calorimeter

The ECAL [147, 148] is a hermetic homogeneous calorimeter made of 75 848 lead tungstate ( $\text{PbWO}_4$ ) crystals. The barrel (EB) covers  $|\eta| < 1.479$  and the two endcap (EE) disks  $1.479 < |\eta| < 3.0$ . The EB (EE) crystal length of 23 (22) cm corresponds to 25.8 (24.7) radiation lengths, sufficient to contain more than 98% of the energy of electrons and photons up to 1 TeV, with the electron and photon separation being possible up to  $|\eta| = 2.5$ , the limit of the region covered by the tracker. The crystal material also amounts to about one interaction length, causing about two-thirds of the hadrons to start showering in ECAL before entering HCAL. To measure and correct for response changes during LHC operation ECAL is equipped with a light monitoring system.

The crystal transverse size matches the small Molière radius of  $\text{PbWO}_4$ , 2.2 cm. This fine transverse granularity makes it possible to fully resolve hadron and photon energy deposits as close as 5 cm from one another, for the benefit of exclusive particle identification in jets. More specifically, the front face of the EB crystals has an area of  $2.2 \times 2.2 \text{ cm}^2$ , equivalent to  $0.0174 \times 0.0174$  in the  $\eta$ - $\phi$  plane. For EE, the crystals are arranged instead



**Figure 2.15:** The CMS ECAL is a homogeneous and hermetic calorimeter [149]. The preshower detector, based on lead absorbers that are equipped with silicon strip sensors, is placed in front of the endcap crystals, to enhance photon identification capabilities.

in a rectangular  $(x, y)$  grid, with a front-face area of  $2.9 \times 2.9 \text{ cm}^2$ . The intrinsic energy resolution of EB was measured with an ECAL supermodule directly exposed to an electron beam, without any attempt to reproduce the inert material of the tracker in front of the ECAL [150]. The energy resolution,  $\sigma$ , is parameterized as a function of the electron energy,  $E$ , as

$$\frac{\sigma}{E} = \frac{2.8\%}{\sqrt{E/\text{GeV}}} \oplus \frac{12\%}{E/\text{GeV}} \oplus 0.3\%, \quad 2.21$$

where the three contributions correspond to the stochastic, e.g., the shower containment, the noise, e.g., deposits from multiple interactions per bunch crossing, and constant, e.g., non-uniformity of the longitudinal light collection, terms, respectively. Because of the very small stochastic term inherent to homogeneous calorimeters, the photon energy resolution is excellent in the 1–50 GeV range typical of photons in jets. The constant term dominates the energy resolution for high-energy electron and photon showers.

The ECAL electronics noise  $\sigma_{\text{noise}}^{\text{ECAL}}$  is measured to be about 40 (150) MeV per crystal in EB (EE). Another important source of spurious signals arises from particles directly ionizing the avalanche photodiodes, aimed at collecting the crystal scintillation light [151]. This effect gives rise to single-crystal spikes with a relative amplitude significantly larger than the scintillation light. Such spikes would be misidentified by a global event description algorithm as photons with an energy up to 1 TeV. Since these spikes mostly affect a single crystal and more rarely two neighboring crystals, they are rejected by requiring the energy deposits to be compatible with arising from a particle shower. This is based on a combination of looser and tighter thresholds on  $E_4/E_1$  and  $E_6/E_2$  ratios, where  $E_1$  ( $E_2$ ) and  $E_4$  ( $E_6$ ) are the energies collected in the considered crystal (crystal pair) and in the four (six) adjacent crystals, respectively. The timing of the energy deposits in excess of 1 GeV is also required to be compatible with the beam crossing time to better than  $\pm 2 \text{ ns}$ .

A much finer-grained detector, known as preshower (E5) and whose fiducial area is

approximately  $1.65 < |\eta| < 2.6$ , is installed in front of each EE disk. It consists of two layers, each comprising a lead radiator followed by a plane of silicon strip sensors. The two lead radiators represent approximately two and one radiation lengths, respectively. The two planes of silicon sensors have orthogonal strips with a pitch of 1.9 mm. When either a photon or an electron passes through the lead, it initiates an electromagnetic shower. The granularity of the detector and the small radius of the initiating shower provide an accurate measurement of the shower position.

Originally, the aim of the superior granularity of ES was (i) to resolve the photons from  $\pi^0$  decays; and (ii) to indicate the presence of a photon or an electron in ECAL by requiring an associated signal in ES. Parasitic signals, however, are generated by the large number of neutral pions produced by hadron interactions in the tracker material, followed by photon conversions and electron bremsstrahlung. These signals affect substantially the ES identification and separation capabilities. In a global event description algorithm, these capabilities can therefore not be fully exploited, and the energy deposited in ES merely is added to that of the closest associated ECAL cluster, if any, and discarded otherwise.

The electron momentum is estimated by combining the energy measurement in ECAL with the momentum measurement in the tracker. The momentum resolution for electrons with  $p_T \approx 45$  GeV from  $Z \rightarrow e^+e^-$  decays ranges from 1.7 to 4.5%. It is generally better in the barrel region than in the endcaps, and also depends on the bremsstrahlung energy emitted by the electron as it traverses the material in front of the ECAL [152].

#### 2.2.4 The hadron calorimeter

The HCAL [153] is a hermetic sampling calorimeter consisting of several layers of brass absorber and plastic scintillator tiles. It surrounds the ECAL, with a barrel ( $|\eta| < 1.3$ ) and two endcap disks ( $1.3 < |\eta| < 3.0$ ). In the barrel, the HCAL absorber thickness amounts to almost six interaction lengths at normal incidence, and increases to about ten interaction lengths at larger pseudorapidities. It is complemented by a tail catcher (H0) that is installed outside the solenoid coil. The H0 material (1.4 interaction lengths at normal incidence) is used as an additional absorber. At small pseudorapidities ( $|\eta| < 0.25$ ), this thickness is enhanced to a total of three interaction lengths by a layer of steel. The total depth of the calorimeter system (including ECAL) is thus extended to a minimum of twelve interaction lengths in the barrel, while the thickness amounts to about ten interaction lengths in the endcaps.

The HCAL is read out in individual towers with a cross section  $\Delta\eta \times \Delta\varphi = 0.087 \times 0.087$  for  $|\eta| < 1.6$  and  $0.17 \times 0.17$  at larger pseudorapidities. The combined (ECAL+HCAL) calorimeter energy resolution was measured in a pion test beam [154] to be

$$\frac{\sigma}{E} = \frac{110\%}{\sqrt{E}} \oplus 9\%, \quad 2.22$$

where  $E$  is expressed in GeV.

The typical HCAL electronics noise  $\sigma_{\text{noise}}^{\text{HCAL}}$  is measured to be  $\approx 200$  MeV per tower. Additionally, rare occurrences of high-amplitude, coherent noise were observed in the HCAL

barrel [155]. Since this coherent HCAL noise would be misinterpreted as high-energy neutral hadrons by a global event description algorithm, the affected events are identified by their characteristic topological features and rejected at the analysis level.

The HCAL is complemented by hadron forward (HF) calorimeters situated at  $\pm 11.2\text{ m}$  from the interaction point that extend the angular coverage on both sides up to  $|\eta| = 5.2$ . The HF consists of a steel absorber composed of grooved plates. Radiation-hard quartz fibers are inserted in the grooves along the beam direction and are read out by photomultipliers. Each HF calorimeter consists of 432 readout towers with a cross section  $\Delta\eta \times \Delta\phi = 0.175 \times 0.175$  over most of the pseudorapidity range, containing long and short quartz fibers. The long fibers run the entire depth of the HF calorimeter (about 165 cm, or approximately ten interaction length), while the short fibers start at a depth of 22 cm from the front of the detector.

In each calorimeter tower, the signals from the short and long fibers are used to estimate the electromagnetic and hadronic components of the shower; photons deposit a significant fraction of their energy in the long-fiber calorimeter segment, whereas hadrons produce on average nearly equal signals in both calorimeter segments. If  $L$  ( $S$ ) denotes the energy measured in the long (short) fibers, the energy of the electromagnetic component, concentrated in the first part of the absorber, can be approximated by  $L - S$ , and the energy of the hadronic component is the complement, i.e.,  $2S$ .

Spurious signals in HF, caused for example by high-energy beam halo muons directly hitting the photomultiplier windows, are reduced by rejecting (i) high-energy  $S$  deposits not backed up by a  $L$  deposit in the same tower; (ii) out-of-time  $S$  or  $L$  deposits of more than 30 GeV, (iii)  $L$  deposits larger than 120 GeV with  $S < 0.01L$  in the same tower; (iv) isolated  $L$  deposits larger than 80 GeV, with small  $L$  and  $S$  deposits in the four neighbouring towers.

### 2.2.5 The muon detectors

Outside the solenoid coil, the magnetic flux is returned through a yoke consisting of three layers of steel interleaved with four muon gaseous detector planes (Fig. 2.16). Drift tube (DT) chambers and cathode strip chambers (CSC) detect muons in the regions  $|\eta| < 1.2$  and  $0.9 < |\eta| < 2.4$ , respectively, and are complemented by a system of resistive plate chambers (RPC) covering the range  $|\eta| < 1.6$ . The reconstruction involves a global trajectory fit across the muon detectors and the inner tracker. The calorimeters and the solenoid coil represent a large amount of material before the muon detectors, and hence induce multiple scattering leaving the tracker to dominate the momentum measurement up to a  $p_T$  of about 200 GeV.

The DTs are segmented into drift cells; the position of the muon is determined by measuring the drift time to an anode wire of a cell with a shaped electric field. The CSCs operate as standard multi-wire proportional counters but add a finely segmented cathode strip readout, which yields an accurate measurement of the position of the bending plane ( $r$ - $\phi$ ) coordinate at which the muon crosses the gas volume. The RPCs are double-gap chambers operated in avalanche mode, and are primarily designed to provide timing

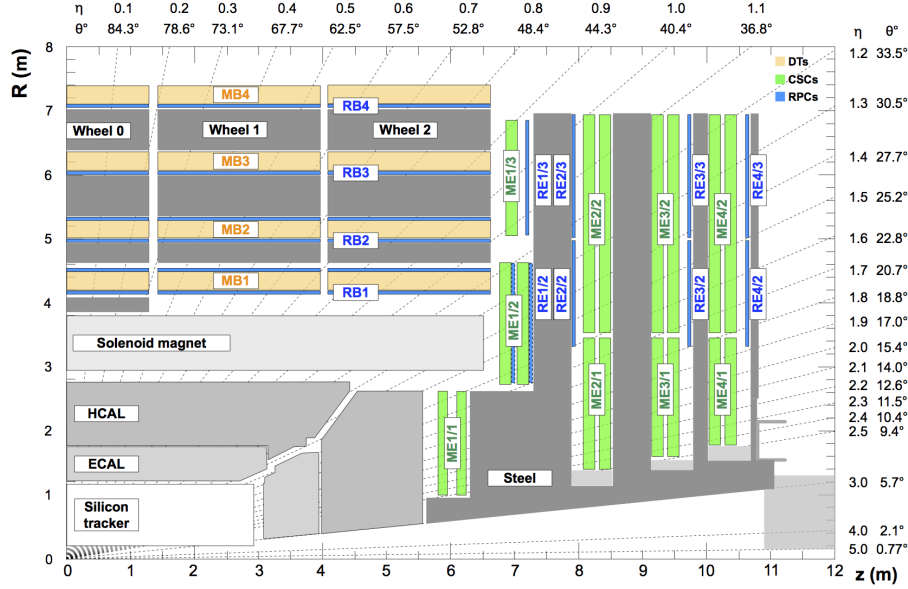


Figure 2.16: The locations of the various muon stations and the steel flux-return disks (dark areas) are shown [156]. Here a quadrant of the CMS detector in the  $r$ - $z$  plane is shown, with the axis parallel to the beam ( $z$ ) running horizontally and the radius ( $r$ ) increasing upward. The interaction point is at the lower left corner.

information for the muon trigger. Three regions can be distinguished, naturally defined by the cylindrical geometry of CMS, i.e., the barrel ( $|\eta| < 0.9$ ), overlap ( $0.9 < |\eta| < 1.2$ ), and endcap ( $1.2 < |\eta| < 2.4$ ) regions.

Table 2.5: Properties and parameters of the CMS muon subsystems during the 2016 data taking period [156].

Muon subsystem	DT	CSC	RPC
$ \eta $ coverage	0.0–1.2	0.9–2.4	0.0–1.9
Number of stations	4	4	4
Number of chambers	250	540	Barrel: 480 Endcap: 576
Number of layers/chamber	$r$ - $\phi$ : 8; $z$ : 4	6	2 in RB1 and RB2 1 elsewhere
Number of readout channels	172 000	Strips: 266 112 Anode channels: 210 816	Barrel: 68 136 Endcap: 55 296
Percentage of active channels	98.4%	99.0%	98.3%

The chambers are arranged to maximize the coverage and to provide overlap where possible. In the barrel, a station is a ring of chambers assembled between two layers of the steel flux-return yoke at approximately the same value  $R$ . There are four DT and four RPC stations in the barrel, labeled MB1–MB4 and RB1–RB4, respectively. Both DT and RPC barrel stations are arranged in five “wheels” along  $z$ . In the endcaps, a station is a ring of chambers assembled between two disks of the steel flux-return yoke at approximately the same value of  $z$ . There are four CSC and four RPC stations in each endcap, labeled ME1–ME4 and RE1–RE4, respectively. Between Run 1 and 2, additional chambers were added in ME4 and RE4 to increase redundancy and improve efficiency. A detailed description



of these chambers, including gas composition and operating voltage, can be found in Refs. [156, 157].

Matching muons to tracks measured in the silicon tracker results in a relative transverse momentum resolution, for muons with  $p_T$  up to 100 GeV, of 1% in the barrel and 3% in the endcaps. The  $p_T$  resolution in the barrel is better than 7% for muons with  $p_T$  up to 1 TeV [156].

### 2.2.6 Luminosity detectors at CMS

A system consisting of five subdetectors (“luminometers”) to monitor and measure the luminosity delivered by LHC is currently in use at the CMS experiment <sup>2</sup>. Based on rate measurements for a variety of observables (see Section 4.2.1), this system includes five luminometers: a) the silicon pixel and strip tracker, b) the HF calorimeter, c) the DT muon detector, d) the Fast Beam Conditions Monitor (BCM1F), and e) the most recently installed and commissioned Pixel Luminosity Telescope (PLT).

#### The tracker-based methods

A pixel-cluster counting (PCC) method uses the rate of pixel clusters in CMS pixel detector to provide a luminosity measurement. It supplied the primary offline luminosity measurement for CMS in 2015–2016, since the large area of the pixel detector and the relatively low occupancy provides a measurement with good statistical precision, and the stability of the measurement over time is typically good. Because the CMS trigger bandwidth (see Section 3.1) available for collecting the data used for this measurement is limited, the statistical precision for a single 23 s period is not as high as for the rest of the luminometers, but integrated over longer time periods this is not anymore an issue for the PCC luminosity.

Two corrections are typically applied to the PCC measurement to account for two discrete effects: the first type accounts for the signal from a hit spilling over into the next bunch crossing after a colliding bunch, while the second type accounts for an exponentially decaying “afterglow” for several bunches following a colliding bunch caused by activation of the surrounding detector material. These effects are measured using data from empty bunches, which should nominally have zero luminosity, and corrections are derived and applied to the raw luminosity. Since these corrections vary over the course of a run they are measured as a function of time and applied in a time-dependent manner.

A second method relies on the primary vertex counting, imposing additional requirements to eliminate the background and retain good reconstruction efficiency. This method is simple and robust, but becomes less linear at high values of instantaneous luminosity. There are two competing effects. On the one hand, primary vertices from two collisions occurring close to one another in space are merged, leading to an undercounting of vertices at progressively high instantaneous luminosity. On the other hand, a large number of tracks can produce fake vertices, leading to overcounting. The precision with which these

<sup>2</sup>Although originally conceived for radiation protection the Radiation Monitoring System for the Environment and Safety (RAMSES) [158] can be used for luminosity measurements.

effects are currently understood falls short of the level needed for precision luminosity studies. Vertex counting nonetheless serves as a useful cross check, especially for the dedicated beam-separation scans, where the instantaneous luminosity is typically low.

### **The HF measurement**

The HF luminosity measurement uses a dedicated readout system installed in the HF calorimeter. The HF provided the primary online luminosity measurement for CMS during Run 1 of the LHC and has continued to provide excellent performance throughout Run 2. The latest algorithm applied in 2017, for the first time, uses the sum of the transverse energy  $\sum E_T$  (HFET), which provides better performance at higher instantaneous luminosity than the occupancy-based algorithm (HFOC), i.e., based on the average fraction of empty towers.

Similarly to the PCC luminosity, afterglow corrections are applied: for HFET amount to approximately 4% in the bunch immediately following a colliding bunch and 0.5% in the next following bunch, with the corrections for subsequent bunches less than about 0.1%.

### **The PLT detector**

The PLT [159, 160] is a dedicated system for measuring luminosity using silicon pixel sensors, installed at the beginning of 2015. There are a total of 48 sensors arranged into 16 “telescopes,” eight at either end of CMS outside the pixel endcap, where each telescope contains three sensor planes arranged nearly parallel to the beam pipe. The sensors measure  $8 \times 8$  mm, divided into 80 rows and 52 columns, although only the central region of the sensors is used to reduce the contribution from background.

Over the course of the time, accumulated radiation damage in the sensors may result in a higher-than-expected loss of efficiency in PLT. The effects of this damage are compensated for by increasing the high voltage applied to the sensors; however, there might still be several periods where the PLT exhibits low efficiency. Corrections for these efficiency losses are applied offline.

### **The BCM1F detector**

BCM1F [161, 162] measures both luminosity and machine-induced background (MIB). It consists of a total of 24 sensors mounted on the same carriage as the PLT. The sensors in BCM1F consist of a total of 10 silicon sensors, 10 polycrystalline diamond (pCVD) sensors, and 4 single-crystal diamond (sCVD) sensors. The pCVD and sCVD sensors use split-pad metallization, with each sensor having two readout channels, to keep the overall occupancy low given the expected conditions in Run 2. The BCM1F readout features a fast readout with 6.25 ns time resolution; the precise time measurement, in conjunction with the position of BCM1F 1.8 m from the center of CMS, allows hits from collision products to be separated from hits from MIB, because the incoming background is separated in time from the outgoing collision products.

**The DT measuremet**

The DT luminosity measurement uses the rate of muon track stubs in the muon barrel track finder. While the DT measurement is available online, the DT algorithm does not provide bunch-by-bunch measurements and is thus applicable only for the total luminosity measurement. In Run 2, the DT measurement has generally been stable and linear, as long as the track finder itself is not changed, so it can provide a complementary offline reference measurement.



## Contents

---

<b>3.1 The online event filter</b> . . . . .	<b>67</b>
3.1.1 L1 trigger overview . . . . .	67
3.1.2 HLT overview . . . . .	69
3.1.3 Electron and photon triggers . . . . .	69
3.1.4 Muon triggers . . . . .	71
3.1.5 Beam position timing trigger . . . . .	72
<b>3.2 Hit, track, and primary vertex reconstruction</b> . . . . .	<b>73</b>
3.2.1 Hit reconstruction performance . . . . .	74
3.2.2 Track reconstruction performance . . . . .	75
3.2.3 Primary vertex reconstruction performance . . . . .	77
3.2.4 Reconstruction of the LHC luminous region . . . . .	79
<b>3.3 Physics object reconstruction</b> . . . . .	<b>81</b>
3.3.1 Pileup mitigation and treatment of the underlying event . . . . .	82
3.3.2 Muon reconstruction and identification . . . . .	84
3.3.3 Electron reconstruction and identification . . . . .	87
3.3.4 Jet reconstruction and noise-jet rejection . . . . .	93
3.3.5 Identification of heavy-flavor jets . . . . .	99
3.3.6 Reconstruction and calibration of $\vec{p}_T^{\text{miss}}$ . . . . .	101

---

To select events of potential physics interest, the CMS filter (“trigger”) system is composed of the first level (L1), a custom hardware trigger, and the second level, a high-level trigger (HLT) that consists of custom software routines. Information on the bunch structure and the timing of the incoming beams is obtained from the beam pickup for timing devices that are found on either side of the interaction point and used for suppressing the noise in triggers with high background.

The CMS track and primary vertex reconstruction software aims to reconstruct the trajectories of charged particles while traversing the CMS silicon tracker. Despite the high track density, the pixel tracker is well-suited to this purpose, with its high granularity giving excellent position and momentum resolution in three dimensions and low channel occupancy. Tracks are further used to reconstruct the primary interaction vertices in each event. For vertices with many tracks, characteristic of interesting events, the achieved vertex position resolution is excellent in each of the three spatial dimensions. For the first time at a hadron collider, a fully-fledged particle-flow (PF) reconstruction algorithm has been developed, tuned to the CMS detector, and consistently used in physics analyses. For each collision, final-state particles are identified and reconstructed by the PF algorithm that provides a global event description and leads to unprecedented CMS performance for electron and muon identification, jet reconstruction, and missing transverse momentum

( $p_T^{\text{miss}}$ ) determination among others. This approach also allows particles from multiple interactions (pileup) to be identified and enables efficient pileup mitigation methods.

The original contribution from the present thesis to the material related to the following Chapter and listed in [Scientific output](#) and [Internal notes](#) is:

- The measurement of the lepton selection (trigger, reconstruction, and identification) efficiency using a “tag-and-probe” method in same-flavor dilepton pp and pPb events enriched in Z boson candidates.
- The measurement of the lepton isolation performance in pPb collisions due to the presence of the underlying event.
- The optimization of the lepton isolation performance in pp collisions (at  $\sqrt{s} = 13$  TeV) due to the presence of the QCD multijet background.
- The measurement of the resolution in the reconstructed primary-vertex positions in event topologies typical for the dedicated beam-separation scans.
- The offline measurement and monitoring of the CMS luminous region during the dedicated beam-separation scans.
- The measurement of the expected efficiency for the correct identification of a b jet as a function of the probability to be misidentified as light-flavor or gluon jet in pp collisions.
- The treatment of the bias in the  $p_T^{\text{miss}}$  measurement by correcting the  $p_T$  of the jets to the particle level jet  $p_T$  using the jet energy correction procedure in pp collisions.
- The uncertainty estimate in  $p_T^{\text{miss}}$  factorizing it into its components (electrons, muons,  $\tau$  leptons, photons, and PF candidates not associated with any of the previous physics objects) and varying each object within its scale and resolution uncertainties in pp collisions.

## 3.1 The online event filter

CMS is a multipurpose detector designed for precise measurements of various physics objects, like leptons, photons, and jets, in pp and nuclear collisions. At design luminosity, the pp interaction rate exceeds 1 GHz; only a small fraction of these collisions contain events of interest, and only a small fraction of those can be stored for offline analysis. The conditions for nuclear collisions, that can be currently delivered with a bunch spacing of at least 100 ns<sup>1</sup>, are typically different from those in the pp case. For example, the instantaneous luminosity delivered by LHC in the 2016 proton-nucleus operational period at  $\sqrt{s_{\text{NN}}} = 8.16 \text{ TeV}$  was about  $10^{29} \text{ cm}^{-2}\text{s}^{-1}$ , resulting in interaction rates of hundreds of kHz, i.e., much lower than in typical pp conditions and with a negligible probability for simultaneous interactions. It is up to the trigger system to select the interesting events for offline storage from the bulk of the (inelastic) collision events.

To select events of potential physics interest, the CMS trigger system [163] utilizes two levels: the first level (L1) is a custom hardware trigger, and the second level, a high-level trigger (HLT), consists of custom software routines. At L1 interesting events are selected at a rate of under 100 kHz based on data from calorimeters and muon detectors. This equates to a rate reduction by a factor of about 400 compared to the bunch crossing frequency.

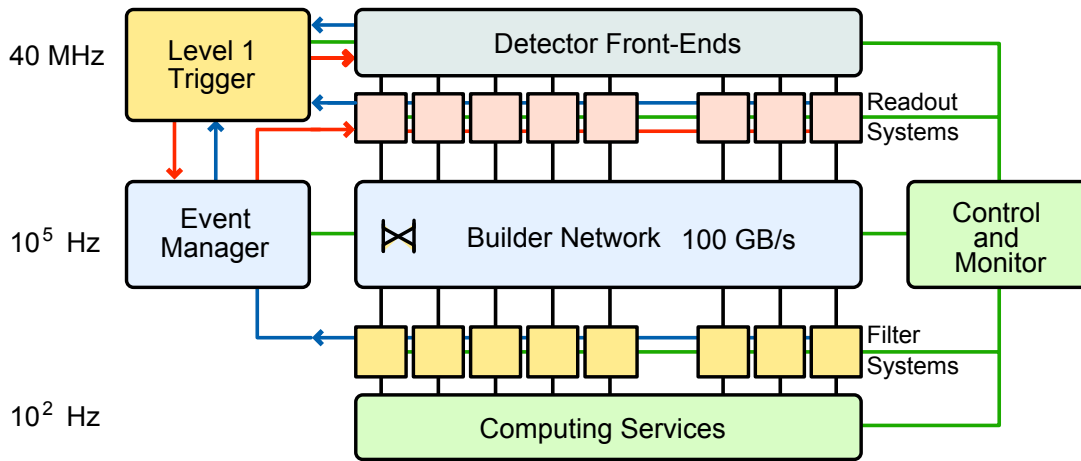
The HLT processor farm of commercial computers consists of a streamlined version of the offline reconstruction and operates on the event data built by the central data acquisition (DAQ) system (Fig. 3.1), performing a more sophisticated event selection and further decreasing the event rate from around 100 kHz to about 0.1-1 kHz. During LS1 a significant upgrade of the CMS DAQ system was performed in order to cope with the increased number of readout channels as well as with the larger expected data fragments, while the general architecture remained the same [164]. In addition to collecting collision data, the trigger and DAQ systems record information for the data-quality monitoring (DQM) that is meant for identifying and diagnosing problems of the detector in real time (“online”).

The thresholds of the trigger level can be adjusted during data taking in response to the value of the LHC instantaneous luminosity, and the overall output rate can be further manipulated by reducing (“prescaling”) the number of events that pass the selection criteria.

### 3.1.1 L1 trigger overview

The L1 trigger (currently) takes input from the calorimeters and the muon system to select the events of physics interest. Trigger primitives are generated on the front-ends of the subdetectors and are combined to physics objects such as muons, electrons or photons, and jets, as well as event-level information such as missing transverse energy, before a final decision is rendered in the global trigger (GT) within about  $4 \mu\text{s}$  after the collision. The specification for CMS electronics is to operate with a trigger rate at L1 up

<sup>1</sup>For the first time, injectors provided a reduced bunch spacing of 75 ns in 2018.



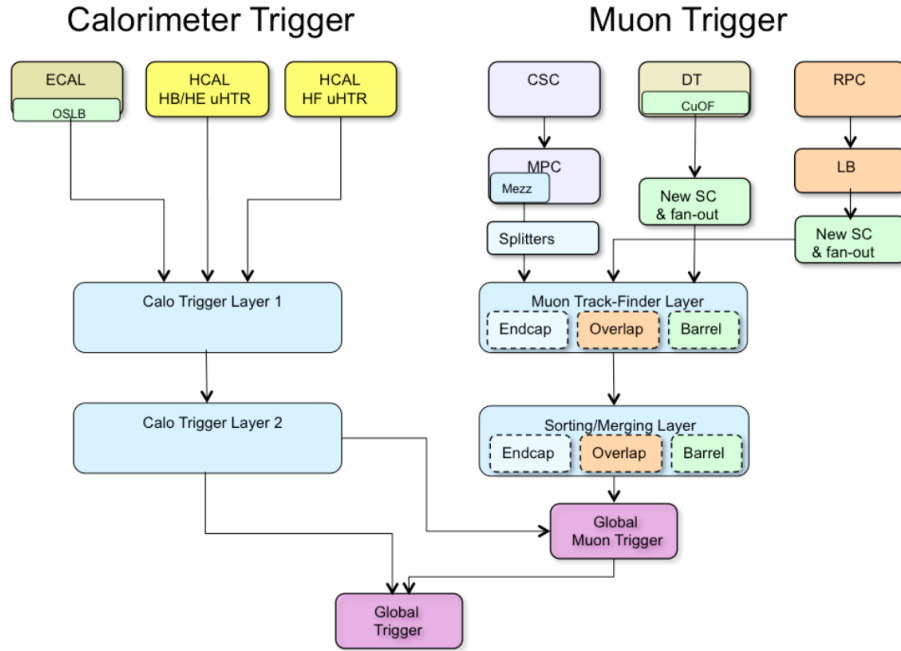
**Figure 3.1:** Simplified schematic view of the key building blocks for a single slice of the CMS DAQ architecture [72]. L1 trigger reduces the incoming event rate of 40 MHz provided by LHC to about 100 kHz, the maximum rate the detector front-ends can withstand. Readout units (RUs) receive event fragments at this rate and provide them to the builder units (BUs). The final event selection is then operated in the filter unit (FU) with the HLT algorithms being executed within the FU using full granularity event data from all subdetectors, while the overall event flow—as illustrated with the solid arrows—is directed by the event manager (EVM). In addition to filtering the L1 output stream, FUs also provide online DQM. Computing services include a host of monitoring services, storage and the interface of the DAQ to the offline environment. The output event rate of the filter units is in the order of 0.1–1 kHz, and is stored on tape.

to 100 kHz, i.e., the limit for low “deadtime” readout operation (the percentage of time during data taking when collisions occur but no triggers can be recorded). With increasing energy, luminosity performance, and pileup, either a substantial increase in trigger thresholds would have been enforced to fit within the 100 kHz limit, exerting a detrimental impact on the physics reach of the experiment, or major upgrades to accommodate the required readout rate would have been required.

To avoid raising the L1 trigger thresholds, a refined data reduction architecture has been introduced during the long shutdown 1 by retaining offline-like features, e.g., subtracted energy sums (“isolation”) for electromagnetic and muon objects, improved position and  $p_T$  resolution for muon tracks, the introduction of invariant-mass calculation, etc. In particular, the calorimeter trigger was redesigned to consist of two processing layers, named “Layer 1” and “Layer 2,” and the muon trigger chain is split into three parallel track finding systems to reconstruct muon tracks using all muon detectors in a given geometrical region. While there were separate track finding systems for the RPC, CSC, and DT detectors in the previous trigger, the upgraded system introduces a barrel, overlap, and endcap muon track finder. The GT finally determines whether a readout process shall be initiated based on data received from the calorimeter trigger chain as well as the muon tracks, with a maximum of 512 separate selections, while the task of system control and synchronization is separately implemented in the Trigger Control and Distribution



System (TCDS). This is illustrated in Fig. 3.2.



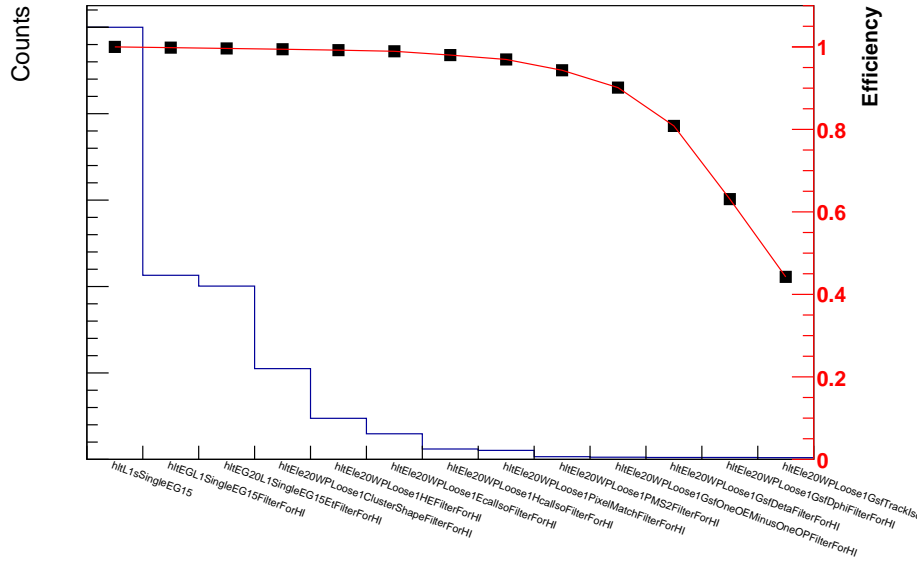
**Figure 3.2:** Overview of the CMS L1 trigger system [165]. The calorimeter trigger is organized into a time-multiplexing trigger, allowing the full event information to be processed in the Layer-2 processors, after preprocessing and multiplexing in Layer 1. The input to the calorimeter trigger system originates from HCAL, HF, and ECAL. In the muon trigger, whose input is obtained from RPC, CSC, and DT, the paradigm changed from a subdetector-based track finding system to a system reconstructing all tracks in a given geometric region. The Global Muon Trigger (GMT) is the final stage of the upgraded muon trigger chain that receives multiple muon objects, but transmits to the Global Trigger (GT) only the eight best muons.

### 3.1.2 HLT overview

The HLT is implemented in software, and further refines the purity of the physics objects. The HLT event selection is performed in a similar way to that used in the offline processing. For each event, objects such as leptons, photons, and jets are reconstructed, and identification criteria are applied. The presence of high- $p_T$  leptons or photons strongly indicates interesting hard collisions, and hence particular attention has been devoted to an efficient set of triggers for these processes.

### 3.1.3 Electron and photon triggers

Electrons and photons (EG or “electromagnetic objects”) are reconstructed primarily using ECAL energy deposits, with little energy deposited in the hadron calorimeter. Electrons can be distinguished from photons based on the presence of tracks that point to electrons and lack thereof for photons. At L1, since only information from the calorimeter is available, no distinction can be made between  $e$  and  $\gamma$ ; only at the HLT level, tracks can be used to resolve this ambiguity.



**Figure 3.3:** Performance of the internal stages of the lowest- $E_T$  unprescaled single-electron trigger during the pPb 2016 run. From left to right the rate is shown for steps associated to L1 seeding, cluster shape, calorimetric isolation, H/E ratio, pixel matching, energy-momentum and directional compatibility, and track isolation.

The electron and photon identification at HLT begins with a regional reconstruction of the deposited energy in the ECAL crystals around the EG candidates retrieved from L1. This is followed by building a “supercluster” (SC) using offline reconstruction algorithms (see Section 3.3.3). Electron and photon candidates are initially selected based on the  $E_T = \sum_i E_i \sin \vartheta_i$  of the SC, with  $E_i$  the energy seen by the calorimeters for the  $i$ th cluster and the sum running over all particles emitted into a fixed solid angle in the event, and on criteria based on properties of the energy deposits in the calorimeters. The common selection requirements include [166]

#### Common selection observables

- $\sigma_{\eta\eta}$ : a cluster shape variable equals to the root-mean-square of the width in  $\eta$  of the shower.
- **isolation**: additional, i.e. after footprint removal, energy deposits in blocks of ECAL and HCAL are measured in a region around the EG candidate
- **ratio to the HCAL energy**: in a fixed cone size centered on the SC relative to the SC energy.

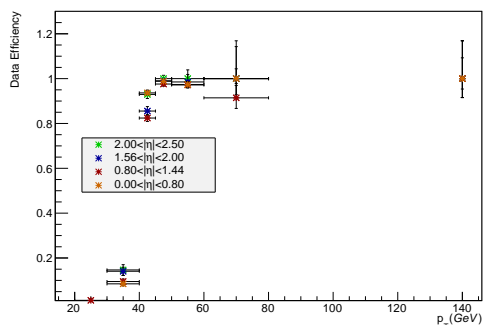
After the common selection, the online electron candidates are further subjected to requirements involving the tracker. The presence of a reconstructed track compatible with the SC is the basis of

#### Electron and photon candidates observables

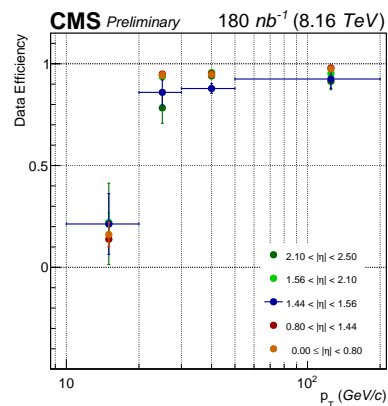
- **pixel matching**: the energy and position of the SC to propagate a hypothetical trajectory through the magnetic field under each charge hypothesis to search for compatible hits in the pixel detector. Full silicon tracks are then reconstructed from the resulting pixel seeds.
- **electron tracking**: based on a simple Kalman filter technique [167] it is comple-

mented by the Gaussian-Sum Filtering (GSF) algorithm, which better parametrizes the highly non-Gaussian electron energy loss, and the reconstructed luminous region (see Section 3.2.4) position.

- **energy-momentum compatibility:** the inverse momentum of the electron tracks must be compatible with the SC inverse energy.
- **directional compatibility:** their direction at the last tracker layer should match the SC position in  $\eta$  and  $\phi$ .
- **track isolation:** requirements with respect to the tracks reconstructed around the electron candidate are applied if required for rate reasons.



**Figure 3.4:** The single-electron trigger efficiency at HLT as a function of the offline reconstructed  $E_T$ , separately in the EB and EE regions, with the lowest unrescaled threshold of  $E_T = 40$  GeV in pp collisions at  $\sqrt{s} = 5.02$  TeV [AN1].



**Figure 3.5:** The electron trigger efficiency at HLT as a function of the offline reconstructed  $E_T$ , separately in the EB and EE regions, with the lowest unrescaled threshold of  $E_T = 20$  GeV in pPb collisions at  $\sqrt{s_{NN}} = 8.16$  TeV [AN2].

The lowest-threshold inclusive single isolated electron filter at the 2016 pPb run, corresponding to instantaneous luminosity of about  $10^{29} \text{ cm}^{-2} \text{ s}^{-1}$ , had a threshold of  $E_T > 20$  GeV. Figure 3.3 shows how the rate is gradually reduced by the filtering steps of this trigger, along with the relative contribution of each step. Using a “tag-and-probe” technique (see Appendix B), efficiencies are computed with respect to a standard offline selection (Figs. 3.4 and 3.5).

### 3.1.4 Muon triggers

The muon HLT combines information from both the muon and the tracker detectors to identify muon candidates and determine their  $p_T$ . The muon HLT algorithm is composed of two main steps, level-2 (L2), which uses information from the muon system only, and level-3 (L3), which combines measurements from both tracker and muon detectors.

**Level-2** The reconstruction of a track in the muon spectrometer starts from seeding DT and CSC segments. Each seed is then used to reconstruct a track using measurements

(hits and segments) from all the muon detectors using the Kalman filter technique. The track reconstruction is followed by the removal of possible duplicates of the same muon candidate checking that tracks do not share any hits. The luminous region position is used to constrain the track parameters to improve the transverse momentum resolution. The multiplicity and relevant parameters of successfully reconstructed L2 muons are used to filter the event. The main selection is based on the muon  $p_T$ , while the number of muon chambers and measurements used in the track fit, e.g., goodness-of-fit  $\chi^2$ , can also be used to suppress misreconstructed muons.

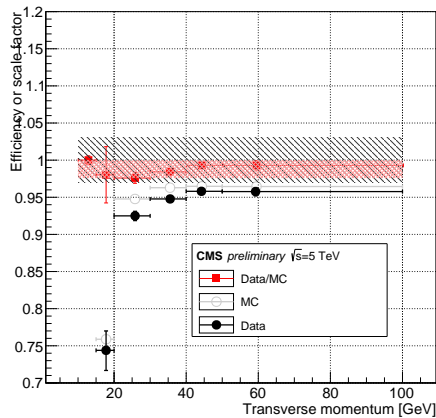
**Level-3** The L3 muon reconstruction exploits the excellent momentum and vertex resolution of the inner silicon tracker to further improve the momentum resolution at high  $p_T$ . Owing to timing and software constraints, the full tracker reconstruction is not performed. Instead, the L3 muon trigger algorithm consists of three main steps: (i) seeding of tracker reconstruction starting from L2 information, (ii) reconstruction in the tracker, and (iii) combined fit in the tracker and muon systems. These tracks and the L2 muons are propagated to a common surface, e.g., the innermost layer of the muon system, and their compatibility is evaluated using criteria, such as angular separation or goodness-of-fit  $\chi^2$ . If one or more L3 muons are successfully reconstructed, their number and parameters are used to filter the event; the main selection is based on the muon  $p_T$ , whereas additional track parameters, such as pixel hits and impact parameter, can be used to suppress misreconstructed muons.

**Isolation** The isolation of L3 muons can be also considered and is evaluated combining information from the silicon tracker, ECAL, and HCAL. Tracks reconstructed in the silicon tracker in a fixed cone size around the L3 muon, and are summed up with ECAL and HCAL deposits. The calorimeter deposits are typically corrected for the average energy density,  $\rho$ , in the event using the FASTJET technique [168].

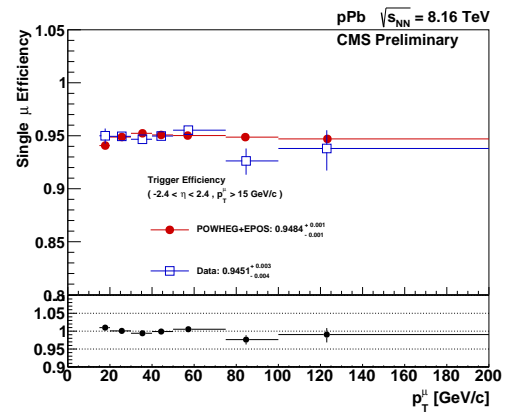
The used inclusive single muon filter for the 2015 reference pp and 2016 pPb runs had thresholds of  $p_T > 15$  and  $> 12$  GeV, respectively, and no isolation requirement. Efficiencies are measured in data and compared to simulation, and are shown to be highly-efficient and well-understood (Figs. 3.6 and 3.7).

### 3.1.5 Beam position timing trigger

Several subdetectors may not provide trigger primitives, but still, generate simple binary logic signals for their inclusion in the trigger menu logic. For instance, the two LHC beam position monitors closest to the interaction point for each LHC experiment are reserved for timing measurements which are called the Beam Pick-up Timing eXperiment (BPTX) detectors [169]. For CMS, they are located at a distance of approximately 175 m on either side of the interaction point (BPTX+ and BPTX-). The trigger typically selects “zero-bias” events, i.e., valid bunch crossings, using the digitized BPTX signal. Requiring a logical conjunction of BPTX+ with BPTX- (BPTX\_AND) a coincidence of the signals from the detectors on either side is established. To suppress noise in triggers with high background



**Figure 3.6:** The single-muon trigger efficiency at L3 as a function of the off-line reconstructed  $p_T$  with a threshold of  $p_T = 15$  GeV in pp collisions at  $\sqrt{s} = 5.02$  TeV [AN1, AN3, AN4].



**Figure 3.7:** The single-muon trigger efficiency at L3 as a function of the off-line reconstructed  $p_T$  with a threshold of  $p_T = 12$  GeV in pPb collisions at  $\sqrt{s_{NN}} = 8.16$  TeV [AN5].

a coincidence with BPTX\_AND is thus required.

Since the interaction probability per bunch crossing during nuclear collisions is small, a complementary trigger to select hadronic interactions is typically deployed. In that case, the collision events are selected online by requiring the coincidence of signals from both the BPTX devices, hence indicating the presence of both proton and lead (or lead and lead) bunches crossing the IP, and at least one energy deposit above a readout threshold on either side of HF. The selection offline can be further refined by repeating the energy deposition requirement on each of the two sides of HF, and possibly imposing a requirement on the reconstructed PV multiplicity. This selection that is based on coincidences between the signals from the +z and -z sides of BPTX and HF is referred to as a “minimum-bias” trigger.

### 3.2 Hit, track, and primary vertex reconstruction

The scientific goals of CMS impose stringent requirements on the performance of the tracking system. Reconstructing tracks in a high-occupancy environment is challenging; it requires to attain high track-finding efficiency while keeping the fraction of “fake” tracks small. Fake tracks are falsely reconstructed tracks that may be formed from a combination of unrelated (random) hits or from a genuine particle trajectory that gets badly reconstructed from the inclusion of spurious hits. In addition, the data acquisition system of the strip detector, which runs algorithms on off-detector electronics (modules of the front-end driver (FED) [170]), must control throughput level sufficiently fast to be used not only for offline event reconstruction, but also for the CMS HLT.

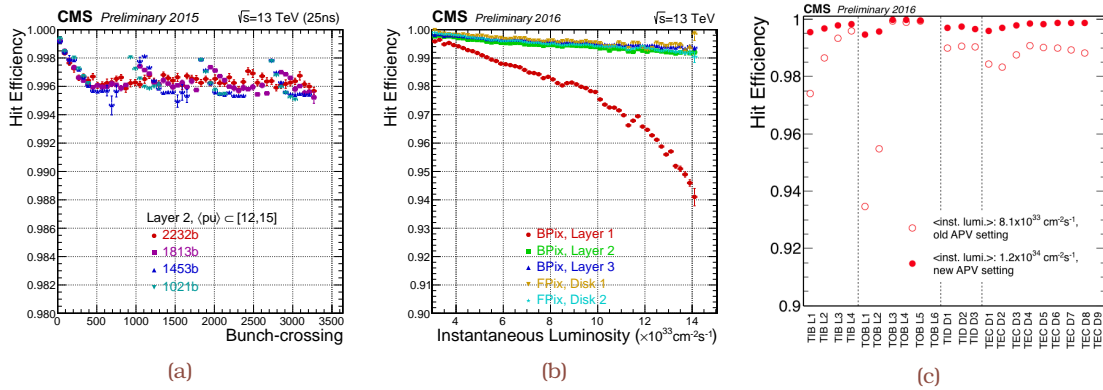
Hits created by charged particles crossing the tracker’s sensitive layers must be reconstructed. The algorithms used to reconstruct tracks from these hits, along with the performance obtained in terms of track-finding efficiency, the proportion of fake tracks,

and track parameter resolution, are explained in Ref. [145].

Primary vertices from collisions are distributed over the luminous region colloquially known as the “beamspot.” This is intimately linked to tracking since, on the one hand, the beamspot and primary vertices are found using reconstructed tracks, and on the other hand, an approximate knowledge of their positions is needed before track finding can be initiated.

### 3.2.1 Hit reconstruction performance

“Local” reconstruction consists of clustering zero-suppressed signals above specified thresholds in pixel and strip channels into hits, and then estimating the cluster positions and their uncertainties<sup>2</sup>. As a result, information for only a small fraction of the channels in any given event is retained for offline storage. The hit efficiency is the probability to find a cluster in a given silicon sensor that has been traversed by a charged particle. In the pixel tracker, the average efficiency for reconstructing hits is typically >99% (Fig. 3.8a), albeit the hit efficiency depends on the instantaneous luminosity, as shown in Fig. 3.8b. The systematic uncertainty in these measurements is estimated to be below 0.5% [171]. Several sources of loss can be identified, e.g., the limited size of the internal buffer of the readout chips causing a dynamic inefficiency that increases with the instantaneous luminosity, readout errors signaled by the FED modules depending on the rate of beam-induced background, etc.



**Figure 3.8:** Hit efficiency for different LHC pp filling schemes as a function of (a) the bunch crossing identification number for the second barrel layer and (b) instantaneous luminosity for all barrel and forward disks of the silicon pixel detector [171]. (c) Hit efficiency of the silicon strip detector planes. During late 2015 and early 2016, the strip tracker observed a loss of hits due to the saturation of the APV25 readout chip [172]. Measurements on last detector planes, i.e., T0B L6 and TEC D9, have been omitted because of the bias (underestimation) induced by the analysis method.

The overall hit efficiency in the strip tracker is about 99.8% (Fig. 3.8c). This number is compatible with the 0.2% fraction of defective channels observed during the construction of the strip tracker [145]. During late 2015 and early 2016, the higher instantaneous

<sup>2</sup>In PbPb collisions, the number of produced particles depends strongly on the geometrical overlap of the Pb ions. As a consequence, to avoid bias, the tracker is read out without hardware zero suppression with the latter being performed offline.

luminosity values at LHC increased the detector occupancy. Charges generated near the back-plane of the sensitive volume of the thicker silicon sensors is inefficiently collected by the APV25 readout chip [173], and corrections are applied to compensate for the cluster barycenter shift along the direction perpendicular to the sensor plane. The increased detector occupancy saturated the APV25 readout chip, that integrates the collected charge in a narrow time window near the back-plane of the thicker silicon sensors, resulting in high deadtime and a further loss of hits. After having changed the drain speed of the pre-amplifier, thus allowing for faster recovery, the efficiency was fully recovered.

### 3.2.2 Track reconstruction performance

Tracking algorithms reconstruct charged particles over the full pseudorapidity range of the tracker  $|\eta| < 2.5$ , finding tracks with  $p_T$  as low as 0.1 GeV, or produced as far as 60 cm from the beam line. Initially, it was thought that track finding should be seeded using hits in the outer layers of the tracker, where the channel occupancy is relatively low [143]. However, later it has been broadly perceived that the pixel tracker is better suited to this purpose featuring high granularity, and hence providing algorithms with excellent resolution in the three spatial dimensions and an even lower channel occupancy, despite the higher track density [145]. In addition, many particles lose a significant fraction of their energy interacting with the tracker material. To ensure high efficiency the track finding process uses trajectory seeds created in the inner region of the tracker, facilitating the reconstruction of low-momentum tracks that are deflected by the strong magnetic field before reaching the outer part of the tracker.

In a typical LHC event containing jets, the track-finding procedure yields a significant fraction of fake tracks. The fake rate can be reduced substantially through quality requirements. Tracks are selected on the basis of the number of layers that have hits, whether the estimated  $\chi^2/\text{dof}$  of their fit is good, and how compatible they are with originating from a primary interaction vertex. The selection criteria can be summarized as [145]

- A requirement on the minimum number of layers in which the track has at least one associated hit. This differs from selections based on the number of hits on the track, because more than one hit in a given layer can be assigned to a track, as in the case of layers with overlapping sensors or double-sided layers in which two sensors are mounted back-to-back.
- A requirement on the minimum number of layers in which the track has an associated three-dimensional hit, i.e., in the pixel tracker or matched hits in the strip tracker.
- A requirement on the maximum number of layers intercepted by the track containing no assigned hits, not counting those layers inside its innermost hit or outside its outermost hit, nor those layers where no hit was expected because the module was known to be malfunctioning.
- $\chi^2/\text{dof} < a_0 N_{\text{layers}}$ .
- $|d_0^{\text{BS}}|/\delta d_0 < (a_3 N_{\text{layers}})^\beta$ .
- $|z_0^{\text{PV}}|/\delta z_0 < (a_4 N_{\text{layers}})^\beta$ .

- $|d_0^{\text{BS}}|/\sigma_{d_0}(p_T) < (a_1 N_{\text{layers}})^\beta$ .
- $|z_0^{\text{PV}}|/\sigma_{z_0}(p_T, \eta) < (a_2 N_{\text{layers}})^\beta$ .

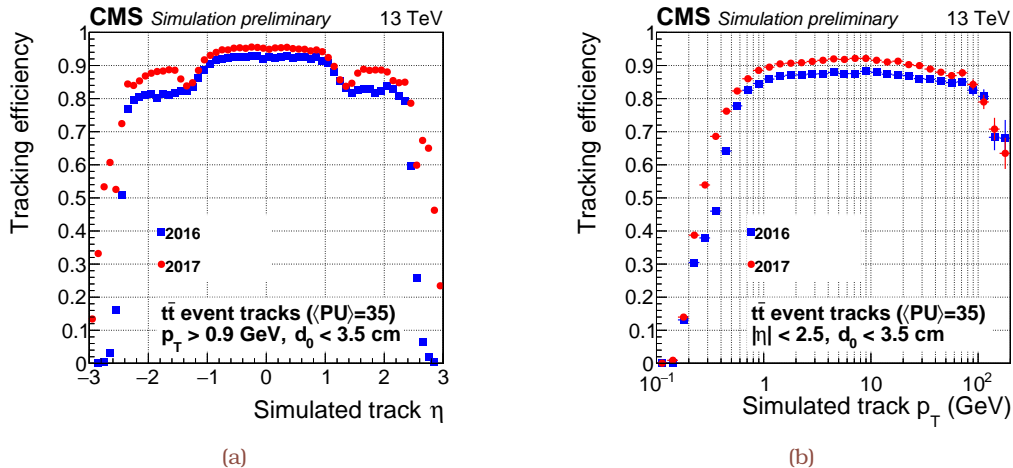
The parameters  $a_i$  and  $\beta$  are configurable constants. The track’s impact parameters are  $d_0^{\text{BS}}$  and  $z_0^{\text{PV}}$ , where  $d_0^{\text{BS}}$  is the distance from the center of the beamspot in the plane transverse to the beam line and  $z_0^{\text{PV}}$  is the distance along the beamline from the closest pixel vertex; the impact parameter uncertainties,  $\delta d_0$  and  $\delta z_0$ , are calculated from the covariance matrix of the fitted track trajectory. The “high-purity” track selection criteria provide stringent requirements that reduce the efficiency and fake rate and are listed in Table 3.1 for each iteration.

**Table 3.1:** Parameter values [145] used in selecting high-purity tracks reconstructed in each step; the number of layers that contain hits assigned to tracks, the parameter  $a_0$  that controls selection criteria based on  $\chi^2/\text{dof}$ , the parameters  $a_i$  and  $\beta$  that define the compatibility of impact parameters with the interaction point. Iterations 2 and 3 use two paths that emphasize track quality (Trk) or primary vertex compatibility (PV). A track produced by these iterations is retained if it passes either of these criteria. These quality criteria are adjusted to maintain high efficiency and low fake rate as a function of track  $p_T$  and  $N_{\text{layers}}$ .

Iteration	$\beta$	Min layers	Min 3D layers	Max lost layers	$a_0$	$a_1$	$a_2$	$a_3$	$a_4$
0 & 1	4	4	4	2	0.9	0.30	0.35	0.40	0.40
2 Trk	4	5	3	1	0.5	0.90	0.90	0.90	0.90
2 PV	3	3	3	1	0.9	0.85	0.80	0.90	0.90
3 Trk	4	5	4	1	0.5	1.00	1.00	1.00	1.00
3 PV	3	3	3	1	0.9	0.90	0.90	1.00	1.00
4	3	6	3	0	0.3	1.00	1.00	1.00	1.00
5	3	6	2	0	0.25	1.20	1.10	1.20	1.10

For prompt, charged particles of  $p_T > 1$  GeV, under typical LHC conditions and in simulated inclusive  $t\bar{t}$  events (Fig. 3.9), the average track reconstruction efficiency, i.e., the fraction of reconstructed tracks that can be associated with the corresponding simulated charged particles, is expected approximately to be 92 (94)% in the barrel region of the phase-0 (phase-I) tracker. It decreases to about 80 (85)% at higher pseudorapidity with most of the inefficiency caused by hadrons undergoing nuclear interactions in the tracker material. In the same  $p_T$  range, the fraction of falsely reconstructed tracks, i.e., the fraction of reconstructed tracks that are not associated with any simulated particle, is at the few percent level. In the central region, tracks with  $1 < p_T < 100$  GeV have a resolution in  $p_T$  of better than 1.5%. In this momentum range, the resolution in the track parameters is dominated by multiple scattering. The resolution in their transverse and longitudinal impact parameters remains approximately constant at 160 and 180 (90 and 100)  $\mu\text{m}$  respectively, in the central region of the phase-0 (phase-I) tracker, while it progressively deteriorates towards higher pseudorapidity.





**Figure 3.9:** Track reconstruction efficiency—the number of reconstructed tracks sharing at least 75% of their assigned hits with the simulated particles divided by the total number of generated particles—as a function of the simulated track (a)  $\eta$  and (b)  $p_T$  [172]. The phase-I pixel detector (installed in 2017) exhibits an improved performance relative to the phase-0 detector all over the  $\eta$  and studied  $p_T$  spectrum. The reconstructed tracks are simulated based on PYTHIA (v.8 [50, 51]) using the 4C tune [174], and are required to meet the “high-purity” requirement (Table 3.1). The number of pileup interactions superimposed on each simulated  $t\bar{t}$  event is randomly generated from a Poisson distribution with a mean value of 35.

### 3.2.3 Primary vertex reconstruction performance

The goal of primary vertex (PV) reconstruction is to measure the location, and the associated uncertainty, of all hadron-hadron, hadron-nucleus, or nucleus-nucleus interaction vertices in each event, including the event vertex from the hard scattering and any vertices from pileup (PU) collisions, using the available reconstructed tracks. It consists of three main steps

- selecting the tracks
- clustering the tracks that appear to originate from the same interaction vertex
- fitting the position of each vertex using its associated tracks.

The identification of the event vertex from the hard scattering is a key element for the reconstruction and identification of b jets among others. The vertex position is estimated with an adaptive vertex fit [145] using a collection of tracks compatible with originating from the same interaction. During Run 1 operations, with an average of 21 additional pp collisions per bunch crossing, the event PV was identified as the reconstructed vertex with the largest  $\sum p_T^2$  of its associated tracks. When the number of additional interactions in the same bunch crossing increases, the chosen PV is not always identified correctly as the vertex corresponding to the hard interaction.

A more robust method has been developed to meet the requirements associated with higher PU. The recently introduced technique [175] consists of replacing the individual tracks contributing to  $\sum p_T^2$  by pseudojets, i.e., by clustering the tracks originating from the same vertex using the anti- $k_T$  jet clustering algorithm (see Section 3.3.4) with a dis-

tance parameter of 0.4. For each primary vertex,  $\sum p_T^2$  is computed by using these jets as well as the  $p_T^{\text{miss}}$  at the PV (“track  $p_T^{\text{miss}}$ ”) in order to account for neutral particles. Different weights are applied to pseudojets and missing transverse momentum because of the experimental precision in determining their  $p_T$  magnitude. The PV with the largest weighted  $\sum p_T^2$  is then chosen as the one corresponding to the hard scattering, and the jets associated with it are used to quantify the performance of the b tagging algorithms.

The efficiency for choosing the right PV depends on the event kinematics. It can be estimated in the simulation by comparing its reconstructed  $z$  position to that of the generated hard interaction. Figure 3.10 displays simulated  $t\bar{t}$  events with MADGRAPH5 [176] interfaced to PYTHIA (v6.424 [49], and using the Z2\* [58, 177] tune) for three detector scenarios [175]: phase-I detector with an average pileup of 50 (blue squares), phase-I-“aged,” i.e., with modeling of the effects of radiation damage after  $1 \text{ ab}^{-1}$  of integrated luminosity, and phase-II (green circles) detectors with an average PU of 140 (red triangles) separately considered for the PV sorting algorithms in Run 1 (open symbols) and 2 (full symbols). The Run 1 algorithm for PV sorting would be indeed significantly less efficient. The pixel-detector extension in phase II will provide an improved efficiency for the PV choice in the forward region  $2.4 < |\eta| < 3.0$  compared to the phase-I-aged detector configuration.

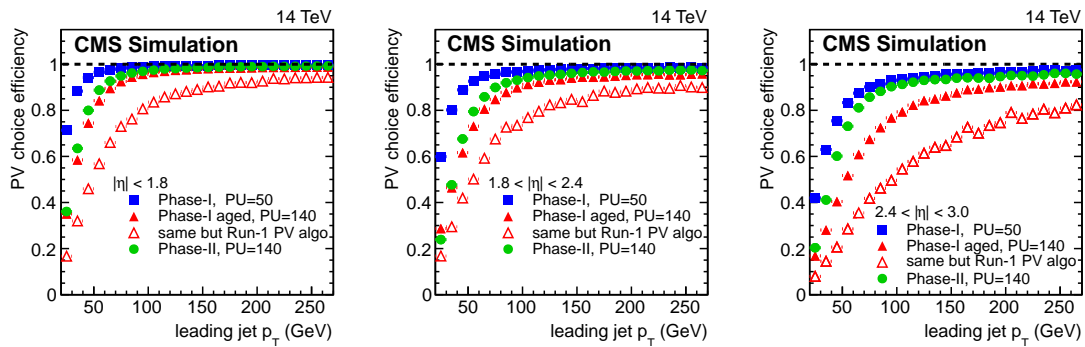
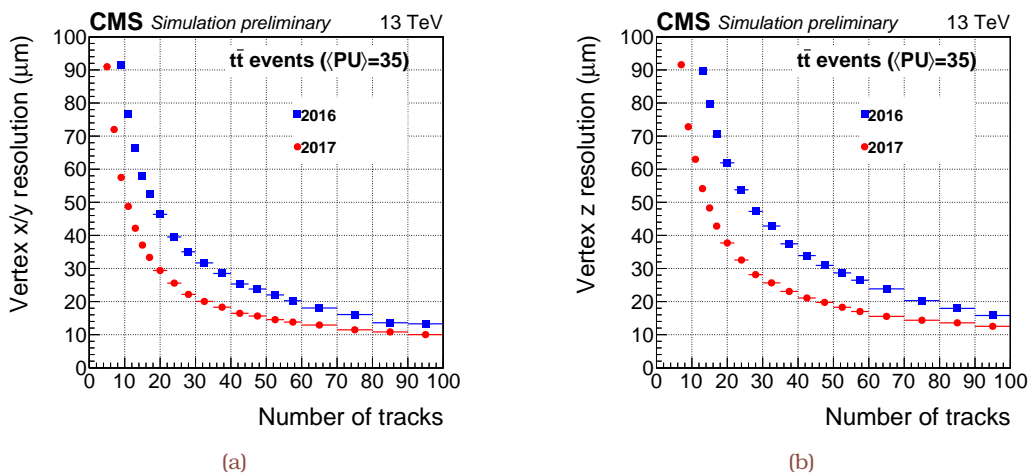


Figure 3.10: Efficiency in choosing the correct PV of the hard interaction as a function of the leading jet  $p_T$  progressively from central (left) to forward (right)  $|\eta|$  regions for  $t\bar{t}$  events simulated with MADGRAPH5 [176] interfaced to PYTHIA (v6.424 [49], and using the Z2\* [58, 177] tune) for three detector scenarios [175].

The resolution in the position of a reconstructed PV strongly depends on the number of tracks used to fit the vertex and the  $p_T$  of those tracks. A track-splitting method [145] is employed for measuring the resolution as a function of the number of tracks associated to the vertex. The tracks used in any given vertex are split into two equal sets. During the splitting procedure, the tracks are first sorted in descending  $p_T$  order, and then combinatorially paired starting from the track with the largest  $p_T$ . For each pair, tracks are randomly assigned to one or the other set of tracks. This ensures that the two sets of tracks have, on average, the same kinematic properties. These two sets of tracks are then fitted independently with an adaptive vertex fitter. Finally, to extract the resolution the distributions in the difference of the fitted vertex positions for a given number of tracks are parameterized using a single Gaussian distribution, whose fitted root-mean-square

(RMS) width is divided by a factor of  $\sqrt{2}$ , because the two used subsets should nominally have the same resolution. The range of the fit is constrained to be within twice the RMS of the distributions.

Results from a study of the PV resolution in transverse and longitudinal directions as a function of the number of tracks associated with the vertex, using jet-enriched simulated data samples, are shown in Fig. 3.11. The resolution approaches  $10\ \mu\text{m}$  in  $x$  and  $y$ , and it is about  $12\ \mu\text{m}$  in  $z$  for primary vertices using at least 50 tracks. For zero- or minimum-bias events, the resolutions are worse across the full range of the number of tracks used to fit the vertex, and less than about 20 and  $25\ \mu\text{m}$  in transverse and longitudinal directions, respectively, for primary vertices reconstructed using at least 50 tracks.

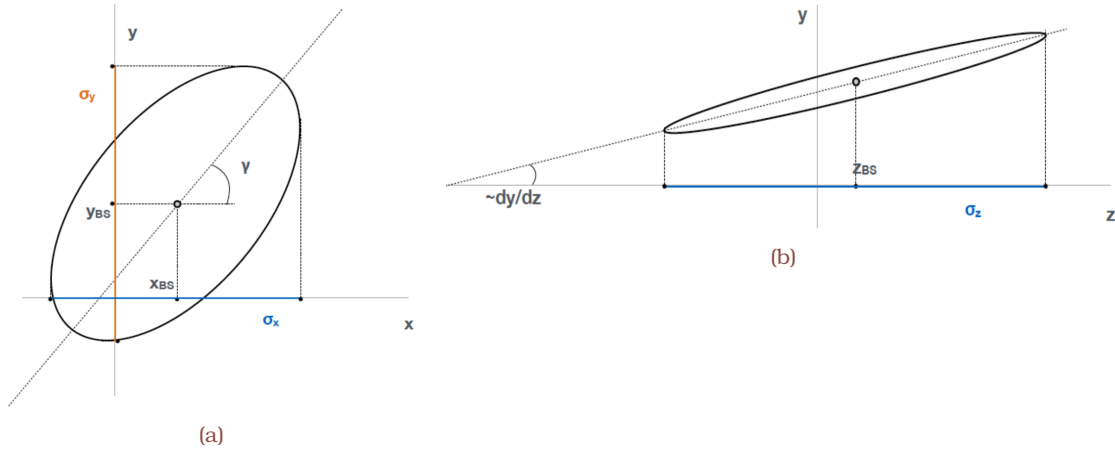


**Figure 3.11:** Primary vertex transverse (a) and longitudinal (b) resolution in events simulated with PYTHIA (v.8 [50, 51]) using the 4C tune [174] as a function of the number of tracks used in the adaptive vertex fit [178]. The reconstructed tracks associated to primary vertices are selected with the high-purity requirement of Ref. [145], while vertices within a radial (longitudinal) distance of 24 (2) cm are retained. The phase-I pixel detector (installed in 2017) exhibits an improved performance relative to the phase-0 2016.

### 3.2.4 Reconstruction of the LHC luminous region

The measurement and continuous monitoring of the luminous region, where the two LHC beams collide at a given interaction point, play a crucial role both for the trigger selection and the event reconstruction. Studying the luminous region is of great interest for the experiments given the interplay between the yearly integrated luminosity performance and the detector event reconstruction efficiency, which depends on pileup, and are routinely performed to optimize parameters, like the levelling time or Fill duration, for different filling scheme scenarios, e.g., Ref. [179]. Given the size of transverse (longitudinal) dispersion, typically a few  $\mu\text{m}$  (cm), the position of the luminous region provides an excellent estimate of the position of the interaction point. This is of primary importance especially for the track reconstruction; the beamspot position is used to constrain the track fitting, when the primary vertices of the event are not yet determined, and to constrain the track clustering in the longitudinal direction for reconstructing the primary

vertices of the event. The precise determination of the beamspot position allows to monitor real-time the position of the beams, and hence minimizing the radiation dose in the tracker and providing the accelerator operators with valuable feedback. To keep the exposure to ionizing radiation uniform in  $\phi$  it is desirable to keep the beam at the center of the tracking detector. Last but not least, the beamspot parameters, as measured in data, are deployed to MC event sample simulation of the PV distribution.



**Figure 3.12:** The position and size of the luminous region at the interaction point [AN6] are measured with respect to the CMS reference frame centered on the mechanical support of the tracker. Transverse (a) and longitudinal (b) distributions of the collision region can be parameterized by Gaussian functions with typical widths of the order of few  $\mu\text{m}$  and cm, respectively, and are accompanied by potential beam offsets in  $x$ ,  $y$  and  $z$  from the nominal center of the detector; angles spanning the detector  $z$  axis and the interaction vertex should also be encoded, and are distributed according to  $\beta^*$  and beam emittance.

The measured beamspot parameters are the coordinates of the center ( $x^{BS}$ ,  $y^{BS}$ , and  $z^{BS}$ ), the widths ( $\sigma_x$ ,  $\sigma_y$ , and  $\sigma_z$ ), and the derivatives (slopes) with respect to the  $z$  axis ( $\frac{dx}{dz}$  and  $\frac{dy}{dz}$ ), and can be determined in two ways [180]. The first method is through the reconstruction of primary vertices, which maps out the collisions as a function of  $x$ ,  $y$ , and  $z$ , hence the shape of the beamspot. The position of the center and the size of the luminous region is determined through a fit to the three-dimensional distribution of vertex positions. The second method, inherited from CDF (e.g., Refs. [181] and [182]), utilizes the correlation between the impact parameter ( $d_0$ ) and azimuthal angle ( $\phi$ ) of tracks originating from the same PV. When the center of the beamspot is displaced relative to its expected position, i.e., when the beam is displaced with respect to the detector coordinate system, the  $d_0$ - $\phi$  distribution exhibits a sinusoidal modulation that can be fitted with a parameterization including the beamspot parameters as:

$$d_0(\phi, z_0) = x^{BS} \sin \phi + \frac{dx}{dz} \sin \phi (z_0 - z^{BS}) - y^{BS} \cos \phi - \frac{dy}{dz} \cos \phi (z_0 - z^{BS}). \quad 3.1$$

The two methods are typically checked against each other to provide consistent results, and a combination of both methods is required to measure the full set of beamspot parameters.

The beamspot can be measured with a per-lumisection (LS) granularity, i.e., short periods of about 23 s each, to protect against effects from orbit drifts. Such granularity is crucial for dedicated measurements of the predicted beam displacement at the IP (see Section 4.3.1). A more precise estimate of the beamspot parameters is typically required though, and hence a weighted average is calculated combining measurements from coarser time intervals, the latter colloquially referred to as Intervals of Validity (IOV). This means a consecutive operation can be viewed as an ordered IOV sequence, each with its own beamspot measurement.

Representative  $z^{\text{BS}}$  and  $\sigma_z$  measurements are shown in Fig. 3.13 for the 2015 pp period at 13 TeV. The results are consistent with those expected from the LHC tuning that centered the  $z^{\text{BS}}$  coordinate to zero from Fill 4386 onward. The longitudinal width has been measured to be typically between four and five cm. Its reduction throughout a Fill can be explained in terms of emittance evolution (shrinkage), that has been independently detected by dedicated “emittance” scans and transverse beam profile monitors [183], and is consistent with the high synchrotron radiation damping at 13 TeV [184]. The longitudinal width is practically unaffected by the absence of the magnetic field because of its larger magnitude, whereas the transverse widths exhibit much larger variations between 0 and 3.8 T [185].

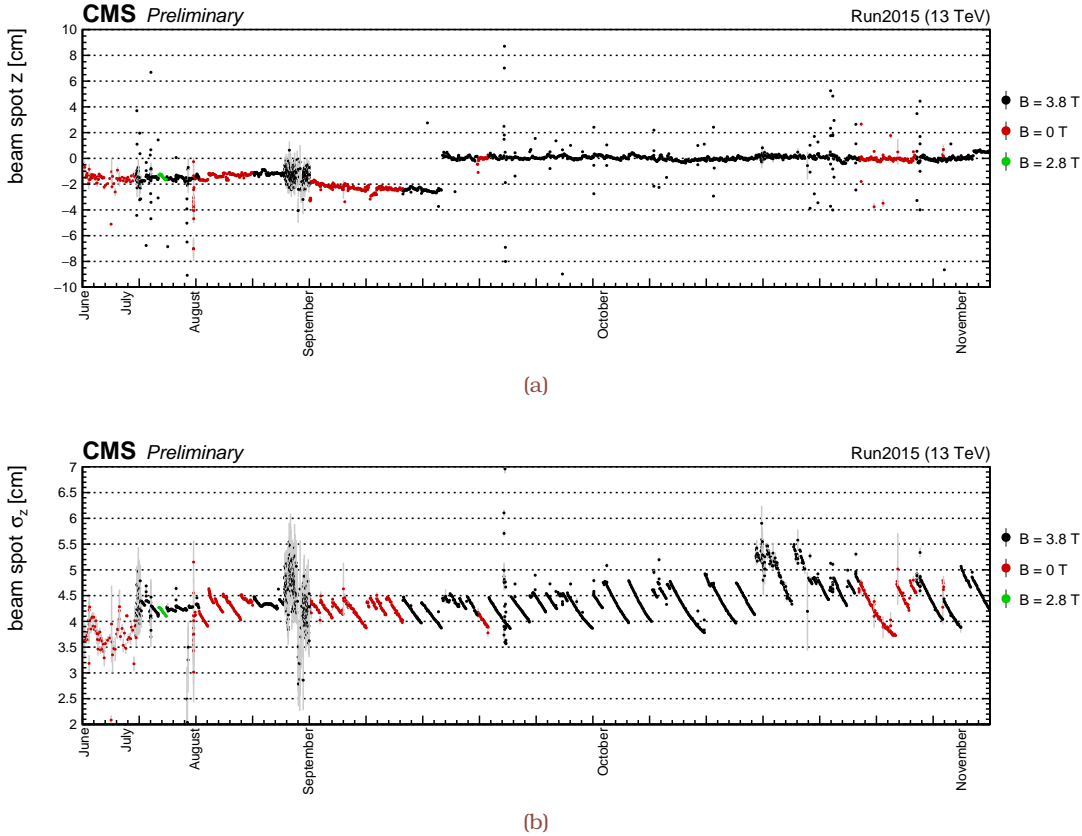
### 3.3 Physics object reconstruction

The concept of cylindrical detection layers, nested around the beam axis, formulates the reconstruction of “physics objects” from signals collected by a given subdetector as:

- The reconstruction of “isolated” photons and electrons primarily concerns the ECAL.
- The identification of muons is principally based on the information from the muon detectors.
- The energy of the jets, consisting of hadrons and photons, can be inclusively measured by the calorimeters, i.e., without any attempt to separate individual jet particles. The same argument applies to  $\vec{p}_T^{\text{miss}}$  reconstruction.
- The “tagging” of jets originating, e.g., from the b quark hadronization or hadronic  $\tau$  decays is based on the properties of the pertaining charged particle tracks, and thus mostly involves the tracker.

A significantly improved event description is achieved by correlating the basic elements from all detectors (tracks and clusters) to identify each final-state particle, and finally combining the relevant measurements to reconstruct the particle properties on the basis of this identification. This universal approach is called the “particle-flow” (PF) reconstruction. The PF concept was developed and used for the first time by the ALEPH experiment at LEP [186].

A key ingredient in this approach is the fine spatial granularity of the detector layers since coarse-grained detectors may cause the signals from different particles to overlap, thereby reducing the particle identification and reconstruction efficiencies. The level of ability to individually identify the particles from the hard scatter is assessed by the extent of the proton or ion debris, the particles from pileup interactions, the particle proximity



**Figure 3.13:** Examples of measured beamspot parameters,  $z^{\text{BS}}$  and  $\sigma_z$ , at IP5 using pp collisions at  $\sqrt{s} = 13 \text{ TeV}$  [185]. Data were recorded with the CMS magnet at 3.8 (black dots), 2.8 (green dots) or 0 T (red dots). The absence of the magnetic field, and hence the partial loss of information about  $\vec{p}_T^{\text{miss}}$ , impacts all the errors of the beamspot parameters returned by the fit methods. The widths are affected too since they are strongly correlated with the primary vertex errors. Low pileup Fills—4266, 4268, and 4269 at the end of August—are more prone to larger statistical uncertainty due to the lower number of reconstructed tracks. The error bars indicate the statistical uncertainty of the employed fit methods.

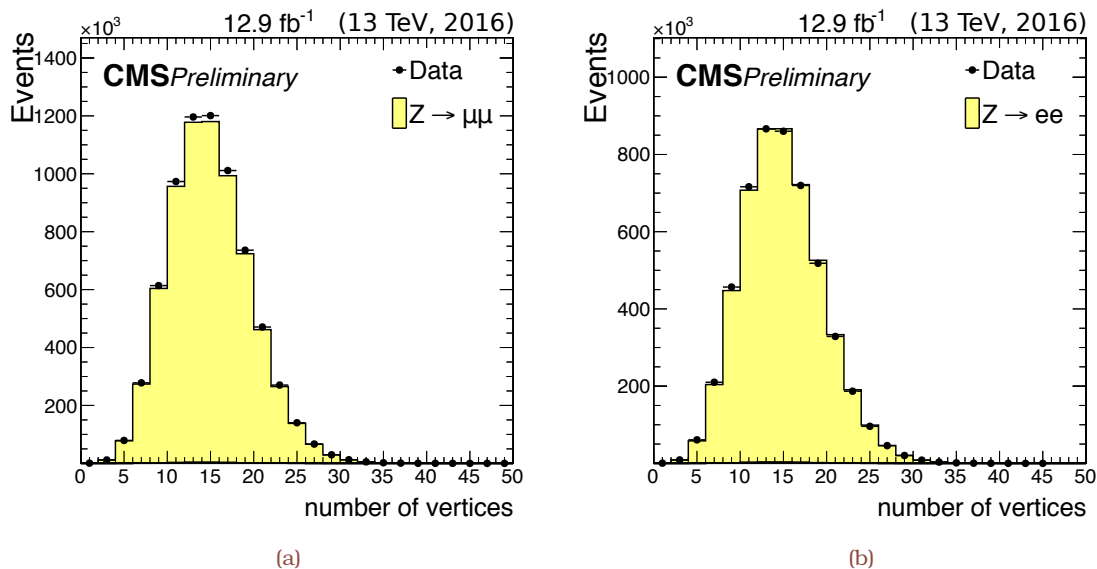
inside high-energy jets, the secondary interactions in the tracker material, etc. In CMS, the PF reconstruction was ready for use in physics analyses in June 2010, and was further implemented at HLT and in heavy ion collision analyses in 2011 [138]. Since then, practically all CMS physics results have been based on PF reconstruction.

### 3.3.1 Pileup mitigation and treatment of the underlying event

The PF algorithm was initially designed without taking into account pileup. During standard LHC running conditions, simultaneous interactions occur per bunch crossing whose average number depends on the specifications of the data-taking period. These interactions are randomly spread along the beam axis around the center of the CMS coordinate system, presumably following a normal distribution with a standard deviation of few cm (see Section 3.2.4 for the RMS of the luminous region). The number of pileup interactions can be estimated, e.g., from the PV multiplicity—the vertex reconstruction

efficiency is about 70% for pileup interactions [187]—and the instantaneous luminosity of the given bunch crossing, with the inelastic pp cross section [188] as additional input. Typically the simulated events are assigned a weight such that the distribution of the simulated multiplicity matches the observed distribution, as displayed in Fig. 3.14. The total uncertainty is dominated by the uncertainty in the total inelastic pp scattering cross section measurement [188], which affects the pileup profile in the simulated sample. Less dominant is the uncertainty associated with the luminosity measurement.

Particles produced in pileup interactions give rise to additional PF constituents, e.g., charged hadrons, photons, and neutral hadrons. This results in an average additional  $p_T$  of few GeV per pileup interaction and unit area in the transverse plane. As a consequence, reconstructed particles from pileup affect, among others, jets,  $\vec{p}_T^{\text{miss}}$ , and the isolation of charged leptons. The measured energy deposits in the calorimeters used as input for particle reconstruction may also be directly affected by pileup interactions.



**Figure 3.14:** Multiplicity of reconstructed primary vertices in  $Z \rightarrow \mu\mu$  (a) and  $Z \rightarrow ee$  (b) candidate events at  $\sqrt{s} = 13$  TeV [189] using a data sample recorded in early 2016 (up to period “Run2016C”). For the remainder of the year, LHC achieved instantaneous luminosity values of up to  $1.53 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ , with an average number of collisions occurring at one proton beam bunch crossing about 27 and a maximum of more than 50, assuming a total inelastic cross section of 80 mb.

As noted in Section 3.2.3, the PV with the largest weighted  $p_T$  (LV) is considered as the one corresponding to the hard scattering, whereas the other reconstructed vertices are referred to as the pileup vertices. Charged hadrons reconstructed within the tracker acceptance can be identified as coming from pileup by associating their track with a pileup vertex. If identified as coming from pileup, these charged particles are removed from the record of reconstructed particles that are used to form physics objects. This widely used algorithm is called “pileup charged-hadron subtraction,” and is referred to as CHS. All neutral and charged particles associated with LV or not associated with any PV are kept in the CHS procedure.

Photons and neutral hadrons, as well as all reconstructed particles outside the tracker acceptance, cannot be associated with one of the reconstructed PV using the CHS technique though. To mitigate the impact of these particles the uniformity of the  $p_T$  density of pileup interactions in the transverse plane facilitates the average  $p_T$  contributions expected from pileup to be subtracted. The  $p_T$  density from pileup interaction can be calculated with jet clustering techniques, such as FASTJET, with the list of all reconstructed particles as input. As an alternative, this contribution can be estimated locally, e.g., around a given lepton, from the expected ratio of the neutral to the charged energy from pileup, typically 0.5 (see Section 3.3.2). After the end of Run 1, advanced pileup mitigation techniques have been explored, e.g., Ref. [190, 191], and become increasingly important (see Appendix C.0.1 for a dedicated study) with the larger number of pileup interactions observed during the LHC Run 2.

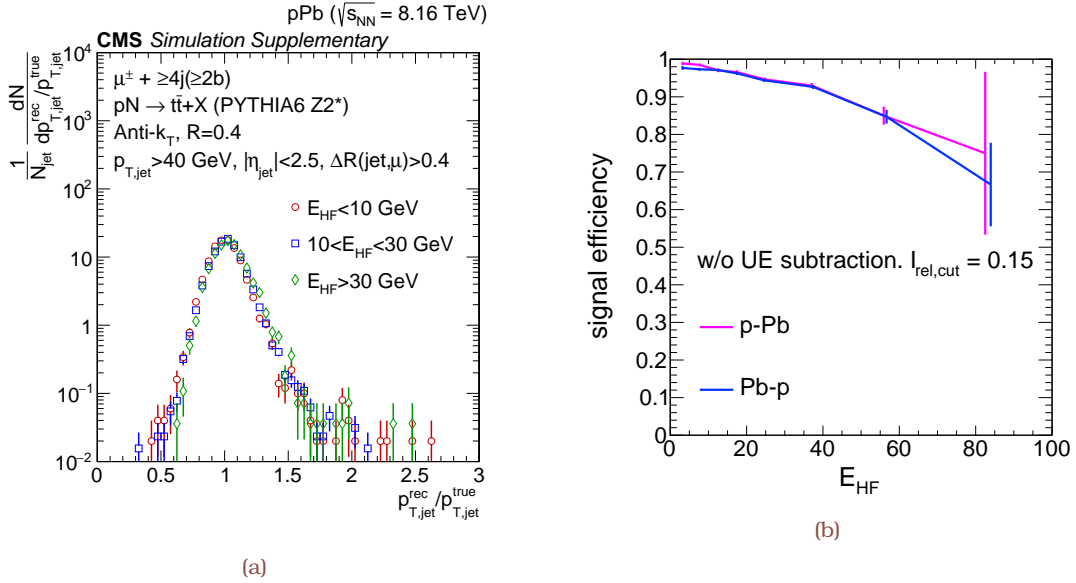
Fluctuations in the transverse energy flow of secondary particles produced in nuclear collisions result in increased soft background relative to hadron-hadron collisions. To account for the influence of these underlying events the initial hard scattering is embedded into simulated events using, e.g., the EPOS-LHC (v.3400) generator [192] that is tuned to reproduce global event properties as measured in data such as the total particle multiplicities, charged-hadron spectra, and anisotropic transverse flow. The procedure involves preprocessing the digitized (“raw” [193]) format of the events, rendering the vertex distribution along the luminous region as observed in data. More specifically, for each event that satisfies a minimum-bias trigger (see Section 3.1.5) selection, the embedding is done by mixing the simulated digital signal information from hard scattering and EPOS. These events are then propagated through the standard reconstruction and analysis chain using the GEANT4 package [194] to simulate the detector response.

The effect of the underlying event on the reconstruction of jets and the isolation of charged leptons is shown in Figs. 3.15a and 3.15b, respectively, using simulated  $t\bar{t}$  events passing the analysis selection of Ref. [TH1] (Section 6.3). The jet energy scale is shown for a various selection of the summed transverse energy in the forward calorimeters in the Pb-going direction, which is a measure of the uncorrelated underlying event produced in pPb collisions. The performance of charged lepton isolation is studied in data using high- $p_T$  muons that are required to form a pair of opposite sign with an invariant mass between 80 and 100 GeV. The efficiency loss due to the isolation is studied as a function of the energy in HF in the Pb-going direction. The muon detection efficiency decreases when the underlying event increases, considering a constant isolation requirement (Eq. (3.2)) of 0.15. For the average HF energy ( $\sim 21$  GeV), the muon isolation efficiency is about 0.93, and uncertainty of 4% is estimated taking a variation of two standard deviations around the mean.

### 3.3.2 Muon reconstruction and identification

In the standard CMS reconstruction procedure, tracks are first reconstructed independently in the tracker (“tracker track”) and the muon system (“standalone-muon track”), and then used as input for muon track reconstruction. “Global muon tracks” are built





**Figure 3.15:** (a) Jet energy scale for jets in simulated  $t\bar{t}$  events using pPb collisions at  $\sqrt{s_{\text{NN}}} = 8.16$  TeV. Reconstructed jets overlapping with the isolated muon are excluded [TH1]. (b) The efficiency loss due to the muon isolation as a function of the energy in the HF subdetector. Similar conclusions are drawn for electrons [AN2].

by matching standalone-muon with tracker tracks. Reconstructed muons are then processed by the PF algorithm, that applies a set of selection criteria to candidates reconstructed with the standalone, global, or tracker muon algorithms [156]. The requirements are based on various quality parameters from the muon reconstruction, and information from other subdetectors is also used, as described in the following. Emphasis is exclusively given on muons that are reconstructed as both belonging to the tracker and global classes, and also selected by the PF algorithm.

The set of variables and selection criteria are defined to allow an optimum balance between efficiency and misidentification rate of about 95% and 1%, and using simulated  $Z \rightarrow \mu\mu$  and QCD multijet events [156], respectively. This “tight” selection thus aims to identify prompt muons, i.e., originating from LV, and suppress contributions from light and heavy flavor decays as well as the misidentification from charged hadrons.

#### Muon reconstruction observables

- Number of hits per track: The tracker track should have hits from at least six layers of the tracker including at least one pixel hit.
- $\chi^2/\text{dof}$  of track fit: The global muon fit must have  $\chi^2/\text{dof} < 10$  and include at least one hit from the muon system.
- Degree of matching between tracker and standalone track: The tracker muon must have segment matching in at least two of the muon stations. The muon segment compatibility is computed by propagating the tracker track to the muon system, and evaluating both the number of matched segments in all stations and the closeness of the matching in position and direction.

#### Global event observables

- $|d_0|$  and  $|d_z|$ : A tight muon must be compatible with LV, having a transverse and

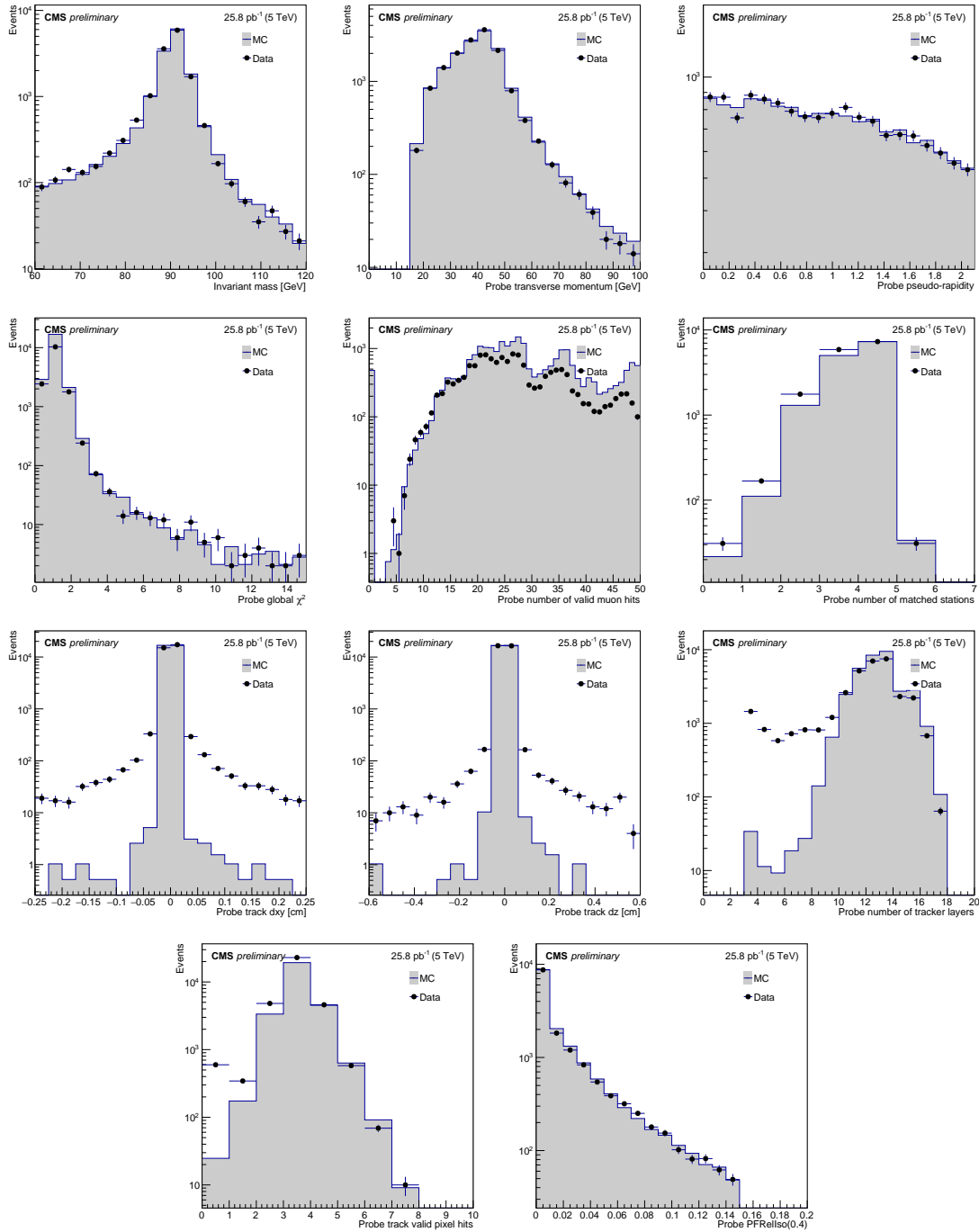


Figure 3.16: Muon kinematic (invariant mass,  $p_T$ , and  $|\eta|$ ) and identification variables for  $Z \rightarrow \mu\mu$  events at  $\sqrt{s} = 5.02$  TeV. Data (Table B.1) are found to be in good agreement with the MC (Table 5.1) simulation—scaled to the amount of data—in all the considered variables. Identification variables are shown prior to any requirement [AN3, AN1, AN4], and hence the non-simulated background is expected to populate the discrepant regions.

longitudinal impact parameter of  $|d_0| < 0.2$  and  $|d_z| < 0.5$  cm, respectively.

To further distinguish prompt charged leptons from those originating, e.g., from semileptonic hadron decays, the scalar  $p_T$  sum of surrounding particles, often referred

to as the “isolation,” is a widely used discriminating variable. Although charged particles associated with pileup vertices can be removed from the isolation, following the CHS paradigm, the contamination of neutral particles from pileup interactions still holds. Various techniques to mitigate this effect have been developed. A widely used example is the so-called  $\delta\beta$  correction, which estimates the contribution of neutral particles from pileup based on the contribution of charged particles associated with pileup vertices. That is, the muon isolation,  $I^\mu$ , is defined as

$$I^\mu = \sum_{\Delta R(i,j) < 0.3, 0.4}^{\text{CH(LV)}} p_T^j + \max \left( 0, \sum_{\Delta R(i,j) < 0.3, 0.4}^{\text{NH}} p_T^j + \sum_{\Delta R(i,j) < 0.3, 0.4}^{\text{PH}} p_T^j - \frac{1}{2} \sum_{\Delta R(i,j) < 0.3, 0.4}^{\text{CH(PU)}} p_T^j \right), \quad 3.2$$

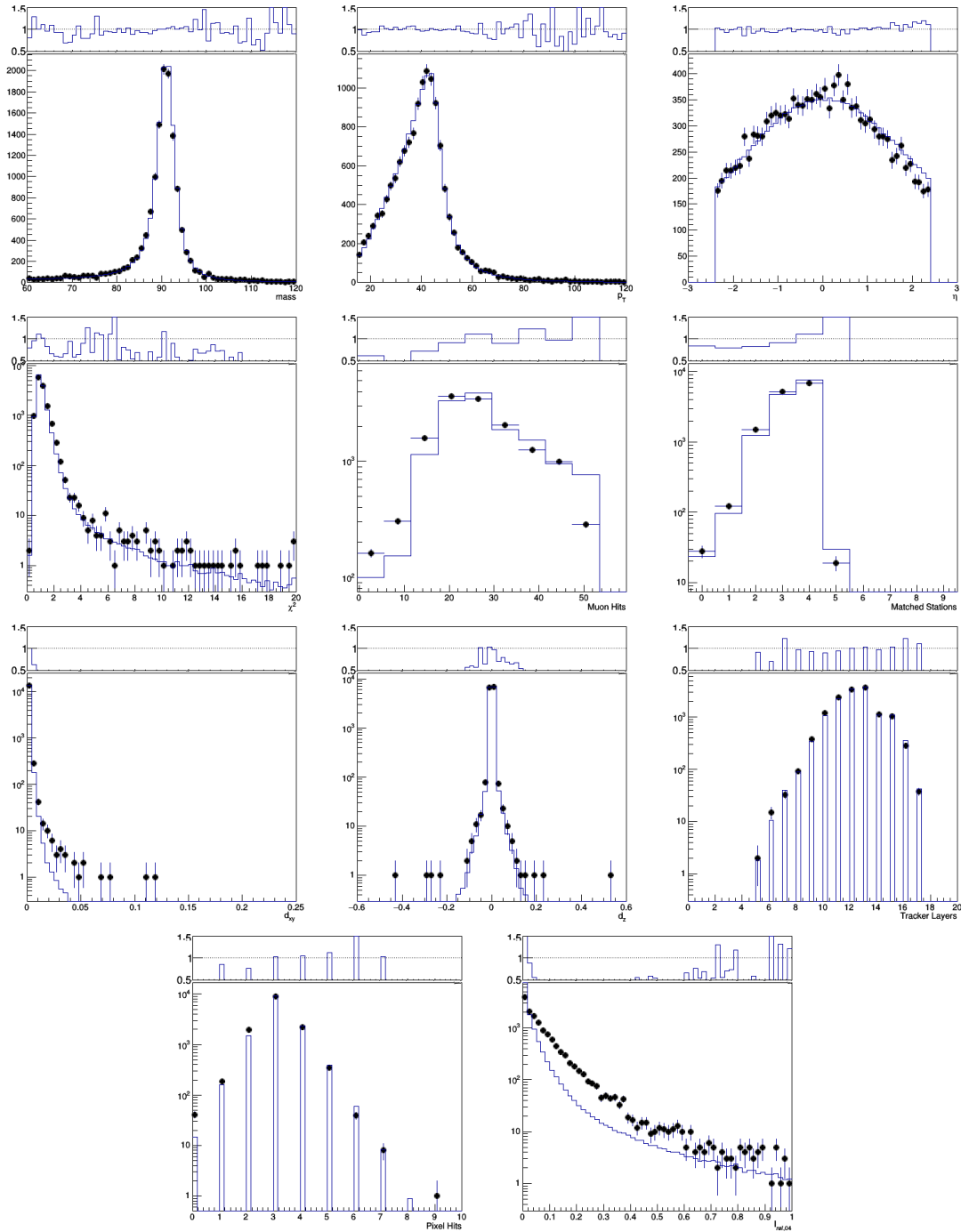
where each  $p_T$  sum is performed over the particles within a cone  $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$  of size 0.3 (in pPb collisions) or 0.4 (in pp collisions), “CH (LV)” and “CH (PU)” are charged particles associated with LV and pileup vertices, respectively, and “NH” and “PH” are neutral hadrons and photons. In this technique, the pileup contamination from neutral particles is estimated as a half, motivated by the isospin limit in which jets are approximately composed of 1/3 neutral and 2/3 charged pions. As such, the  $\delta\beta$  correction subtracts the contamination on average. The isolation is typically further divided by the transverse momentum of the lepton, and it is referred to as the “relative isolation” ( $I_{\text{rel}}^\mu$ ). The default value entering the tight identification using the PF isolation  $I_{\text{rel}}^\mu$  of  $\Delta R < 0.4$  is 0.15, whereas it has been further optimized for final states involving single  $t$  and  $\bar{t}$  quarks in view of Refs. [TH12] and [TH13] (see Appendix C).

Figures 3.16 and 3.17 show muon kinematic and identification variables of the probe leg in the tag-and-probe pair, as well as the invariant mass of the tag-and-probe system, comparing data and simulation using pp and pPb collisions at  $\sqrt{s} = 5.02$  and  $\sqrt{s_{\text{NN}}} = 8.16$  TeV, respectively.

### 3.3.3 Electron reconstruction and identification

Excellent electron reconstruction and selection efficiencies, together with small misidentification probability over a large phase space, and good momentum resolution are of great importance in many analyses. One of the main challenges for precise reconstruction of electrons in CMS is the tracker material (Fig. 2.14), which causes significant bremsstrahlung along the electron trajectory. In addition, this bremsstrahlung spreads over a large volume because of the CMS magnetic field. Dedicated techniques have been developed (Ref. [152] and references therein) to group the energy deposited in contiguous  $5 \times 5$  arrays of clusters in ECAL, i.e. the final global cluster (supercluster), to build the electron track using a dedicated version of the standard Kalman filter (Gaussian sum filter), and to associate the two inputs for estimating the electron properties. The energy calibration and resolution in ECAL and general issues in track reconstruction are discussed in Refs. [149] and [152], respectively.

The related procedures to identify prompt isolated electrons, as concisely described in the following, are optimized using simulation and commissioned with data. By convention, “energy” and “momentum” refer to the energy of the electromagnetic shower initiated



**Figure 3.17:** Muon kinematic (invariant mass,  $p_T$ , and  $\eta$ ) and identification variables for  $Z \rightarrow \mu\mu$  events at  $\sqrt{s_{\text{NN}}} = 8.16$  TeV. Data (Table B.1) are found to be in good agreement with the MC (Table 6.1) simulation—scaled to the integral of data—in most of considered variables, except for the isolation due to the presence of the underlying event [AN2]. The slight mismodeling in  $|d_0|$  can be attributed to the non-simulated background and imperfect simulation of the tracker alignment.

by the electron in ECAL and to the track momentum measurement in the tracker, respectively, while the term “electron momentum” describes the combined information, taking into account the bremsstrahlung photons spatially compatible with originating from the

Variable	Veto		Loose		Veto		Medium	
	Barrel	Endcap	Barrel	Endcap	Barrel	Endcap	Barrel	Endcap
$\sigma_{\eta\eta}$	0.0114	0.0352	0.0103	0.0301	0.0115	0.037	0.00998	0.0298
$\Delta\eta_{in}$	0.0152	0.0113	0.0105	0.0081	0.00749	0.00895	0.00311	0.00609
$\Delta\phi_{in}$	0.216	0.237	0.1150	0.1820	0.228	0.213	0.103	0.045
$H/E$	0.181	0.116	0.1040	0.0897	0.356	0.211	0.253	0.0878
$I_{rel}^{\rho,corr}$	0.126	0.144	0.04	0.05	0.175	0.1590	0.07	0.08
$ 1/E - 1/p $	0.207	0.174	0.1020	0.1260	0.299	0.15	0.134	0.13
$ d_0 $	0.0564	0.2220	0.0261	0.1180	0.0564	0.2220	0.0118	0.0739
$ d_z $	0.4720	0.9210	0.4100	0.8220	0.4720	0.9210	0.3730	0.6020
Missing inner hits	2	3	2	1	2	3	1	1
Pass conversion veto	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

**Table 3.2:** Requirements for the rectangular-based identification of electrons in pp (“veto” and “loose”) and pPb (“veto” and “medium”) collisions at  $\sqrt{s} = 5.02$  [AN7] and  $\sqrt{s_{NN}} = 8.16$  TeV [AN2], respectively.

electron track.

Variables that provide discriminating power against electron background sources, mainly originating from photon conversions, jets misidentified as electrons, or from semileptonic decays of b and c quarks, can be grouped into three main categories, and are listed in Table 3.2.

#### Purely calorimetric observables

- $\sigma_{\eta\eta}$ : The lateral extension of the shower along the  $\eta$  direction is expressed in terms of the variable  $\sigma_{\eta\eta}$ , which is defined as  $(\sigma_{\eta\eta})^2 = [\sum(\eta_i - \bar{\eta})^2 w_i] / \sum w_i$ . The sum runs over the 5×5 matrix of crystals around the highest- $E_T$  crystal of the SC, and  $w_i$  is a weight that depends logarithmic on the contained energy. The positions  $\eta_i$  are expressed in units of crystals, which has the advantage that the variable-size gaps between ECAL crystals (in particular at modules boundary) can be ignored. The discrimination power of the lateral shower extension is due to the fact that bremsstrahlung strongly affects the energy deposition along the  $\phi$  direction.
- $H/E$ : The fraction of energy deposited in HCAL is expected to be small, as electromagnetic showers are essentially fully contained in ECAL and ES.

#### Tracking observables

- Missing inner hits: The PF algorithm can exploit the pattern of track hits. When photon conversions take place inside the volume of the tracker, the first hit on electron tracks from the converted photons is often not located in the innermost layer of the tracker, and missing hits are therefore expected in that region. For prompt electrons, whose trajectories start from the LV, no missing hits are expected in the inner layers.
- Pass conversion veto: In addition to the missing hits, photon conversion candidates can also be rejected using a fit to the reconstructed GSF tracks. Since the photon is massless, and momentum transfer is in general small, the conversions have a well-defined topology, with tracks that have essentially the same tangent at the conversion vertex. The strategy for rejecting these candidates consists of fitting the track pairs to a common vertex, incorporating this topological constraint, and then

rejecting the converted photon candidates according to a  $\chi^2$  probability of the fit.

- $|d_0|$  and  $|d_z|$ : The impact parameters of the electron, such as the transverse ( $|d_0|$ ) and longitudinal ( $|d_z|$ ) distance to the LV at the point of closest approach in the transverse plane are used to reject secondary electrons.

#### Track-cluster association observables

- $\Delta\eta_{in}$  and  $\Delta\phi_{in}$ : The track-cluster association criterion is designed to preserve high efficiency and reduce misidentification probability, and it is therefore not very restrictive along the direction of the track curvature affected by bremsstrahlung. For ECAL-seeded electrons, this requires a geometrical matching between the GSF track and the SC, such as:
  - $|\Delta\eta| = |\eta_{SC} - \eta_{in}^{extrap}| < 0.02$ , with  $\eta_{SC}$  being the SC energy-weighted position in  $\eta$ , and  $\eta_{in}^{extrap}$  the track  $\eta$  extrapolated from the innermost track position and direction to the position of closest approach to the SC,
  - $|\Delta\phi| = |\phi_{SC} - \phi_{in}^{extrap}| < 0.15$ , with analogous definitions for  $\phi$ .
- $|1/E - 1/p|$ : The track-cluster association includes not only geometrical but also SC energy-track momentum matching. The  $|1/E - 1/p| = |1/E_{SC} - 1/p_{in}|$  variable is typically considered, where  $E_{SC}$  is the SC energy and  $p_{in}$  is the track momentum at the point of the closest approach to LV.

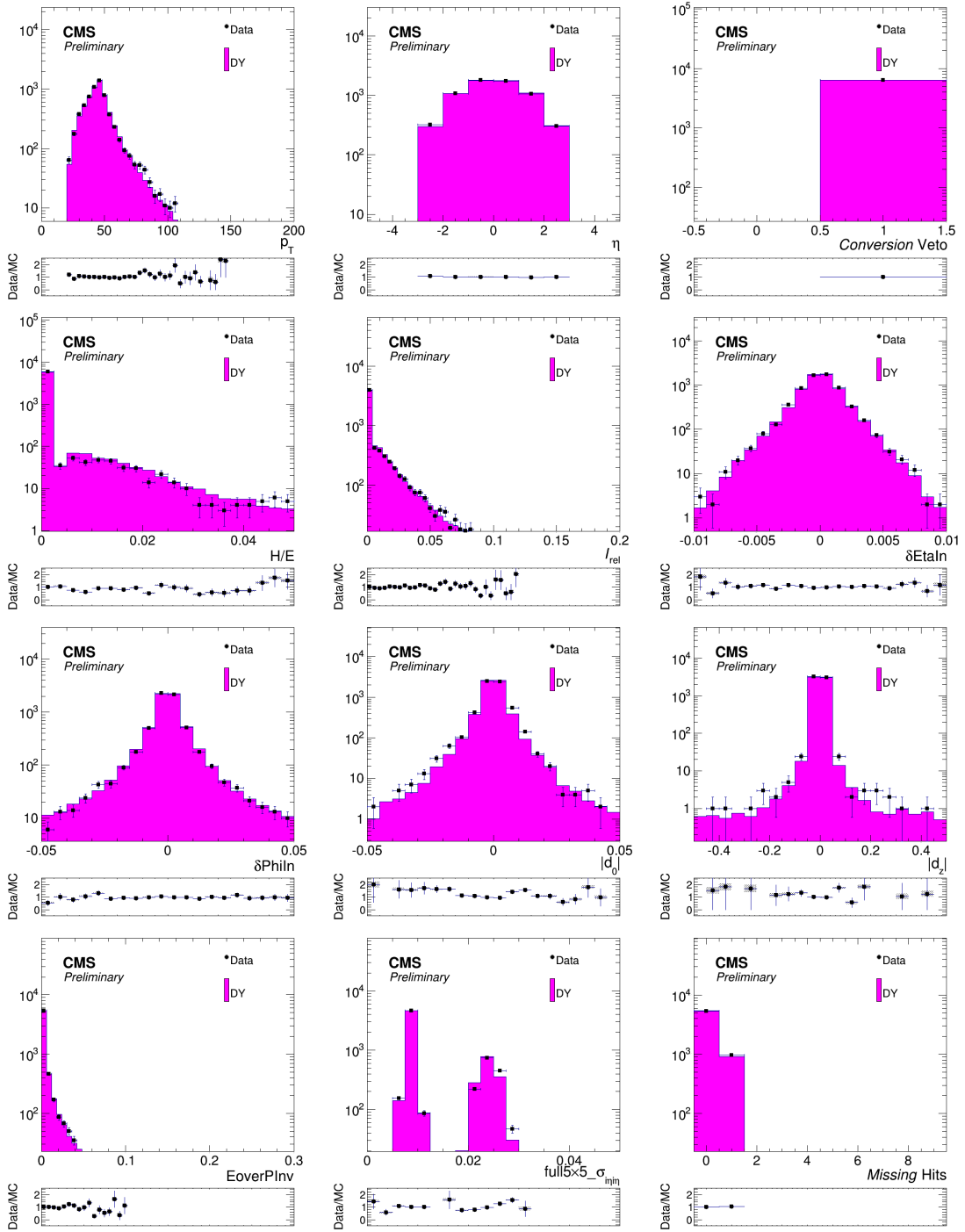
Electron candidates from misidentified jets or semileptonic decays of b or c quarks have significant energy flow near their trajectories, and hence requiring electrons to be isolated from such nearby activity greatly reduces these sources of background. Two isolation techniques are used in CMS. The simplest one is referred to as “detector-based isolation,” and relies on the sum of energy depositions either in ECAL or HCAL around each electron trajectory, or on the scalar sum  $p_T$  of all tracks reconstructed from LV. These sums are usually computed within cone radii of  $\Delta R = 0.3$  or  $0.4$  around the electron direction, and are used at HLT to reject jets misidentified as electrons given their decent performance.

Offline, the PF technique is used to extend the definition of isolation quantities, i.e., rather than using energy measurements in independent subdetectors, the isolation is defined using the PF candidates

$$I^e = \sum_{\Delta R(i,j) < 0.3}^{CH(LV)} p_T^j + \max \left( 0, \sum_{\Delta R(i,j) < 0.3}^{NH} p_T^j + \sum_{\Delta R(i,j) < 0.3}^{PH} p_T^j - p_T^{PU} \right), \quad 3.3$$

where the sums run over the charged PF candidates, neutral hadrons and photons, within a chosen  $\Delta R$  cone around the electron direction. The charged candidates are required to originate from LV, and  $p_T^{PU}$  is the correction related to event pileup.

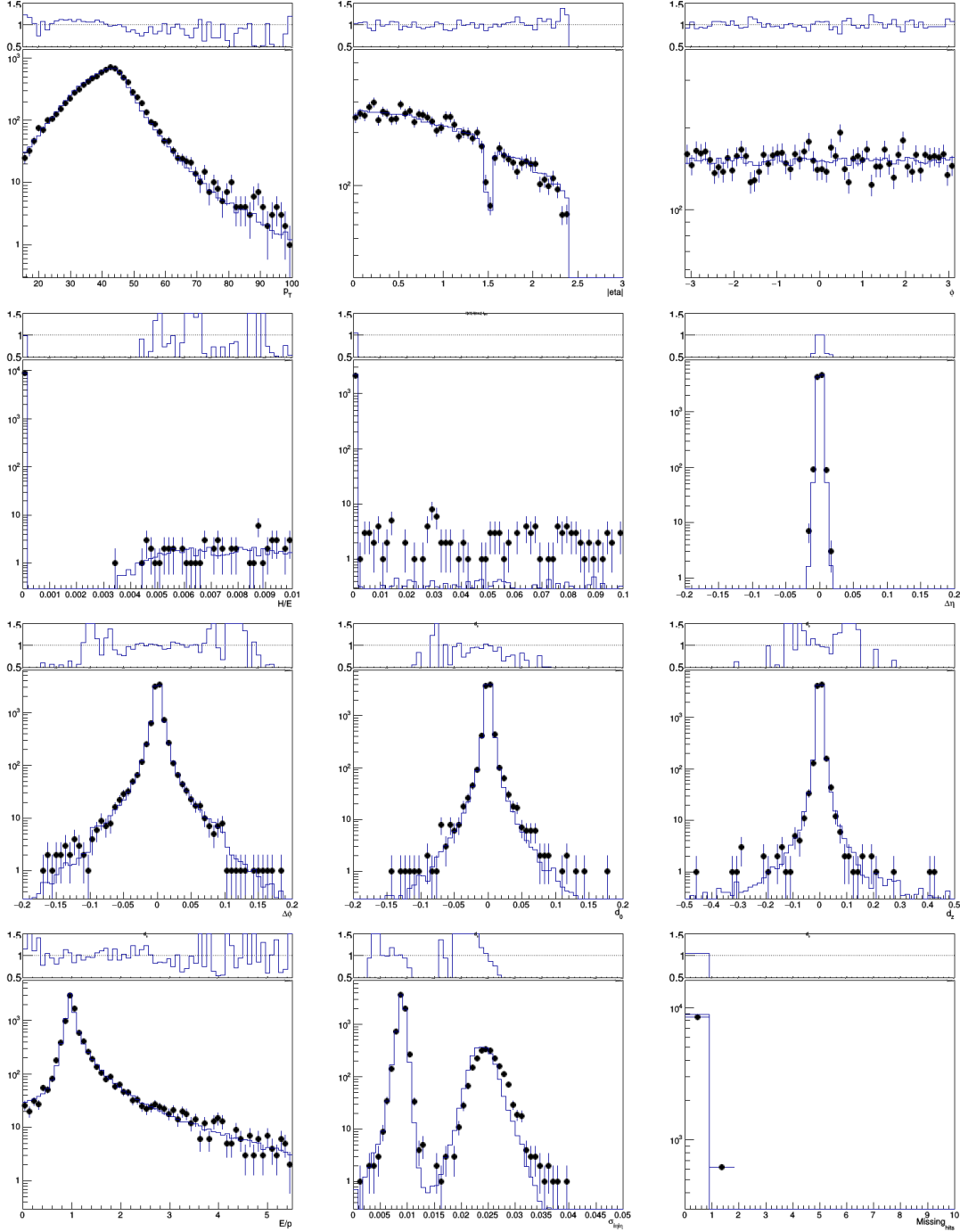
The contribution from pileup in the isolation cone is computed using the FASTJET technique, assuming  $p_T^{PU} = \rho A_{eff}$ ;  $\rho$  is the average energy density, defined as the median of the energy density distribution for all particles within  $|y| < 5$  and a fixed grid size of 0.6. Using the median instead of the mean makes  $\rho$  effectively less sensitive to low energy pseudojets composed of nonphysical particles with infinitesimal momenta and random



**Figure 3.18:** Electron kinematic ( $p_T$ ,  $|\eta|$ , and  $\phi$ ) and identification variables (Table 3.2) for  $Z \rightarrow ee$  events at  $\sqrt{s} = 5.02$  TeV [AN7]. Data (Table B.1) are found to be in good agreement with the MC (Table 5.1) simulation—scaled to the amount of data—in all the considered variables. The slight mismodeling in  $|d_0|$  can be attributed to the non-simulated background and imperfect simulation of the tracker alignment.

direction, and hence reducing the bias from low pileup energy densities. The effective area  $A_{\text{eff}}$  in  $(\eta, \phi)$  plane is defined, for each component of the isolation, as  $(\Delta R)^2$  scaled by the ratio of the slopes for  $\rho$  and for the considered component against the PV multiplicity.

The dependence of both the charged and neutral components on the PV multiplicity is expected to be almost linear.



**Figure 3.19:** Electron kinematic ( $p_T$ ,  $|\eta|$ , and  $\phi$ ) and identification variables for  $Z \rightarrow ee$  events at  $\sqrt{s_{NN}} = 8.16$  TeV [AN2]. Data (Table B.1) are found to be in good agreement with the MC (Table 6.1) simulation—scaled to the integral of data—in all the considered variables. The variable depicted in the middle of the second row is  $I^{\text{CH}} - I^{\text{NH}} - I^\gamma$ .



### 3.3.4 Jet reconstruction and noise-jet rejection

QCD predicts that large-distance nonperturbative effects can be decoupled from the hard reaction so that energetic partons fragment into collimated streams or “jets” of hadrons, which inherits energy and momentum from their parent parton. A prescription is then required to define what “collimated” means, i.e., to unambiguously decide upon what objects can be ascribed to a jet given some measure of distance. The outcome of the jet-clustering procedure must deal with the cancellation of collinear and soft singularities appearing in QCD, namely, the algorithm has to treat the dependence on (i) the splitting or merging of collinear parton four-vectors and (ii) the addition of arbitrarily soft partons to the list of objects to be clustered.

Sequential-recombination algorithms belong to a class of infrared and collinear safe inclusive jet algorithms for hadron colliders, and iteratively combine the closest pairs of physics objects to obtain massive jets using as a measure of distance

$$d_{ij} = \min(p_{T,i}^{2p}, p_{T,j}^{2p}) \frac{\Delta R_{ij}^2}{R^2} \quad 3.4$$

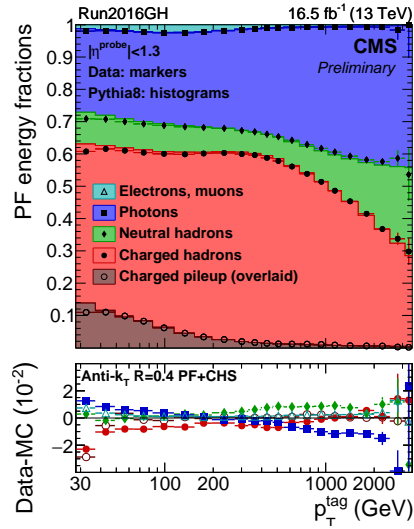
with  $\Delta R_{ij} = \sqrt{|y_i - y_j|^2 + |\phi_i - \phi_j|^2}$  and fixed distance parameter  $R$ ; a continuous parameter  $p$  sets the power of the  $p_T$  scale relative to the geometrical distance;  $i$  and  $j$  entities with rapidities  $y_i$  and  $y_j$ , and azimuthal angles  $\phi_i$  and  $\phi_j$ , respectively, are only combined as long as  $d_{ij} < p_{T,i}^{2p}$ ; otherwise, the jet is promoted to a final jet and ignored in subsequent steps. Considering  $p = -1$  yields the so-called “anti- $k_T$  clustering algorithm” that is infrared and collinear safe, yet resilient with respect to soft radiation concerning the jet boundaries [195].

In CMS, jets are reconstructed using the anti- $k_T$  clustering algorithm, as implemented in the FASTJET package [196], with a nominal distance parameter of 0.4, and  $i$  and  $j$  entities corresponding to different types of physics objects. We focus on three types of jets that are distinguished based on the nature of simulated and detector information (Fig. 1.7): simulated particle-level, calorimeter, and PF jets.

The simulated particle-level jets are built by applying the clustering procedure to all stable (lifetime  $c\tau > 1$  cm) particles excluding neutrinos. The exclusion of neutrinos is a convention adopted by CMS, facilitating the definition of the jet response in a way that is experimentally accessible and significantly reduces response differences between heavy-flavor (c, b) and light-quark (u, d, s) or gluon jets (see Section 3.3.5), caused by neutrinos produced in semileptonic decays of heavy-flavor hadrons<sup>3</sup>. Although the lifetime of heavy hadrons on average is shorter than  $c\tau = 1$  cm, they are considered, together with their decay products, as the particles for jet clustering. To prevent these generated hadrons from affecting the susceptibility of jets to additional underlying-event and pileup radiation [197] the modulus of the hadron four-momentum is set to a small number, retaining only the directional information, a procedure known as “ghost association.” Jets coming

<sup>3</sup>The neutrino fraction leads to an additional systematic uncertainty in the fragmentation of heavy hadrons relative to the original b quarks that is typically considered in, e.g., measurements of the inclusive b jet cross section or the top quark mass.

from pileup interactions are tentatively identified as those missing a geometrically match to a simulated particle-level jet.



**Figure 3.20:** Jet PF composition studied with QCD dijet events in the end of 2016 pp data taking (period “Run2016 GH”) for fully corrected jets using pp collisions [198]. Data are compared to QCD multijet simulation. All types of PF particles are considered: leptons, photons, neutral and charged hadrons. The latter are split into hadrons clustered into PF CHS jets (see Section 3.3.1) and energy removed by CHS before jet clustering, overlaid on top of the “PF+CHS” jet energy fraction for visualization.

The calorimeter (CALO) jets are reconstructed from energy deposits in the calorimeter towers alone. A calorimeter tower consists of one or more HCAL cells and the geometrically corresponding ECAL crystals. Calorimeter jets result from a relatively simplistic yet robust approach and were widely used in the early CMS analyses. With the improvement of the understanding of the detector and the commissioning of the reconstruction with data, the performance of the PF reconstruction has proven to be reliable, and hence the PF jets are used in the majority of recent analyses.

The PF jets are reconstructed by clustering the four-momentum vectors of PF candidates, i.e., muon, electron (“charged EM”), photon (“neutral EM”), charged and neutral hadron candidates. The PF jet momentum and spatial resolutions are greatly improved with respect to CALO jets, as the use of the tracking detectors and high granularity of ECAL improves the energy resolution through the independent measurements of charged hadrons and photons inside a jet, which together constitute more than  $\approx 85\%$  of the average jet energy (Fig. 3.20). In reconstructing the PF candidate four-momentum, photons are assumed massless and charged hadrons are assigned the charged pion mass [16].

### Jet energy correction strategy

Similar to all experimentally-reconstructed objects, jets need to be calibrated in order to have the correct energy scale (JES): this is the purpose of the jet energy corrections (JEC). The detailed understanding of both JES and the  $p_T$  resolution of the jets (JER) is of crucial importance for many physics analyses, and a leading component of the systematic

uncertainty. Jets are successively corrected for the offset energy coming from pileup, the detector response to hadrons, and residual differences between data and simulation as a function of their  $p_T$  and  $\eta$ . These corrections also depend on the jet flavor and the distance parameter, albeit the latter is fixed throughout all the analyses of the thesis. The consecutive JEC steps are illustrated in Fig. 3.21.

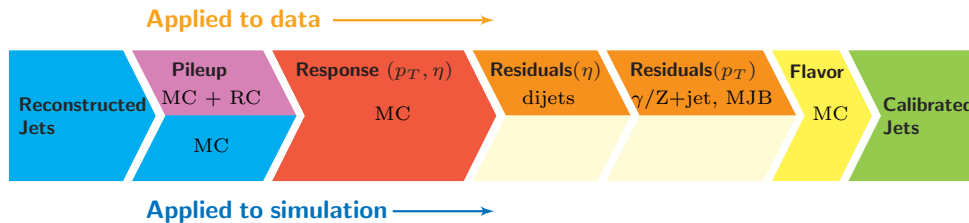


Figure 3.21: Consecutive stages of JEC, as applied in data and MC simulation [187].

Briefly, the jet response corrections are determined based on simulated QCD dijet flavor mixture as a function of jet  $p_T$  and  $\eta$ . Corrections for residual differences between data and detector simulation as a function of  $\eta$  are determined from a  $p_T$ -balance technique using QCD dijet events. The  $\eta$ -dependent corrections are estimated from QCD dijet events relative to a jet of similar  $p_T$  in the reference (barrel) region of  $|\eta| < 1.3$ . These corrections include a  $p_T$  dependence of the JES, and up to the limit of available dijet data. The absolute scale, together with its  $p_T$  dependence, is measured in  $\gamma$ +jet events, and it is typically cross-checked with Z+jet events. The uncertainty in the jet-flavor composition are also derived from MC simulations. The first step in the chain of the factorized JEC approach, i.e., the pileup offset correction, subtracts the energy not associated with the hard collision in the bunch crossing. The excess energy, which includes contributions from electronic noise in the calorimeters, “in-time”—extra interactions within the same bunch crossing—as well as “out-of-time” pileup—energy integrated from the preceding and subsequent bunch crossings—is not expected to be significant in the studied LHC conditions.

Simulated particle-level jets are geometrically matched to jets reconstructed in the detector such that their angular distance in  $(\eta, \phi)$  plane is less than the distance parameter  $R$  used in the clustering. The performance of the MC-based corrections that are used to remove the bulk of the nonuniformity in  $\eta$  and the nonlinearity in  $p_T$  are verified by calculating JES. The latter is defined as the mean value of a Gaussian fit to the distribution of  $p_T^{\text{reco}}/p_T^{\text{gen}}$ ; a mean value close to unity signifies a good agreement between the corrected reconstructed and particle-level jet  $p_T^{\text{reco}}$  and  $p_T^{\text{gen}}$ , respectively. This has been verified in the realm of Ref. [199], and additionally checked using b jets for the purpose of the current thesis, as shown in Fig. 3.22 as a function of  $\eta^{\text{gen}}$  in bins of  $p_T^{\text{gen}}$ . The JER variation is extracted by the  $\sigma$  parameter of the same Gaussian fit; the factorized dependence of JER on jet  $p_T$  and  $\eta$  is displayed in Fig. 3.22. The jet  $p_T$  resolutions have been independently determined with QCD dijet events, as discussed in Ref. [199].

To measure the response of a jet at any  $\eta$  relative to the response in the region  $|\eta| < 1.3$  the dijet  $p_T$ -balance technique is used. For the measurement of relative jet energy response, two leading in  $p_T$  jets are selected in the pseudorapidity interval of  $|\eta| < 3.0$ .

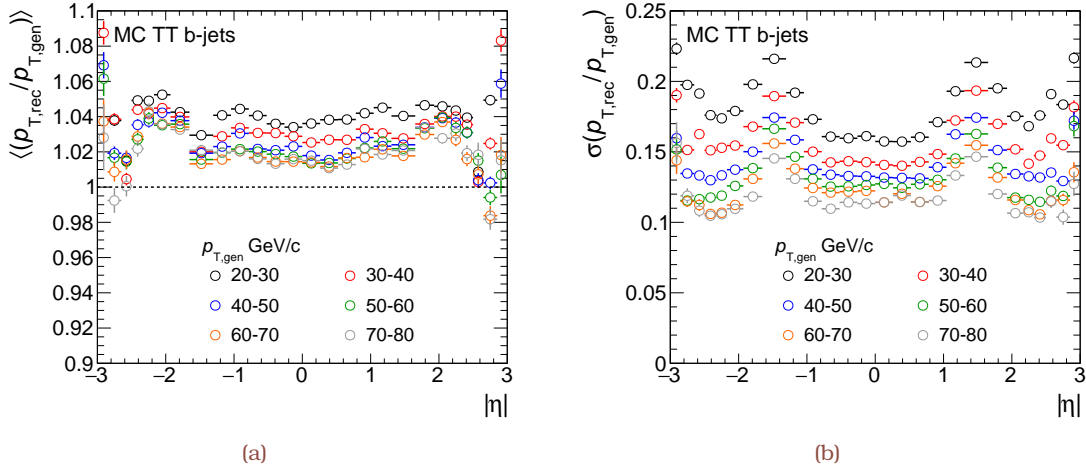


Figure 3.22: Jet energy scale (a) and resolution (b) in different  $p_T$  bins for jets identified as associated to the hadronization of a b quark in a simulated  $t\bar{t}$  event sample (Table 6.1) at  $\sqrt{s} = 5.02$  TeV [AN3].

The reference jet is required to be within  $|\eta^{\text{ref}}| < 1.3$ , with the probe jet at arbitrary  $\eta^{\text{probe}}$ . If both jets are within  $|\eta| < 1.3$ , one jet is randomly chosen as the reference jet, whereas, if both jets are not within  $|\eta| < 1.3$ , the event is discarded. A dijet balance quantity,  $B$ , is then defined as

$$B = \frac{p_T^{\text{probe}} - p_T^{\text{ref}}}{p_T^{\text{avg}}}, \quad 3.5$$

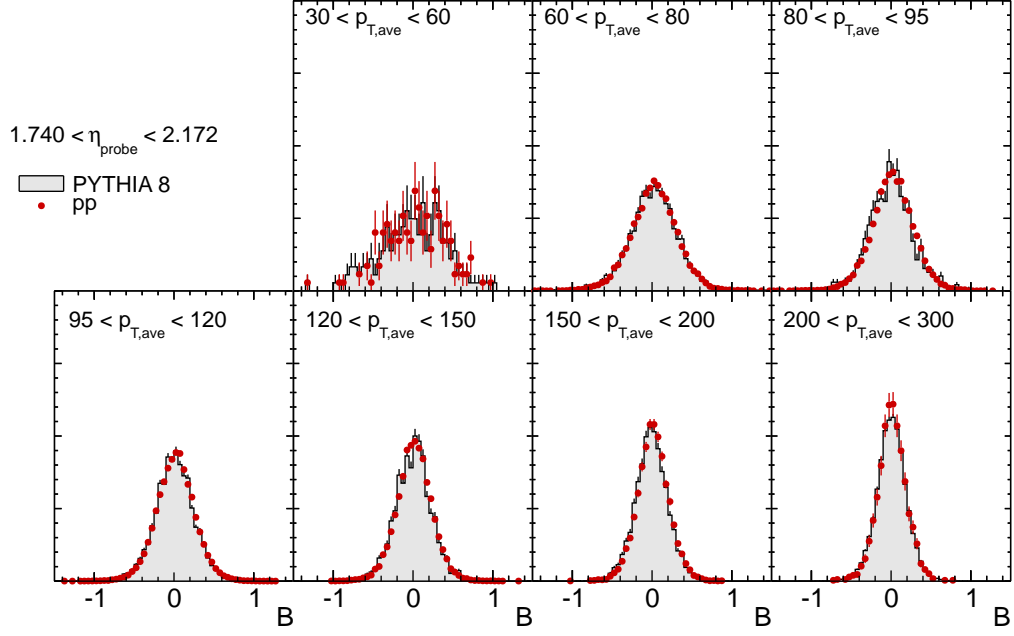
where  $p_T^{\text{avg}}$  is the average  $p_T$  of the two leading jets

$$p_T^{\text{avg}} = \frac{p_T^{\text{probe}} + p_T^{\text{ref}}}{2}. \quad 3.6$$

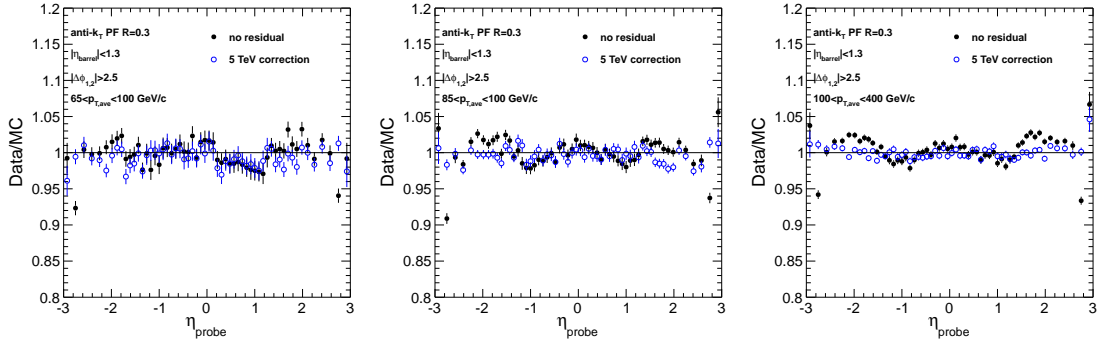
The correction, calculated from the response  $R = (2 - \langle B \rangle)/(2 + \langle B \rangle)$ , is derived as a function of  $p_T^{\text{ave}}$  in separate  $\eta$  bins, chosen according to calorimeter tower granularity and merging two neighboring cells to increase the event count. Figure 3.23 shows an example of dijet balance distributions from data and MC simulation within  $1.740 < \eta^{\text{probe}} < 2.172$  for a selection of  $p_T^{\text{avg}}$ . Using balance distributions the response in data and MC are compared, and after data are corrected to match simulation by a multiplicative jet by jet factor of  $R_{\text{MC}}/R_{\text{Data}}$ , the closure is examined as shown in Fig. 3.24.

The absolute JES is verified using  $\gamma$ +jet events; a photon in the barrel region  $|\eta| < 1.479$  and  $p_T^\gamma > 40$  GeV is selected. The jets used in the  $\gamma$ +jets sample are required to lie in the pseudorapidity interval  $|\eta^{\text{probe}}| < 1.3$ . The leading away side jet with  $p_T^{\text{probe}} > 10$  GeV is selected as the probe jet. The  $\gamma$ +jets sample is dominated by QCD multijet background, in which jets can disguise a photon. To suppress this type of background the photon candidate is required to be isolated based on stringent selection criteria [200]. The average value of the transverse momentum ratio  $p_T^{\text{probe}}/p_T^\gamma$  distribution in a given  $p_T^\gamma$  bin, is used to determine the absolute response  $R_{\text{abs}}$ , that is shown in Fig. 3.25.

The fractions of the jet energy carried by certain types of PF candidates, along with

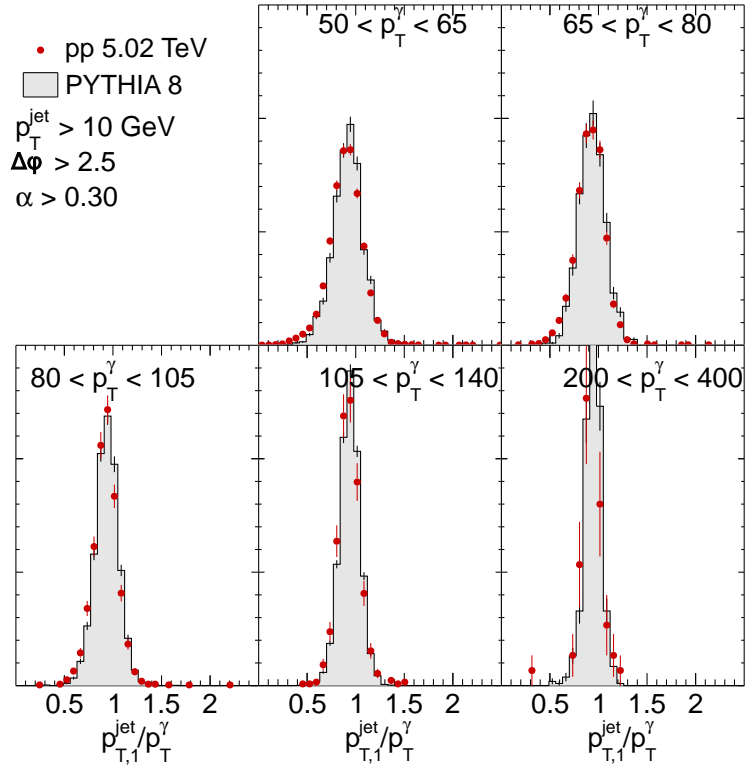


**Figure 3.23:** Balance distributions in PYTHIA simulation and data are shown for the  $1.749 < \eta < 2.172$  range in bins of  $p_T^{\text{ave}}$  for PF jets. The variables  $a = p_T^{\text{third}}/p_T^{\text{avg}}$  and  $|\Delta\phi_{jj}| = \phi^{\text{probe}} - \phi^{\text{ref}}$  are used to suppress three-jet events in the illustration [AN8].



**Figure 3.24:** Response ratios in PYTHIA simulation and data as a function of  $\eta^{\text{probe}}$  for PF jets found in  $|\eta| < 3.0$ , before and after the residual correction to the data is applied to account for the discrepancy between data and simulation in the jet response [AN8]. In pp collisions the results from dijet balance technique is cross-checked with the  $p_T^{\text{miss}}$  projection fraction (MPF) method as described in Ref. [187].

the number of PF candidates, clustered into a jet are used to discriminate between noise jets and physical jets. The jet energy fraction and multiplicity variables are sensitive to different sources of noise from ECAL and HCAL. Table 3.3 presents these PF jet identification (ID) criteria. Three PF jet ID working points (WPs) are defined: “loose,” “tight,” and “tight lepton veto.” The loose and tight WPs are designed to remove jets originating from calorimetric noise, while the “tight lepton veto” WP additionally rejects potential background from misreconstructed electron or muon candidates, effectively resolving also the ambiguity between isolated lepton candidates and jets reconstructed from single lepton



**Figure 3.25:** Response distributions for different  $p_T^\gamma$  ranges in PYTHIA simulation for PF jets [AN8]. To reject  $\gamma \geq 2$  jets events the variables  $a^\gamma = p_T^{\text{subleading}}/p_T^\gamma$  and  $\Delta\phi_{j\gamma}$  are used in the illustration. A cross-check of the absolute response obtained in MC is performed using  $Z(ee, \mu\mu)$ +jet events with the same requirements on  $\Delta\phi$  and  $a$  as in the  $\gamma$ +jet final state. The average ratio of the response in simulation and data is found to be close to unity in both balance studies.

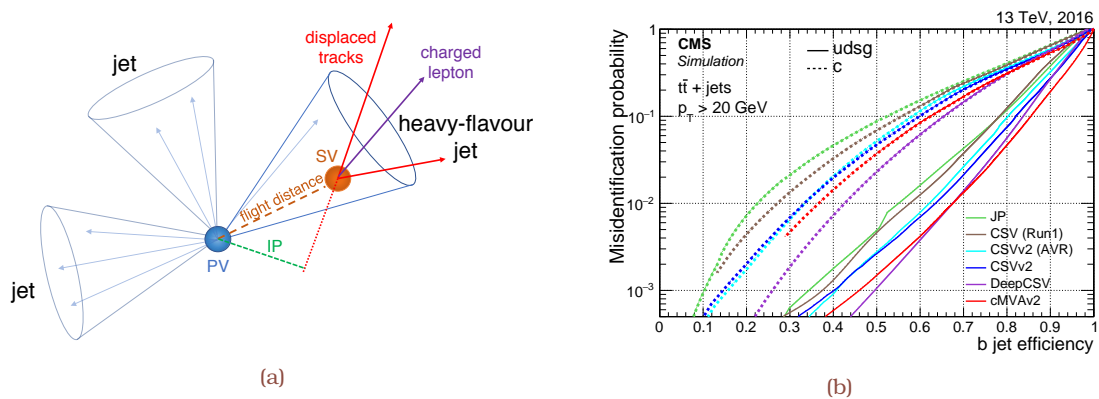
candidates.

**Table 3.3:** The three PF jet ID criteria; the “charged” variables extend up to about  $|\eta| < 2.4$  since there is no tracker coverage outside of this region, whereas the “neutral” variables extend to the  $\eta$  region up to  $|\eta| < 5.2$  [191].

Jet variables	Pseudorapidity range	Loose	Tight	Tight lepton veto
Charged hadron fraction	$ \eta  < 2.4$			$>0.0$
Charged multiplicity	$ \eta  < 2.4$			$>0$
Charged EM fraction	$ \eta  < 2.4$	$<0.99$	$<0.99$	$<0.9$
Muon fraction	$ \eta  < 2.4$			$<0.8$
Neutral hadron fraction	$ \eta  < 2.7$	$<0.99$	$<0.9$	$<0.9$
Neutral EM fraction	$ \eta  < 2.7$	$<0.99$	$<0.9$	$<0.9$
Neutral multiplicity	$2.7 <  \eta  < 5.2$			$<0.9$
	$2.7 <  \eta  < 3.0$			$>2$
	$3.0 <  \eta  < 5.2$			$>10$

### 3.3.5 Identification of heavy-flavor jets

Heavy-flavor jet identification techniques exploit the properties of hadrons clustered into the jet to discriminate between those originating from  $b$  quarks against  $c$  or light-flavor quarks and gluons. Owing to the lifetime of  $b$  hadrons, typical displacements of a few mm to cm—depending on their momentum—are expected, thus giving rise to displaced tracks out of which a secondary vertex (SV) can be reconstructed (Fig. 3.26a). In addition the decay products from  $b$  quarks have larger average  $p_T$  relative ( $p_{T,rel}$ ) to the jet axis—spanned by the primary vertex and the direction of the jet momentum—than the other jet constituents because of their larger mass and harder fragmentation compared to the light-flavor quarks and massless gluons. Apart from the properties of the reconstructed SV and displaced tracks, the presence of charged leptons is also exploited for heavy-flavor jet identification.



**Figure 3.26:** (a) Illustration of a heavy-flavor jet with a secondary vertex (SV) from the decay of a  $b$  hadron [201]. The displacement of tracks, including possibly an identified lepton, with respect to the primary interaction vertex is characterized by their impact parameter (IP). The IP can receive both positive and negative values, with a positive sign indicating that the track is produced “upstream,” namely, the angle between the impact parameter vector and the jet axis is less than  $\pi/2$ . (b) Misidentification probability for  $c$  and light-flavor jets against  $b$  jet identification efficiency [201].

One of the algorithms, implemented by CMS to select jets originating from  $b$  quarks and used in the current thesis, combines the information of displaced tracks with the information on SV associated with the jet using a multivariate (MVA) technique. The combined secondary vertex algorithm [201] is a variant of the technique already developed during Run 1 [202], mainly after adopting a different reconstruction algorithm for SV [203] and extending the list of input variables (Table 3.4). The baseline selection consists of tracks having an angular distance with respect to the jet axis of  $\Delta R < 0.3$ ; any combination of two tracks compatible with the mass of the  $K_S^0$  meson in a window of 30 MeV is rejected; jets that have neither a selected track nor a SV are assigned the lowest MVA score.

The efficiency (misidentification probability) to correctly (wrongly) label a jet with flavor  $f$  is defined as the number of  $f$  jets fulfilling the “tagging” requirement divided by the total number of  $f$  jets. Figures 3.27 and 3.28 show the  $b$  tagging efficiency as a function of the misidentification probability against either  $c$  or light-flavor jets in simulated  $t\bar{t}$  and QCD

**Table 3.4:** Input variables used for the Run 1 version of the CSV [202] compared to the CSVv2 [201] algorithm. To avoid discrimination between jet flavors caused by different jet  $p_T$  and  $\eta$  distributions all variables are reweighted to obtain the same spectrum for all jet flavors in the training sample, which consists of inclusive QCD multijet events in three independent vertex categories, depending on the presence and association of SV to selected jets or tracks. The symbol “ $\checkmark$ ” (“ $\times$ ”) means that the variable is (not) used in the respective version of the algorithm. The order of variables is arbitrary.

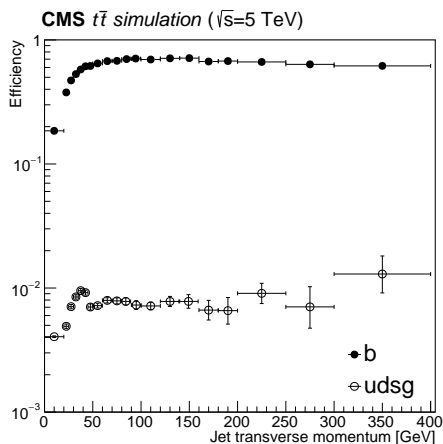
Input variable	CSV (Run 1)	CSVv2
SV two-dimensional flight of distance significance	$\checkmark$	$\checkmark$
SV multiplicity	$\times$	$\checkmark$
Track $\eta_{rel}$	$\checkmark$	$\checkmark$
Corrected SV mass	$\checkmark$	$\checkmark$
Number of tracks originating from SV	$\checkmark$	$\checkmark$
SV energy ratio	$\checkmark$	$\checkmark$
$\Delta R(SV, jet)$	$\times$	$\checkmark$
Three-dimensional IP significance of the first four tracks	$\checkmark$	$\checkmark$
Track $p_{T,rel}$	$\times$	$\checkmark$
$\Delta R(track, jet)$	$\times$	$\checkmark$
Track $p_{T,rel}$ ratio	$\times$	$\checkmark$
Track distance	$\times$	$\checkmark$
Track decay length	$\times$	$\checkmark$
Vector sum track $E_T$ ratio	$\times$	$\checkmark$
$\Delta R(summed tracks, jet)$	$\times$	$\checkmark$
First track two-dimensional IP significance above c quark mass threshold	$\times$	$\checkmark$
Number of selected tracks	$\times$	$\checkmark$
Jet $p_T$	$\times$	$\checkmark$
Jet $\eta$	$\times$	$\checkmark$

multijet events requiring jets with  $p_T > 20 \text{ GeV}$  and  $|\eta| < 2.4$  in pp and pPb collisions, respectively. For the latter the b tagging efficiency is integrated over the jet  $p_T$  and  $\eta$  distributions in the  $t\bar{t}$  and QCD multijet samples. Three standard working points are defined, namely, “loose,” “medium,” and “tight.” These correspond to thresholds on the discriminator after which the misidentification probability is around 10%, 1%, and 0.1%, respectively, for light-flavor jets, and using pp collisions at 13 TeV (Fig. 3.26b). The efficiency for correctly identifying b jets in simulated  $t\bar{t}$  events for each of the three working points of the various taggers is also summarized in Table 3.5.

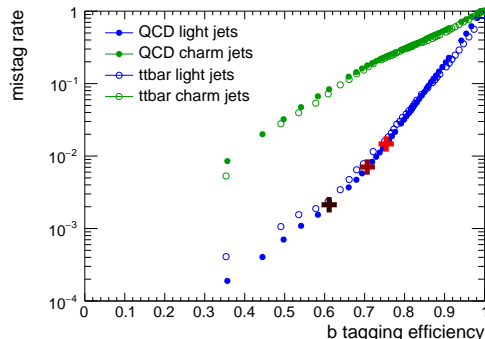
The tagging efficiency depends on the jet  $p_T$  and  $\eta$ , whereas the dependence with respect to the number of pileup interactions is expected to be small for the considered conditions. The efficiency for correctly identifying b jets is maximal for jets with  $p_T \approx 100 \text{ GeV}$  and decreases at low- and high- $p_T$  values, respectively, as indeed illustrated in Fig. 3.27. The lower efficiency at low jet  $p_T$  is due to the larger uncertainty on the track impact parameter resolution. At high jet  $p_T$ , there are two main effects. First, the misidentification probability for light-flavor jets increases, as can be also seen in Fig. 3.27 within the uncertainty, because of the larger number of tracks present in the jet. Second, at higher jet transverse momenta, jets are more collimated, resulting in merged hits in the innermost layers of the tracker. This effect impacts the track reconstruction efficiency and hence also the b jet identification efficiency. Owing to the higher track reconstruction



efficiency and the better resolution of the track parameters at small  $|\eta|$  values (Fig. 3.9a), the algorithm is more efficient in identifying b jets in the barrel region of the silicon tracker.



**Figure 3.27:** Identification efficiency and misidentification probability for b and c and light-flavor jets, respectively, as a function of the jet  $p_T$  in pp collisions at  $\sqrt{s} = 5.02$  TeV determined from  $t\bar{t}$  simulation [AN3, AN1, AN4].



**Figure 3.28:** Misidentification probability for c and light-flavor jets against b jet identification efficiency in pPb collisions at  $\sqrt{s_{NN}} = 8.16$  TeV separately determined from  $t\bar{t}$  and QCD multijet simulations [AN2].

**Table 3.5:** Main algorithms, operating (“working”) points, and corresponding efficiency for b jets with  $p_T > 20$  GeV in simulated  $t\bar{t}$  events using pp collisions at  $\sqrt{s} = 13$  TeV. The numbers are for illustrative purpose since the b jet identification efficiency is integrated over the  $p_T$  and  $\eta$  distributions of jets [201]. For comparison, the most efficient algorithms in Run 1 and 2, the CSV and DeepCSV taggers, respectively, achieve b jet tagging efficiencies of about 85 (70)% and 84 (68)% for a light-parton misidentification probability of approximately 10 (1.5)% and 11 (1.1)%, using the same loose (medium) event selection [202].

Tagger	Working point	$\varepsilon_b$ (%)	$\varepsilon_c$ (%)	$\varepsilon_{udsg}$ (%)
Combined secondary vertex (CSVv2)	Loose	81	37	8.9
	Medim	63	12	0.9
	Tight	41	2.2	0.1
Deep combined secondary vertex (DeepCSV)	Loose	84	41	11
	Medium	68	12	1.1
	Tight	50	2.4	0.1

### 3.3.6 Reconstruction and calibration of $\vec{p}_T^{\text{miss}}$

The missing transverse momentum vector is defined as the negative vector sum of the momenta of all observed final-state particles in any event, projected onto the plane perpendicular to the direction of the proton beams, i.e.,  $\vec{p}_T^{\text{miss}} \equiv -\sum \vec{p}_T$ . Its magnitude is referred to as  $p_T^{\text{miss}}$ . By momentum conservation,  $\vec{p}_T^{\text{miss}}$  is also equal to the total transverse

momentum of all unobserved particles, such as neutrinos. In CMS several distinct and complementary algorithms to reconstruct  $\vec{p}_T^{\text{miss}}$  have been developed [189] (and references therein). The  $\vec{p}_T^{\text{miss}}$  reconstructed with the PF technique is used primarily. It is defined as the negative vector sum over the transverse momenta of all PF particles. A less commonly used method is the calorimetric  $\vec{p}_T^{\text{miss}}$  that is calculated using the energies contained in calorimeter towers and their directions relative to the center of the detector.

Different effects could lead to biased  $p_T^{\text{miss}}$  values like minimum energy thresholds in the calorimeters, inefficiencies in the tracker, nonlinearity of the response of the calorimeters for hadronic particles, etc. This bias is significantly reduced by correcting the  $p_T$  of jets to the particle-level  $p_T$  using jet energy corrections (see Section 3.3.4),

$$\vec{p}_T^{\text{miss,corr}} = \vec{p}_T^{\text{miss}} - \vec{\Delta}_{\text{jets}} = \vec{p}_T^{\text{miss}} - \sum_{\text{jets}} (\vec{p}_{T,\text{jet}}^{\text{corr}} - \vec{p}_{T,\text{jet}}), \quad 3.7$$

where the superscript “corr” refers to the corrected values. The sum extends over all jets with an electromagnetic energy fraction below 0.9 (Table 3.3) and a corrected  $p_T > 15$  GeV, and it is colloquially referred to as “type-I” correction. In addition, if a muon reconstructed using the outer tracking system overlaps with a jet, its four-momentum is subtracted from the four-momentum of the jet. In this thesis exclusively the corrected  $\vec{p}_T^{\text{miss}}$  is used, and therefore the prefix “corr” is typically omitted to improve clarity, unless specified otherwise.

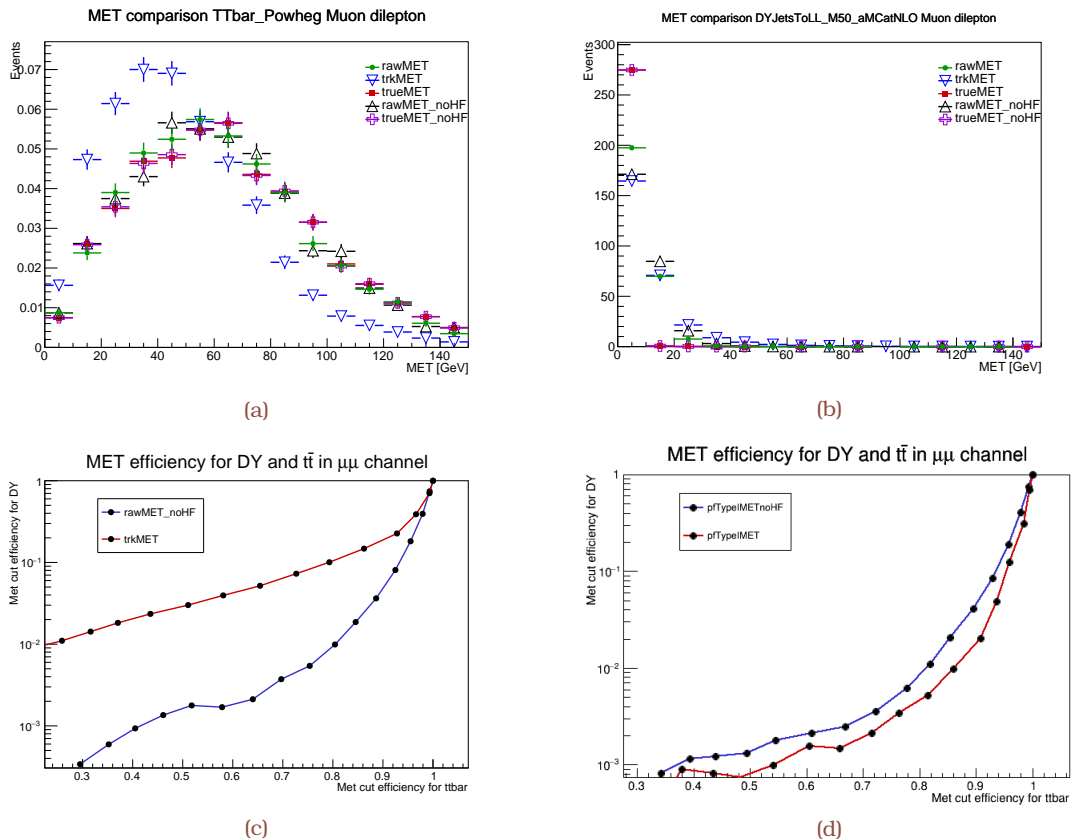
**Table 3.6:** Functional forms of the resolutions in the  $p_T$  measurement for each PF-candidate class contributing to the unclustered energy, whose uncertainty is the dominant source in the estimation of  $p_T^{\text{miss}}$ .

PF-candidate type	Resolution function
Charged hadron	$(0.00009 \times p_T)^2 + (0.0085 / \sqrt{\sin \times (2 \times \arctan(e^{-\eta}))})^2$
Charged and neutral EM (ECAL)	$(0.03/p_T) \oplus 0.001$
Neutral hadron (HCAL barrel)	$\min(0.25, (0.8/p_T) \oplus 0.05)$
Neutral hadron (HCAL endcap)	$\min(0.30, (1/p_T) \oplus 0.04)$
Neutral hadron (HF)	$(1/p_T) \oplus 0.05$

The uncertainty in  $p_T^{\text{miss}}$  has strong dependence on the exact topology of the final state under study. Since  $p_T^{\text{miss}}$  relies on the accurate measurement of the reconstructed physics objects, namely, muons, electrons, photons,  $\tau$  leptons, jets, and unclustered energy—defined as the contribution from PF candidates not clustered into any of the previous physics objects—the associated uncertainty is estimated based on a factorized approach. The contribution from each physics object is independently recalculated within its scale and resolution boundaries, meaning the uncertainty in  $p_T^{\text{miss}}$  is evaluated by comparing the recalculated  $p_T^{\text{miss}}$  to its nominal value.

The uncertainty due to unclustered energy is evaluated based on the momentum resolution of each PF candidate. More specifically, the  $p_T$  measurement of PF charged hadrons is dominated by the tracker resolution, while for PF neutral hadrons, the  $p_T$  resolution is dominated by the resolution of HCAL. The ECAL resolution dominates the PF photon  $p_T$  mea-

surement, whereas HF intrinsic resolution dominates the measurement for PF particles in that region. The largest contributions to the unclustered energy uncertainty are due to the PF neutral hadrons and candidates in HF. Table 3.6 summarizes the functional forms of the resolutions of the individual PF-candidate classes contributing to the unclustered energy.

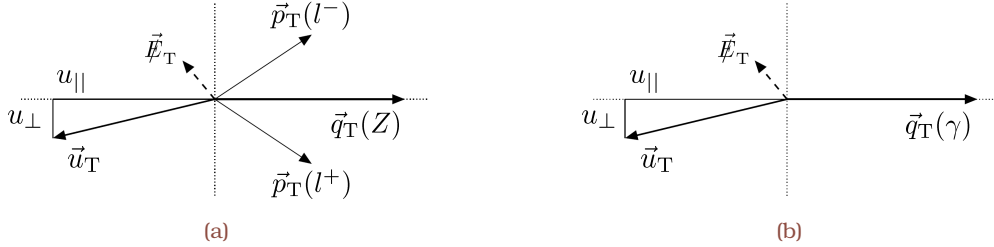


**Figure 3.29:** Generated (“true”)  $p_T^{\text{miss}}$  distributions are compared to  $p_T^{\text{miss}}$  reconstructed considering all PF candidates (“rawMET,” “typeIMET”), only PF candidates from the tracker (“trkMET”), or excluding those from HF (“noHF”) using  $\mu^\pm\mu^\mp$  events in simulated  $t\bar{t}$  (a) and  $Z/\gamma^*$  (b) event samples (Table 5.1) at  $\sqrt{s} = 5.02$  TeV. The fraction of events satisfying a  $t\bar{t}$ -enriched selection (see Section 5.3) and sequentially applying an increasing requirement on  $p_T^{\text{miss}}$  (from left to right and bottom to top for the horizontal and vertical axis, respectively) is shown in (c) and (d) [AN3, AN4].

A typical requirement is imposed using the  $p_T^{\text{miss}}$  magnitude to further reject  $Z/\gamma^*$  background events in the  $t\bar{t}$  dilepton final state. To that end, different definitions of  $p_T^{\text{miss}}$  have been studied, which are displayed in Figs. 3.29a and 3.29b, including the generated, raw and tracking  $p_T^{\text{miss}}$  distributions, with and without the HF contribution separately in  $t\bar{t}$  signal and  $Z/\gamma^*$  background for the event selection of Ref. [TH4] (see Section 5.3). Figures 3.29c and 3.29d show the background rejection against the  $t\bar{t}$  signal efficiency for the different  $p_T^{\text{miss}}$  definitions. The chosen working point is  $p_T^{\text{miss}} > 35$  GeV that corresponds to a signal efficiency of  $\sim 80\%$  with more than 99% of the  $Z/\gamma^*$  contribution being rejected.

The scale and resolution of  $p_T^{\text{miss}}$  can be studied in event samples with an identified Z boson or an isolated photon. Such events have no genuine  $\vec{p}_T^{\text{miss}}$ , and hence the per-

formance can be measured comparing the momenta of the vector boson to that of the so-called “hadronic recoil” system. The hadronic recoil system is defined as the vector  $p_T$  sum of all PF candidates excluding either the vector boson itself or its decay products in the case of the Z boson decay. In Fig. 3.30 the kinematic representation of the  $p_T$  of the vector boson and the hadronic recoil,  $\vec{q}_T$  and  $\vec{u}_T$ , respectively, is given. Momentum conservation in the transverse plane trivially imposes  $\vec{q}_T + \vec{u}_T + \vec{p}_T^{\text{miss}} = \vec{0}$ . Thus this type of events provide a unique event axis and a precise momentum scale. The components of the hadronic recoil parallel and perpendicular to the boson axis are denoted by  $u_{\parallel}$  and  $u_{\perp}$ , respectively. The  $u_{\parallel}$  and  $u_{\perp}$  are used to study the  $p_T^{\text{miss}}$  response and resolution. Specifically, the mean of the  $u_{\parallel}$  (or  $u_{\parallel} + q_T$ ) distribution is used to estimate the  $p_T^{\text{miss}}$  response, whereas the RMS of the  $u_{\parallel}$  (or  $u_{\parallel} + q_T$ ) and  $u_{\perp}$  distributions are used to estimate the resolution of  $u_{\parallel}$  and  $u_{\perp}$ , respectively. Corrections can be then extracted from  $Z/\gamma^*$ - or photon-enriched samples in bins of  $\vec{q}_T$  and applied to the recoil in MC simulation to match the scale and resolution measured in data.



**Figure 3.30:** Illustration of the  $Z/\gamma^*$  (a) and photon (b) event kinematics in the transverse plane. The vector  $\vec{u}_T$  denotes the vector  $p_T$  sum of all particles reconstructed in the event excluding the two leptons from the  $Z/\gamma^*$  decay or the photon [189].

Figure 3.31 shows the  $p_T^{\text{miss}}$  distributions for W boson muon [204] and  $t\bar{t} \ell$ +jets [TH1] decay events at  $\sqrt{s_{\text{NN}}} = 8.16$  TeV. The distributions of the hadronic recoil components that are parallel and perpendicular to the Z boson transverse momentum are fitted in simulation and data using a weighted sum of two Gaussian functions. The mean and resolution values extracted from the recoil fits are used to scale the simulated hadronic recoil distributions to match the performance measured in data. The corrected  $p_T^{\text{miss}}$  distribution is then derived in the MC event samples as the vector sum of  $\vec{u}_T^{\text{corr}}$  and the  $\vec{q}_T$  of the reconstructed leptons from the decay of Z bosons. Owing to the similarity of the Z and W boson production processes, and their similar masses, it is assumed that the recoil distributions do not depend on the particle type.

Last but not least anomalous high- $p_T^{\text{miss}}$  events and/or  $\phi$  modulations of physics objects can arise from reconstruction failures or malfunctioning detector. The anomalous events with artificially large  $p_T^{\text{miss}}$  are typically found to be mostly due to electronic noise in the calorimeters, e.g., showers in ECAL of non collision origin or noise in the HCAL electronics, while an excess of events with  $\phi \approx 0$  or  $\pi$  can be attributed to machine-induced backgrounds, especially the production of muons when beam protons undergo collisions upstream of the detector (“beam halo”). The angular distribution of beam halo events is dictated by the shape of the LHC tunnel and the beamline elements [205]. Event filters are

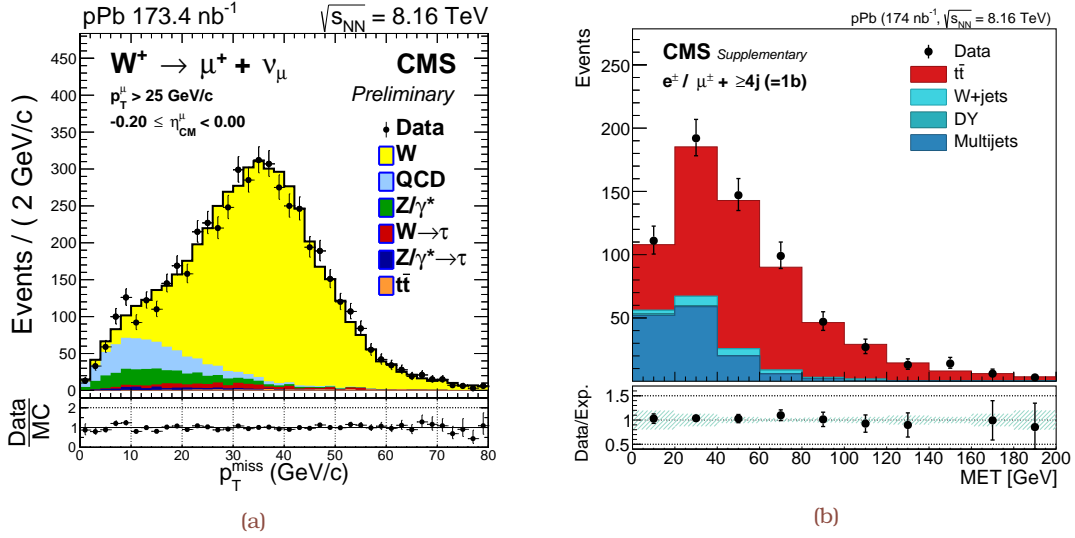


Figure 3.31: The  $p_T^{\text{miss}}$  reconstructed in (a) W boson [204] or (b) top quark [TH1] leptonic decays at  $\sqrt{s_{\text{NN}}} = 8.16$  TeV. They are split into contributions from  $t\bar{t}$ ,  $Z/\gamma^*$ ,  $W$ +jets, and QCD multijet processes. Error bars represent the statistical uncertainty. The lower panels display the data divided either by the result of the unbinned fit to the data (a) or the expectation from the simulation (b). For (a) the  $-0.2 < \eta_{\text{CM}}^\mu < 0.0$  range is defined such that the proton is moving towards positive pseudorapidity.

designed to identify more than 85–90% of the spurious events with a mistag rate of less than 0.1%. The event filtering algorithms can be further combined with the jet identification requirements (Table 3.3), e.g., by imposing the neutral hadron energy fraction of a jet to be less than 0.9, more than 99% of the noise jets are rejected (independent of jet  $p_T$ ) with a negligible mistag rate.

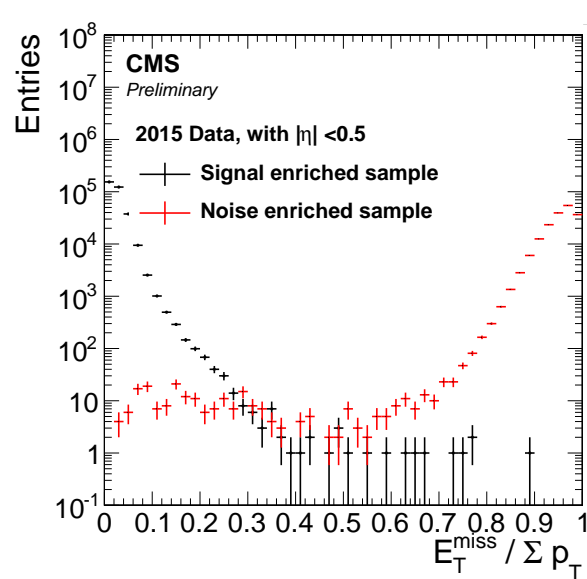
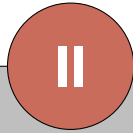


Figure 3.32: The distributions of the  $p_T^{\text{miss}}$  over the scalar  $p_T$  sum ( $\sum p_T$ ) ratio in signal- and noise-enriched events using back-to-back dijet (black) and a minimum-bias selection without applying any PF jet identification requirements (red), respectively [191].





Part B





## Contents

<b>4.1 The concept of luminosity</b> . . . . .	<b>111</b>
4.1.1 Interaction rate determination . . . . .	111
4.1.2 Formalism of absolute luminosity from separating the beams . . . . .	113
<b>4.2 Offline luminosity measurement in CMS</b> . . . . .	<b>116</b>
4.2.1 Experimental setup . . . . .	116
4.2.2 The vdM scan protocols . . . . .	117
4.2.3 The pixel-cluster counting method . . . . .	120
4.2.4 Extraction of the visible cross section . . . . .	121
<b>4.3 Calibration corrections</b> . . . . .	<b>122</b>
4.3.1 Concept and formalism of constant separation LSC scan . . . . .	122
4.3.2 Orbit drift . . . . .	127
4.3.3 Impact of nonfactorizable beam shape . . . . .	128
4.3.4 Beam-beam effects . . . . .	130
4.3.5 Bunch current corrections and normalization . . . . .	131
<b>4.4 Cross-detector stability</b> . . . . .	<b>133</b>
<b>4.5 Conclusions</b> . . . . .	<b>135</b>

A system consisting of five detectors to monitor and measure the luminosity delivered by LHC is currently in use at the CMS experiment. On the one hand, the BCM1F, HF, and PLT detectors are characterized by an independent high-rate data acquisition system, and hence serve as excellent real time monitoring systems for luminosity. On the other hand, the silicon pixel detector and DT feature very low occupancy and good stability over time. Absolute calibrations of the luminometers are established by performing van der Meer (vdM) scans typically with dedicated LHC machine setups. Scanning the two beams through one another in the transverse plane of the detector the vdM technique allows to measure the luminosity per colliding bunch pair directly from the machine parameters.

The calibration of the integrated luminosity delivered to the CMS experiment during the proton-proton (pp), and the lead-proton (Pbp) and proton-lead (pPb) periods at  $\sqrt{s} = 5.02$  and  $\sqrt{s_{\text{NN}}} = 8.16$  TeV, respectively, is presented in this Chapter. Three different subdetectors are used: the HF calorimeter, PLT, and the silicon tracker. Visible cross sections are obtained using the vdM procedure for measuring the luminometer rate as a function of the beam separation. After taking into account the effects due to horizontal-vertical beam correlations, bunch-to-bunch and scan-to-scan variations, spurious charges, adjustments of the length scale measurements provided by LHC magnet currents, and beam-beam effects, an overall uncertainty of 2.3, 3.2 and 3.7% is assessed in pp, Pbp and pPb periods, respectively, while the total “Pbp+pPb” uncertainty is found to be 3.5%, combining the visible cross sections in the two separate periods. Time stability of these calibrations is considered and included in the final systematic uncertainty.

The material in the following Chapter, as documented in [Scientific output](#) and [Internal notes](#), relies almost exclusively on an original contribution. Each study is subjected to an exhaustive internal review from early analysis to publication, including several fixed waypoints that had to be additionally met.

## 4.1 The concept of luminosity

Luminosity has been used in astronomy (cosmology) indicating the amount of electromagnetic energy an astronomical (cosmological) object radiates per unit time. The term has been introduced in particle physics in the early 1960's in the context of the first matter-antimatter collider, Anello di Accumulazione, at the Frascati laboratory accelerating electron ( $e^-$ ) against positron ( $e^+$ ) beams at  $\sqrt{s} = 250$  MeV [206]. The analogy between the definitions in accelerator physics and astronomy, that is driven by a characteristic “source factor” in both cases, rendered luminosity as the proportionality coefficient between the event accumulation rate in a particle collider and the cross section. Luminosity thus quantifies the potential of the collider for delivering a statistically significant sample of any class of events.

Storage-ring beam dynamics suggest operating particle colliders in “bunched mode,” i.e., each of the two beams consists of a string of bunches typically unevenly distributed around the collider ring and numbering a few ten to a few thousand (see Section 2.1.4). To determine the cross section of any given subatomic process at high energy colliding-beam experiments a measurement of the colliding-bunch luminosity must hence be performed. The single-bunch instantaneous luminosity  $\mathcal{L}_b$ , i.e., produced by a single pair of colliding bunches, can be written as

$$\mathcal{L}_b = \frac{R}{\sigma_{\text{ref}}} , \quad 4.1$$

where the interaction rate  $R = \mu f_r$  for any reference process is linearly dependent on the average number of interactions per bunch crossing ( $\mu$ ) and the bunch revolution frequency ( $f_r$ ). In principle, the reference process can be arbitrarily selected, and referring to inelastic collisions as a typical example, the total instantaneous luminosity for inelastic interactions known with an absolute scale  $\sigma_{\text{inel}}$  is given by

$$\mathcal{L} = \sum_{b=1}^{n_b} \mathcal{L}_b = n_b \frac{\langle \mu \rangle f_r}{\sigma_{\text{inel}}} . \quad 4.2$$

In Eq. (4.2) the sum runs over the bunch pairs  $n_b$  colliding at the interaction point (IP), and the mean bunch luminosity is regulated by the bunch-averaged pile-up parameter  $\langle \mu \rangle$ . Therefore the instantaneous luminosity can be determined using any per-bunch-granularity method that measures the ratio  $\mu/\sigma_{\text{ref}}$ , or respectively  $\langle \mu \rangle/\sigma_{\text{ref}}$  for the total instantaneous luminosity. Although luminosity is a macroscopic indicator of the global performance of a collider, the observed bunch-to-bunch intensity and emittance variations in hadron colliders result in a large spread in  $\mathcal{L}_b$ , hence make impractical any bunch-averaged luminosity measurements.

### 4.1.1 Interaction rate determination

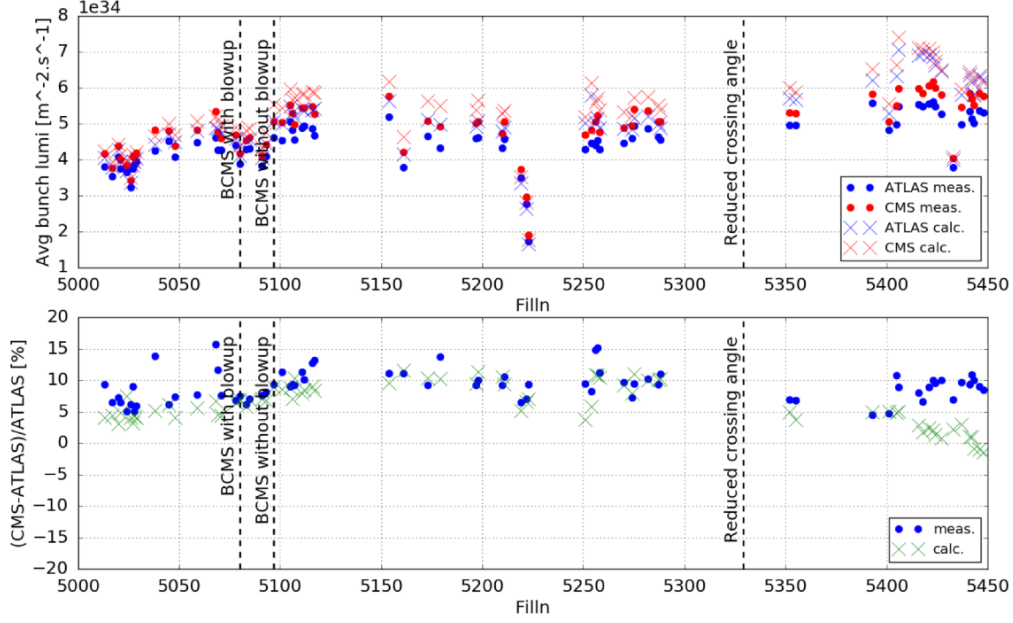
Methods for absolute luminosity determination can be classified as being either direct or indirect. For instance, indirect methods make use of the optical theorem in a simultaneous measurement of the elastic and total cross sections [207] or perform a

comparison with QED processes for which the absolute cross section is well known from theory [208, 209]. Direct methods derive the luminosity from the measurement of the colliding-beam parameters. The analysis described in this thesis relies on two direct methods to determine the absolute luminosity calibration, i.e., the “van der Meer” (vdM) and the “beam-imaging” (BI) scan methods. With the notable exception of the “first” and “second” generation experiments at CERN Intersecting Storage Rings (ISR) facility—including the Louvain-Northwester group [210]—the achieved precision in the luminosity determination at hadron colliders typically ranges from 1 to 15% [73]. The 1% “precision frontier” is not uniquely linked to a fundamental limitation, rather it stems from a complex mix of sources of systematic uncertainty.

The vdM technique exploits the ability to control the beam separation in both transverse coordinates with high precision (see Section 2.1.4), and hence to scan the overlap integral of the colliding beams at different relative beam positions, while measuring the interaction rate. This method—first applied at the CERN ISR [211]—has been widely used by all major LHC experiments during Run 1 [65, 136, 212–214] and more recently at 13 TeV [215–217] [TH3, TH14]. The BI method [218] is based on reconstructing primary vertices from interactions between one beam fixed in the rest-frame of the detector and the other one consecutively moving in  $x$  and  $y$ . The shapes obtained by the distribution of vertices can be analytically convolved with vertex position resolution models, and can be used to determine the overlap integral accounting for genuine nonfactorizabilities.

In both methods, data recorded by CMS are used in conjunction with input from the LHC beam instrumentation. While beam-current monitoring currently achieves sub-percent level precision, single-beam profile measurements, e.g. Ref. [219], are challenging because of instrumental resolution and limitations on optical models of the collider lattice.

Lately, LHC measurements, that resorted to synchrotron-light telescopes as transverse and longitudinal beam-profile monitors, achieved a remarkably good ( $\mathcal{O}(5\%)$ ) agreement with the absolute luminosity scales of both ATLAS and CMS; discrepancies though appeared indicating the impact driven by instrumental calibration. Figure 4.1 shows an example comparison between the peak luminosity values calculated (crosses) using the measured bunch parameters, i.e., transverse emittance, bunch intensity and length, and the average measured peak luminosity provided by ATLAS (blue circles) and CMS (red circles) during 2016 [220]. The bottom plot shows the luminosity imbalance between the two experiments using the same marker convention. During the first part of the year, before the transition to the compression merging and splitting (BCMS) production scheme, very good agreement between the calculated and measured peak luminosity is observed. After the transition to BCMS, even though the calculated and measured imbalance agrees well, the absolute values start to diverge. In the third part, after the crossing angle reduction, a disagreement is observed both in absolute values and in imbalance. An update in the calibration of the beam synchrotron radiation telescope (BSRT) system was performed both before the transition to BCMS and the crossing angle change.



**Figure 4.1:** Using the measured bunch parameters, i.e., transverse emittances, bunch intensity and length, the calculated peak luminosity values (crosses) at the beginning of stable beams are compared to the average measured peak luminosity provided by the ATLAS (blue circles) and CMS (red circles) during 2016 [220]. The drop in peak luminosity for Fills 5219, 5222, 5223, and 5433 corresponds to machine development conditions [221, 222] for studying electron cloud effects on the LHC performance.

#### 4.1.2 Formalism of absolute luminosity from separating the beams

The bunch luminosity  $\mathcal{L}_b$  produced by one colliding bunch pair, with time- and position-dependent density functions,  $\rho_1(x, y, z, t)$  and  $\rho_2(x, y, z, t)$ , is given by [218]

$$\mathcal{L}_b = KN_1N_2f_r \int_{-\infty}^{\infty} \rho_1(x, y, z, t)\rho_2(x, y, z, t) dx dy dz dt, \quad 4.3$$

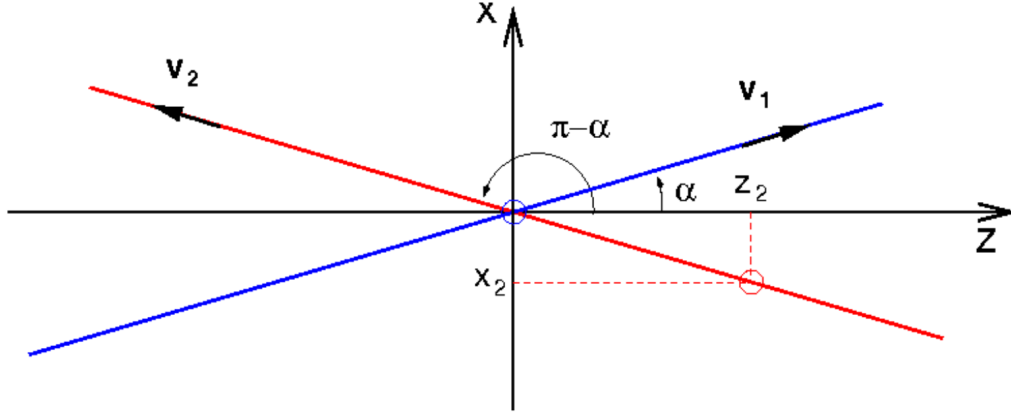
where  $N_1$  and  $N_2$  are the numbers of protons in the two colliding bunches respectively,  $f_r = 11\,2455$  Hz is the bunch orbit frequency around the LHC ring, and  $\rho_{1,2}$  are the bunch proton densities normalized to unity at any time  $t$ . While the bunch populations can be measured to good precision directly, a precise measurement of  $\rho_{1,2}$  is difficult. The kinematic factor  $K$

$$K = \sqrt{|\vec{v}_1 - \vec{v}_2|^2 - \frac{\vec{v}_1 \times \vec{v}_2}{c^2}}, \quad 4.4$$

can be readily attained in case of equal and relativistic beam velocities, i.e.,  $|\vec{v}_1| = |\vec{v}_2| = c$ , and Eq. (4.3) can be thus simplified to

$$\mathcal{L}_b = (2\cos^2\partial_C)N_1N_2f_r \int_{-\infty}^{\infty} \rho_1(x, y, z, t)\rho_2(x, y, z, t) dx dy dz dt, \quad 4.5$$

in terms of a half-crossing angle  $\partial_C$  defined by the two beam trajectories without loss of generality in the  $x$ - $z$  plane (Fig. 4.2). The time-dependent beam overlap  $f = \rho_1\rho_2$  equals simply to the product of the normalized particle-density distributions, and it can



**Figure 4.2:** Definition of coordinates and beam trajectories [218]; the laboratory frame  $xyz$  is defined such that the  $x$  axis points in the direction of  $\bar{v}_1 + \bar{v}_2$ , the  $y$  axis in that of  $\bar{v}_1 \times \bar{v}_2$ , and the  $z$  axis in that of  $\bar{v}_1 - \bar{v}_2$ , where  $\bar{v}_1 = c \hat{v}_1(\sin\alpha, 0, \cos\alpha)$  and  $\bar{v}_2 = c \hat{v}_2(\sin\alpha, 0, -\cos\alpha)$ . The points  $(x_1, z_1)$  and  $(x_2, z_2)$  are the positions of the bunch centers at time  $t = 0$ . For simplicity, the former is on the intersect of the two beam trajectories in the  $x$ - $z$  plane. The third dimension ( $y$ ) is here suppressed.

be approximated as

$$f(x, y, z, t) = \frac{1}{(2\pi)^3 \sigma_{xi} \sigma_{yi} \sigma_{zi}} \exp \left[ - \left( \frac{(x - x_i)^2}{2\sigma_{xi}^2} + \frac{(y - y_i)^2}{2\sigma_{yi}^2} + \frac{(z - ct)^2}{2\sigma_{zi}^2} \right) \right], \quad 4.6$$

assuming perfectly Gaussian bunch profiles. The values  $\sigma_{xi}$ ,  $\sigma_{yi}$ , and  $\sigma_{zi}$  are the transverse and longitudinal beam sizes ( $i = 1, 2$ ) in the frame as defined in Fig. 4.2, whereas  $x_i$  and  $y_i$  correspond to the transverse positions of the bunch centroids at the nominal collision point ( $t = 0$ ). The time-integrated beam overlap distribution reveals cross terms implying that the resulting distribution is not exactly factorizable in a  $x$ - and  $z$ -dependent Gaussian profile. For small crossing angles and  $\sigma_{zi} \gg \sigma_{xi}$ —two conditions typically valid at LHC—the longitudinal dependence of the transverse beam size is raised, and one can always find a rotated reference in the crossing plane in which the beam overlap is the product of two profiles each depending only on one position variable. For beams colliding with a half crossing-angle  $\alpha$  in the  $x$ - $z$  plane, and with a relative transverse offsets  $\Delta x$  ( $\Delta y$ ) in the  $x$  ( $y$ ) direction, it can be therefore shown that integrating Eq. (4.6) leads to

$$\mathcal{L}_b = 2\cos^2 \partial_C N_1 N_2 f_r \int_{-\infty}^{\infty} f(x, y, z) dy dz = \frac{\cos^2 \partial_C N_1 N_2 f_r}{2\pi \Sigma_x \Sigma_y} \exp \left[ - \left( \frac{\Delta x^2}{2\Sigma_x^2} + \frac{\Delta y^2}{2\Sigma_y^2} \right) \right]. \quad 4.7$$

More specifically, the variances  $\sigma_{xi}^2$  and  $\sigma_{yi}^2$  have been convolved to

$$\Sigma_x = \sqrt{(\sigma_{x1}^2 + \sigma_{x2}^2) \cos^2 \partial_C + (\sigma_{z1}^2 + \sigma_{z2}^2) \sin^2 \partial_C}, \quad 4.8$$

in the crossing plane, and

$$\Sigma_y = \sqrt{(\sigma_{y1}^2 + \sigma_{y2}^2)}, \quad 4.9$$

otherwise.

The vdM scan method allows to measure the beam overlap in Eq. (4.6) assuming that the two bunch densities factorize in  $x$  and  $y$ , meaning

$$\int_{-\infty}^{\infty} \rho_1(x, y) \rho_2(x + \Delta x, y + \Delta y) dx dy = \int_{-\infty}^{\infty} \rho_1(x) \rho_2(x + \Delta x) dx \int_{-\infty}^{\infty} \rho_1(y) \rho_2(y + \Delta y) dy. \quad 4.10$$

The estimate of the possible bias introduced by this assumption is calculated in Section 4.3.3. Both sides of Eq. (4.3) can then be integrated independently in  $\Delta x$  and  $\Delta y$ , while the separation in the other direction is kept fixed at arbitrary values  $\Delta y_0$  and  $\Delta x_0$ , respectively. For  $\partial_C = 0$

$$N_1 N_2 f_r \int_{-\infty}^{\infty} \rho_1(y) \rho_2(y + \Delta y_0) dy = \int_{-\infty}^{\infty} \mathcal{L}_b(\Delta x, \Delta y_0) d(\Delta x), \quad 4.11$$

and therefore

$$\int_{-\infty}^{\infty} \rho_1(x) \rho_2(x + \Delta x_0) dx = \frac{\mathcal{L}_b(\Delta x_0, \Delta y_0)}{\int_{-\infty}^{\infty} \mathcal{L}_b(\Delta x, \Delta y_0) d(\Delta x)}. \quad 4.12$$

Likewise for  $y$ . Experimentally the integration over  $\Delta x$  and  $\Delta y$  is implemented by scanning the two beams against each other and the integral in the denominator of Eq. (4.12) is evaluated by measuring the detector rate as a function of the beam-beam separation, the so-called “scan curves.” After replacing the factorized in  $x$  and  $y$  beam overlap integral according to Eq. (4.12), Eq. (4.7) becomes for any head-on collision:

$$\mathcal{L}_b(\Delta x = 0, \Delta y = 0) = N_1 N_2 f_r \frac{R(\Delta x = 0, \Delta y_0) R(\Delta x_0, \Delta y = 0)}{\int_{-\infty}^{\infty} R(\Delta x, \Delta y_0) d(\Delta x) \int_{-\infty}^{\infty} R(\Delta x_0, \Delta y) d(\Delta y)}, \quad 4.13$$

where the luminosity is expressed in terms of the rate  $R(\Delta x, \Delta y)$  measured when the two beams are separated by values  $\Delta x$  and  $\Delta y$ , respectively. The integrals of the scan curves can be obtained from the convolved beam widths  $\Sigma_x$  and  $\Sigma_y$

$$\Sigma_x = \frac{1}{\sqrt{2\pi}} \frac{\int_{-\infty}^{\infty} R(\Delta x, \Delta y_0) d(\Delta x)}{R(\Delta x = 0, \Delta y_0)}. \quad 4.14$$

Likewise for  $y$ . In the case of Gaussian luminosity curves, the convolved beam width coincides with the standard deviation of that distribution. Equation (4.14) is generic though meaning  $\Sigma_x$  and  $\Sigma_y$  depend only upon the area under the luminosity curve. The appealing feature of the vdM method is therefore no assumption about the shape of the scan curve is made. The bunch luminosity at zero separation is extracted from machine parameters by performing a pair of beam-separations scans

$$\mathcal{L}_b(\Delta x = 0, \Delta y = 0) = \frac{N_1 N_2 f_r}{2\pi \Sigma_x \Sigma_y}, \quad 4.15$$

such that the final formula used to measure the so-called “visible cross section” is

$$\sigma_{\text{vis}} = \frac{2\pi \Sigma_x \Sigma_y R(\Delta x = 0, \Delta y = 0)}{N_1 N_2 f_r}. \quad 4.16$$

In contrast to beam-profile measurements,  $\Sigma_x$  and  $\Sigma_y$  are directly determined from fits

of the scan curves based on detector rate measurements during the vdM scans. While the convolved beam widths are the same for all detectors, the peaks of the corresponding scan curves depend on the detector, meaning the exercise is repeated per detector. The comparison of the  $\Sigma_x$  and  $\Sigma_y$  as estimated from the scan curves of different detectors represents a crucial cross-check.

## 4.2 Offline luminosity measurement in CMS

### 4.2.1 Experimental setup

On the one hand, the BCM1F, HF, and PLT detectors are characterized by an independent high-rate DAQ system and thus serve as excellent luminosity monitoring systems online. On the other hand, the silicon pixel and DT detectors feature very low occupancy and good stability over time. These two detectors utilize the standard CMS trigger and DAQ systems, in contrast to the others, which are read out asynchronously with respect to the rest of CMS. The employed techniques for luminosity determination can be classified as follows.

1. “Event” counting: the fraction of bunch crossings is determined during which a specified detector registers an event satisfying a selection requirement. For instance, a bunch crossing can be deemed to contain an event, if at least one interaction in that crossing induces a coincidence of observed hits in the detector. The BCM1F, HFOC, and PLT luminosity measurements fall into this category.
2. “Hit” counting: the hit multiplicity per bunch crossing is registered, e.g., the number of electronic channels or energy clusters above a background threshold. Representative examples are methods exploiting pixel-cluster, vertex and track counting.
3. “Particle” counting: the particle multiplicity per bunch crossing is inferred from reconstructed quantities such as calorimeter-energy distributions or observables sensitive to particle flux traversing the detector. The DT-based and HFET methods have been used successfully and proved reliable similarly to the event- and hit-counting methods.

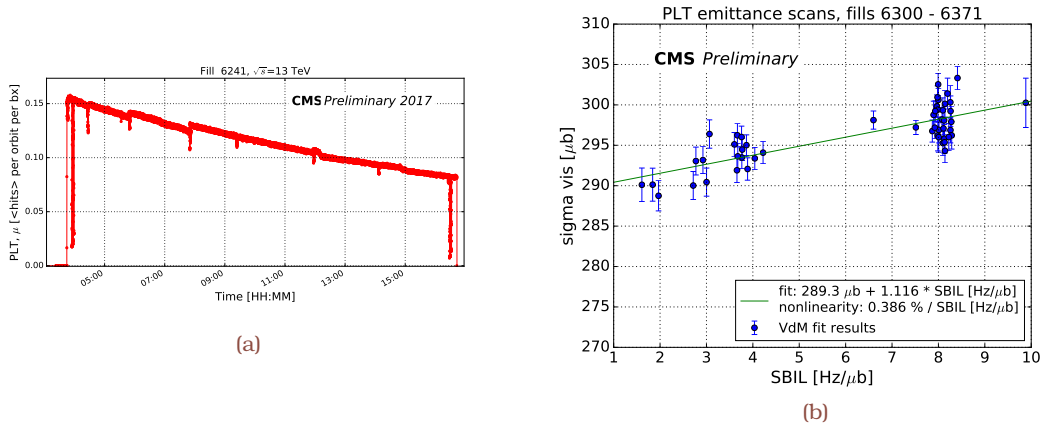
Considering the event counting as the simplest approach, each separate term in the series of Eq. (4.1) can be expressed as a function of a detector-dependent constant, i.e., the visible cross section, which relates the measurable quantity,  $\mu_{\text{vis}}$ , to the absolute bunch luminosity (Fig. 4.3a) as

$$\mathcal{L}_b = \frac{\mu_{\text{vis}} f_r}{\varepsilon \sigma_{\text{inel}}}. \quad 4.17$$

The efficiency  $\varepsilon$  determines the average number of inelastic collisions per bunch crossing to satisfy the selection requirements, while  $\sigma_{\text{vis}} = \varepsilon \sigma_{\text{inel}}$  is a unique luminometer property that depends on detector acceptance. Since the Poisson probability  $P_0(\mu_{\text{vis}})$  for observing zero events in a given bunch crossing out of a total number of  $n_{\text{BC}}$  equals to  $P_0(\mu_{\text{vis}}) = e^{-\mu_{\text{vis}}}$ , the probability of observing at least one occurrence of events ( $n_{\text{FORB}}$ ) in either forward or backward IP region reads  $P = 1 - P_0(\mu_{\text{vis}}) = 1 - e^{-\mu_{\text{vis}}}$ , which can be solved for  $\mu_{\text{vis}} \ll 1$

$$\mu_{\text{vis}} = -\ln\left(1 - \frac{n_{\text{FORB}}}{n_{\text{BC}}}\right). \quad 4.18$$





**Figure 4.3:** (a) Instantaneous luminosity measured by PLT as a function of time [113]. The two deep dips—at the beginning and end of the Fill—correspond to emittance scans. The shallower dips during the Fill correspond to optimization scans, which are performed by the LHC operators to tune the beam positions for complete overlap. (b) Visible PLT cross section as a function of the single-bunch instantaneous luminosity that serves as metric of  $\mu_{\text{vis}}$  [113]. The best-fit function (first-order polynomial) is superimposed, with its parameters displayed on the legend.

The efficiency to detect a single inelastic interaction is constant in the absence of subtle detector effects, and it can be generalized to arbitrary  $n$  simultaneous occurrences

$$\varepsilon_n = 1 - (1 - \varepsilon)^n, \quad 4.19$$

given that the number of interactions in any bunch crossing also obeys a Poisson distribution.

When  $\mu_{\text{vis}} \gg 1$ , event-counting methods start losing sensitivity (Fig. 4.3b) as progressively fewer bunch crossings result in zero observed interactions. This can be alleviated by counting the number of hits ( $n_{\text{hits}}$ ) in a given detector rather than the total number of events. Under the assumption that the number of hits in one interaction follows a binomial distribution, the average probability to register a hit per bunch crossing in one of the detector channels (“CH”) out of  $n_{\text{CH}}$  is similarly given by  $P = 1 - e^{-\mu_{\text{vis}}}$ , which can be solved for the measurable quantity

$$\mu_{\text{vis}} = -\ln\left(1 - \frac{n_{\text{hits}}}{n_{\text{BC}}n_{\text{CH}}}\right). \quad 4.20$$

In particular, the binomial assumption used to derive Eq. (4.20) holds true for pixel-cluster counting (PCC), i.e., the probability to observe a hit in any single channel is independent of the number of hits observed in the other channels.

### 4.2.2 The vdM scan protocols

The beam conditions for vdM calibration are different from those in normal physics Fills, with fewer bunches colliding, lower bunch intensities and typically special optical configuration at the IP. In other words, they are specifically optimized to handle various sources of systematic uncertainty in the calibration procedure. The extrapolation of the

absolute luminosity calibration from the low-rate regime to the nominally high-pileup conditions is given by the intrinsic nonlinearity and cross-detector stability. As a complementary method for measuring the beam size during high-intensity conditions, luminosity scans with a small beam separation—regularly performed at LHC already since the 2015 proton physics operation—can be exploited to further constrain [TH8] the extrapolation uncertainty.

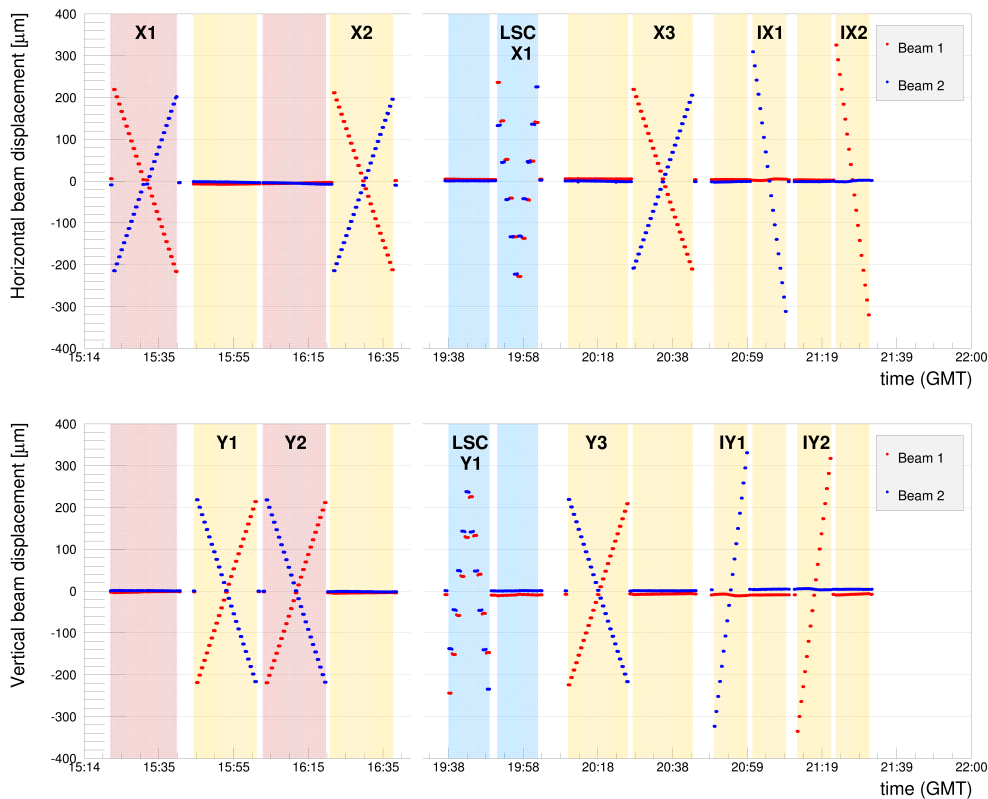
The vdM scans for the luminosity calibration of the pp reference run at  $\sqrt{s} = 5.02$  TeV were performed during the LHC Fill 4634 in November 2015. The LHC beam conditions for Fill 4634 are summarized in Table 4.1. The LHC filling scheme was Multi\_44b\_22\_22\_22\_4bpi12inj with 22 colliding bunch pairs at IP5 widely spread over the orbit to reduce long-range beam-beam effects and detector afterglow. LHC beam optics were adjusted to  $\beta^* = 4$  m and transverse emittance of  $\epsilon_N \approx 3.5 \mu\text{m}$  resulting in relatively wide beams of about  $\sigma_b = 72 \mu\text{m}$ . The two beams were crossing with an angle of  $170 \mu\text{rad}$ .

**Table 4.1:** Summary of the LHC beam parameters for Fill 4634 at  $\sqrt{s} = 5.02$  TeV [TH3]. The vdM scans for the luminosity calibration of the proton-ion runs at  $\sqrt{s_{NN}} = 8.16$  TeV were performed under physics conditions, i.e. low- $\beta^*$  operation, during the LHC Fills 5527 and 5563 in November and December 2016, respectively [TH2].

Beam energy	2510 GeV
Peak luminosity	$2.4 \times 10^{30} \text{ cm}^{-2} \text{ s}^{-1}$
Peak pileup	$\langle \mu \rangle = 0.6$
Injection scheme	Multi_44b_22_22_22_4bpi12inj
Beam 1/2 intensity	$3.2 \times 10^{12}$
Number of bunches in beam 1/2	44
Number of colliding bunches in beam 1/2	22
Beams crossing angle	$170 \mu\text{rad}$
$\beta^*$	4 m

The beam intensities, about  $3 \times 10^{12}$  protons per beam, are measured with the DC Current Transformers (DCCT) [223], whereas the individual bunch currents are measured either with the Fast Beam Current Transformers (FBCT) [224] or the Beam Quality Monitor (BQM) [225]. The amount of elementary charge in the nominally empty bunch slots (“ghost charge”) and the charge circulating in the RF buckets adjacent to the nominal bunch slots (“satellite charge”) are estimated by means of the LHC Longitudinal Density Monitors (LDM) [226]. An independent measurement of ghost charges is provided by the LHCb Collaboration [227]. The beam orbit is monitored using the Diode ORbit and Oscillation System (DOROS) beam position monitor [228]. To maximize the number of events specifically for PCC at large beam separations CMS gates zero-bias triggers on a restricted number of bunch pairs and records events with a bandwidth of about 18 kHz.

Since the determination of  $\sigma_{\text{vis}}$  requires the measurement of the convolved transverse beam sizes, a minimum of two separate beam scans is required, i.e., one where the beams are separated horizontally and a second where the beams are separated vertically. The beams are sequentially moved in a series of scan steps, and data recorded at each step to obtain a statistically significant measurement per luminometer. To help assess the



**Figure 4.4:** Horizontal and vertical beam displacements, measured with the DOROS beam position monitor, during the vdM scan campaign at  $\sqrt{s} = 5.02$  TeV in November 2015 [TH3].

experimental systematic uncertainty two such scan pairs (at least) are performed typically in short succession to provide independent calibrations under similar beam conditions.

The CMS vdM scan program of November 2015, as summarized in Fig. 4.4, consisted of six sequential  $x$ - $y$  scan pairs: two  $x$ - $y$  regular scans, one length scale calibration (LSC) scan, one additional  $x$ - $y$  scan and two BI scans.<sup>1</sup> For the three vdM scan pairs the two beams were put to  $\pm 6 \sigma_b = \pm 434 \mu\text{m}$  with respect to the nominal beam position for head-on collisions and scanned in 25 steps across one another in opposite directions. For the BI scans, beam 1 (2) is kept fixed at nominal position while beam 2 (1) is being separated and moved in 19 steps from  $+4.5$  to  $-4.5 \sigma_b = \pm 325 \mu\text{m}$ , first in  $x$  and then in  $y$ . The time spent at each scan point is about 30 s. For the LSC scans, the two beams are kept with fixed separation of  $94 \mu\text{m}$  and simultaneously moved back and forth in the vertical and horizontal directions to derive a linear length scale correction.

The vdM scan program in November–December 2016 consisted [TH2] of four sequential  $x$ - $y$  scan pairs: two  $x$ - $y$  regular scans, one LSC scan, and one additional  $x$ - $y$  scan to ensure reproducibility. For the three vdM scan pairs the two beams were put to  $\pm 6 \sigma_b = \pm 108 \mu\text{m}$  and scanned in 25 steps across one another in opposite directions. The time spent at each scan point was about 30 seconds. For the LSC scans the two beams are kept

<sup>1</sup>During the first  $x$  scan ( $X_1$ ) and the second  $y$  scan ( $Y_2$ ) issues in the pixel readout prevented from acquiring all the data. In the November 2015 vdM analysis the  $x$  and  $y$  scans are paired as: (1, 2)  $\rightarrow$  ( $Y_1$ ,  $X_2$ ), (3, 4)  $\rightarrow$  ( $Y_3$ ,  $X_3$ ), (5, 6)  $\rightarrow$  ( $IY_1$ ,  $IX_1$ ), and (7, 8)  $\rightarrow$  ( $IY_2$ ,  $IX_2$ ).

with fixed separation of  $18\ \mu\text{m}$  and simultaneously moved back and forth in the vertical and horizontal directions to derive a linear length scale correction. The direction of the higher energy proton beam was initially clockwise and was then reversed, producing two statistically independent data sets (“Pbp” and “pPb”) that respect the usual convention of the proton fragmentation region being probed in CMS, i.e., the proton-going side defining the positive fragmentation region.

### 4.2.3 The pixel-cluster counting method

The PCC method, featuring a very low occupancy of less than a permille even under high-pileup conditions, has already been shown to provide high precision luminosity measurements [65, 217]. As in Run 1 and 2, the rate algorithm uses the mean number of pixel clusters per bunch crossing. All reconstructed pixel clusters are considered in the counting process. However, clusters reconstructed in the innermost barrel layer (Fig. 3.8) are affected by dynamic inefficiency, e.g. memory size (buffer) overflow in pixel readout when the L1 trigger rate is very high, and are thus excluded from further consideration. Generally, this effect is found to be less than 0.4% for the rest of barrel layers and endcap disks during the 2015 running conditions [217], and a 0.4% is taken as a systematic uncertainty. In addition, pixel modules that are not fully operational throughout the entire data-taking period have been omitted from the cluster counting sum. Finally, the pixel front-end driver numbered 33 (FED 33) was excluded from the DAQ system during the  $X_1$  and  $Y_2$  vdM scans in Fill 4634. To minimize scan-to-scan variations and to ensure scan-to-scan reproducibility pixel clusters associated with FED 33 are therefore not taken into account in the vdM analysis or in the luminosity measurement at  $\sqrt{s} = 5.02\ \text{TeV}$ .

During usual data taking conditions, out-of-time response affects the true mean number of pixel clusters. The particle flux through the detectors at LHC is dominated by collision products, whereas the single-beam halo created by aperture losses or beam-gas interactions are typically negligible. The tight bunch spacing induces a peculiar type of “afterglow” background that affects luminosity measurements at the level of up to a few percents during normal physics conditions, depending on the detector and algorithm considered. Bunch pairs collided at the considered IP are called “colliding” BCIDs, while bunches that do not collide at the considered IP are labeled “nonactive.” Figure 4.5 shows the single-bunch instantaneous luminosity (SBIL) as a function of BCID measured by the PCC rate within a typical Fill during the  $\sqrt{s} = 5.02\ \text{TeV}$  conditions from data collected with “random triggers,” i.e., signaling nonactive bunch crossings. While the bunch train structure of the filling scheme is clearly visible, a nonvanishing rate in nonactive bunch slots can be observed. The relative magnitude of such structures, as observed in Fig. 4.5, depends on the instrumental characteristics and the local material distribution. They may also contain a fraction of collisions between a nonactive bunch in one beam and an unbunched component in the opposing beam.

Two out-of-time response effects are distinguished and taken into account. The first effect (“type 1”) is due to a tail of the pixel hit signal leaking into the time integration window of the next 25 ns bunch slot. Such an effect is visible in Fig. 4.5 in the trailing

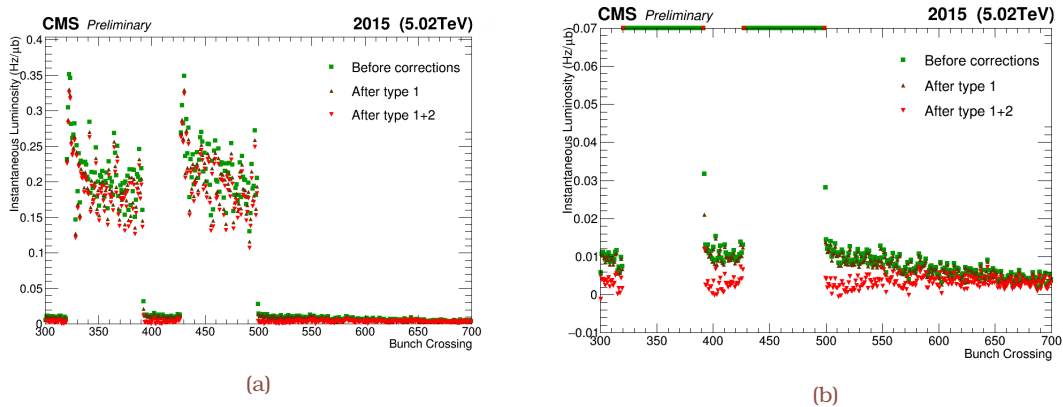


Figure 4.5: Instantaneous luminosity as a function of bunch crossing, showing the impact of type 1 and 2 corrections [TH3]. Plot (b) zooms in on the vertical axis of plot (a).

bunch slots after the trains. The second effect (“type 2”) is due to the exponentially decaying activation of the material surrounding the detector; the long tails in Fig. 4.5 is the activation result from the previous bunch crossings. Since it is assumed that the type 1 effect mainly depends on the final response, while the type 2 effect stems from real activity in the pixel detector, the correction of type 1 effect should precede the one related to type 2.

The correction of type 1 effect is modeled by:

$$C_1(n+1) = cR(n), \quad 4.21$$

where  $C_1(n+1)$  is the type 1 correction in the  $(n+1)$ th bunch,  $R(n)$  is the SBIL response in the  $n$ th bunch, and  $c$  is a constant determined on a Fill-by-Fill basis. The average of the residual correction for the first nonactive bunch slot after a train is considered as a systematic uncertainty. For type 2, the correction model is iteratively built and reads

$$C_2(n+j) = \beta e^{(-\hat{\eta}j)} A(n), \quad j > 0, \quad 4.22$$

where  $C_2(n+j)$  is the type 2 correction in the  $(n+j)$ th bunch due to the activity in the  $n$ th bunch;  $A(n)$  is the true (corrected) activity in the  $n$ th bunch and  $\beta$ ,  $\hat{\eta}$  are the constants determined as follows. Based on SBIL measurements for random trigger samples recorded during 25 ns bunch-spacing periods, the model parameters  $(\beta, \hat{\eta})$  are estimated by minimizing the root-mean-square of SBILs around zero for nonactive bunch slots. The parameters of type 2 correction are determined to be  $\beta = 0.00086$  and  $\hat{\eta} = 0.014$  for all 2015 data [217], and their performance is found to be optimal both within and out of trains.

#### 4.2.4 Extraction of the visible cross section

The size of the beam overlap is measured by fitting the luminometer rate measurements, normalized by the bunch current product, as a function of beam-beam separation.

The scan curves are fitted with a double-Gaussian model with an additional constant term, whose purpose is to subtract pedestals from the background rates. A more accurate estimate of the constant term is obtained measuring the beam overlap by simultaneously fitting the PCC and reconstructed PV (“vertex counting”) rate measurements; PCC data are available only during the November 2015 vdM scan program. The effective horizontal and vertical widths of the beam ( $\Sigma_x$  and  $\Sigma_y$ ) as well as the normalized rates ( $R_x$ ,  $R_y$ ) are obtained per scan per BCID. For each BCID the visible cross sections are then measured using

$$\sigma_{\text{vis}} = 2\pi\Sigma_x\Sigma_y\mu_{\text{vis}}, \quad 4.23$$

where

$$\mu_{\text{vis}} = \frac{1}{2}(R_x + R_y), \quad 4.24$$

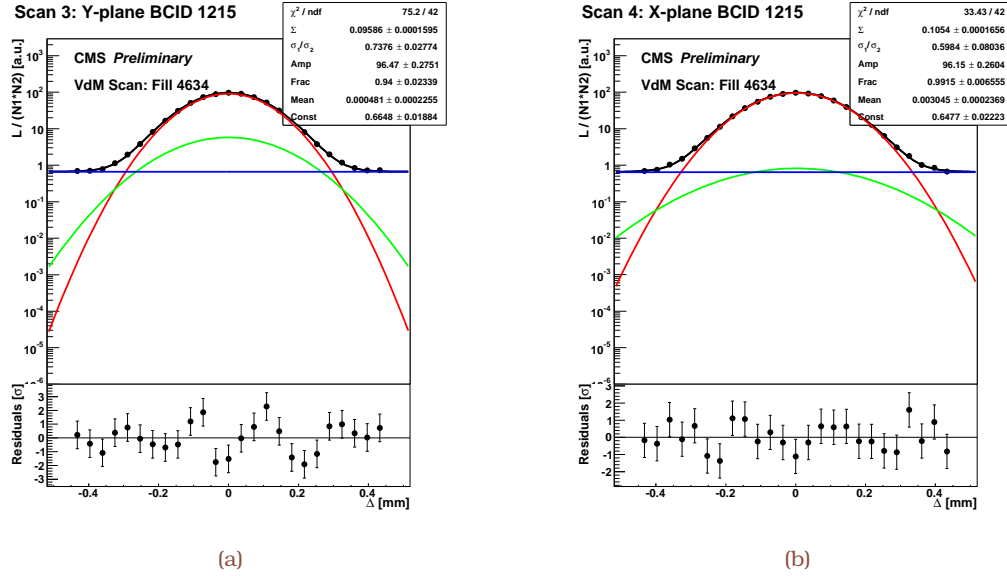
with  $R_{x,y}$  denoting the amplitudes of the fitted scan curves, i.e.,  $R_x \equiv R(\Delta x = 0, \Delta y_0)$  and  $R_y \equiv R(\Delta x_0, \Delta y = 0)$ .

Example fits for BCID 1215 in  $Y_3$  and  $X_3$  are shown in Figs. 4.6 and 4.7 for PCC and vertex counting, respectively, during the November 2015 vdM program. For vertex selection in the vertex counting method the standard CMS definition of good vertices (see Section 3.2) is applied. The fitted  $\Sigma_x$  and  $\Sigma_y$  parameters are shown in Fig. 4.8, while the resulting  $\sigma_{\text{vis}}$  are illustrated in Fig. 4.9. For the November 2015 vdM program all luminometers produce compatible estimates: the percent differences between any two luminometers are statistically consistent with zero. For the November–December 2016 vdM program the percent differences between any two luminometers is consistent with zero at the level of 0.5 and 0.2% in the Pbp and pPb period, respectively, while a scan-to-scan variation is also assigned based on the root-mean-square of the measured  $\sigma_{\text{vis}}$ , which is found to be 0.6 (1.0)% in the Pbp (pPb) period, respectively.

## 4.3 Calibration corrections

### 4.3.1 Concept and formalism of constant separation LSC scan

The ability to measure the convolved beam sizes depends upon the precise knowledge of the distance by which the beams are separated during the scans. If the nominal scale for beam displacement differs from the true displacement scale, a systematic error related to length scale can be introduced. At each scan point, the absolute beam separation is controlled by a set of closed orbit bumps applied locally near the considered IP using steering correctors (Fig. 4.10). During dedicated LSC calibration measurements, that are typically performed using the vdM protocol, i.e., the same optics configuration, both beams are moved in five equidistant steps first in  $x$  and then in  $y$  keeping their nominal separation constant. The nominal beam displacement entered into the accelerator control system can be thereby calibrated against the measured displacement of the luminous centroid. To enhance the accuracy of the beam separation derived from the LHC orbit knobs a correction is applied that is determined by measuring the luminous region movement using the tracker characterized by great resolution (Table 2.4). Since each of



**Figure 4.6:** Examples of fitted scan curves, i.e., normalized rates recorded based on the PCC method as a function of the beam separation ( $\Delta$ ) in (a)  $Y_3$  and (b)  $X_3$  scans in the November 2015 vdM period [TH3]. The applied fit model is represented as the green and red, blue, and black curves corresponding to the two Gaussian components, the constant term, and their sum, respectively. The values and the statistical uncertainty in the fitted parameters are shown on the legend, along with the reduced  $\chi^2$ . The bottom panels include the residuals, i.e., the difference between the measured and fitted values divided by the statistical uncertainty. The fit is performed simultaneously with the rates of Fig. 4.7.

the four bump amplitudes (two beams in two transverse directions) depends on different magnet and lattice functions, in principle, LSC scans can independently determine these four calibration constants. However, since for the vdM scans the beams are moved symmetrically (as opposed to one beam at a time), only the average length scale matters to first order for calibrating the absolute separation.

The beams are offset from one another by about one  $\sigma_b$ , and then moved in equidistant steps; differences in the amounts by which the beams have been moved translate in rate differences as the scan progresses. More specifically, if  $x_i^0$  and  $x_i$  are the nominal and actual horizontal beam displacements, respectively, the LSC correction factors for the beam 1 and 2,  $a_1$  and  $a_2$ , are defined through

$$x_i \equiv a_i x_i^0, \quad 4.25$$

where ideally  $a_i = 1$ . At the start of the scan, the two beams are positioned as shown in Fig. 4.4, i.e., at zero and at a value of  $a_2 s_0$ , respectively, where  $s_0$  is the nominal beam separation, while, as the scan progresses, both beams are moved by a nominal amount of  $\Delta^j$  at the  $j$ th step, meaning

$$x_1^j = a_1 \Delta^j, \quad 4.26$$

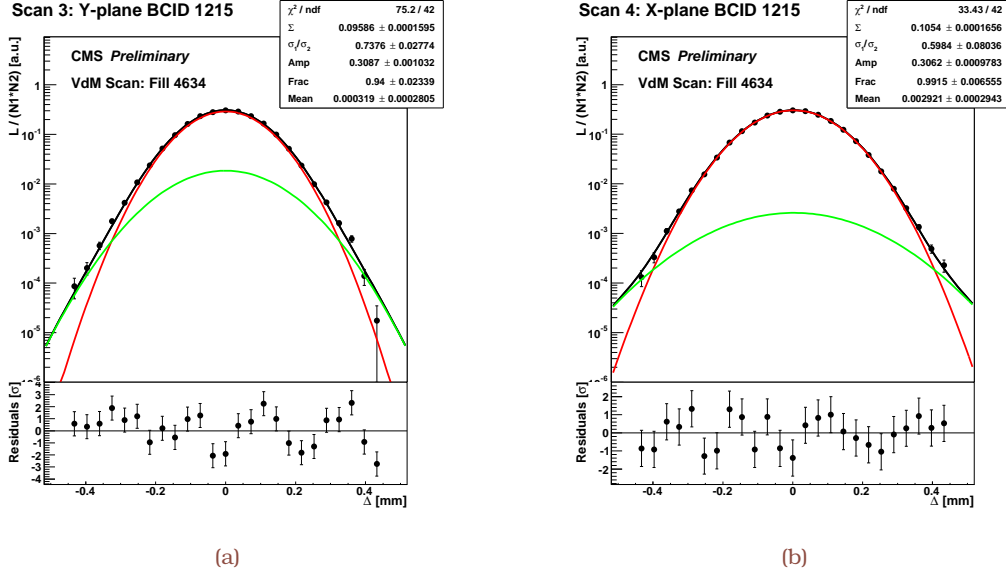


Figure 4.7: Examples of fitted scan curves, i.e., normalized rates recorded based on the vertex counting method as a function of the beam separation ( $\Delta$ ) in (a)  $Y_3$  and (b)  $X_3$  scans in the November 2015 vdM period [TH3]. The applied fit model is represented as the green and red, blue, and black curves corresponding to the two Gaussian components, the constant term, and their sum, respectively. The values and the statistical uncertainty in the fitted parameters are shown on the legend, along with the reduced  $\chi^2$ . The bottom panels include the residuals, i.e., the difference between the measured and fitted values divided by the statistical uncertainty. The fit is performed simultaneously with the rates of Fig. 4.6.

and

$$x_2^j = a_2(s_0 + \Delta^j). \quad 4.27$$

The average position of the two beams, which is the position of the luminous region, reads

$$\bar{x}^j = \left( \frac{a_1 + a_2}{2} \right) \Delta^j + \frac{a_2 s_0}{2} \equiv \bar{a} \Delta^j + s'_0, \quad 4.28$$

while the distance between the two beams is given by

$$\Delta x^j = a_2(s_0 + \Delta^j) - a_1 \Delta^j = (a_2 - a_1) \Delta^j + a_2 s_0 \equiv \epsilon \Delta^j + 2s'_0. \quad 4.29$$

The observed rate,  $\mu(\Delta^j)$ , is related to  $\mu_{\text{vis}}$ , i.e., the rate when beams are fully aligned, via

$$\mu(\Delta^j) \stackrel{\text{Eq.(4.7)}}{\simeq} \mu_{\text{vis}} \left( 1 - \frac{s_0 \epsilon}{\Sigma_{x,y}^2} \Delta^j \right), \quad 4.30$$

where a Taylor expansion of the  $\Delta^j$  variable is assumed using the fact that  $(\Delta x^j)^2 \simeq s_0^2 + 2s_0 \epsilon \Delta^j$  and  $\epsilon \ll 1$ . It becomes thus obvious from Eqs. (4.28) and 4.30 that the values for  $\bar{a}$  and  $\epsilon$  can be obtained by fitting the measured  $\bar{x}^j$  and  $\mu(\Delta^j)$ , respectively, as a function



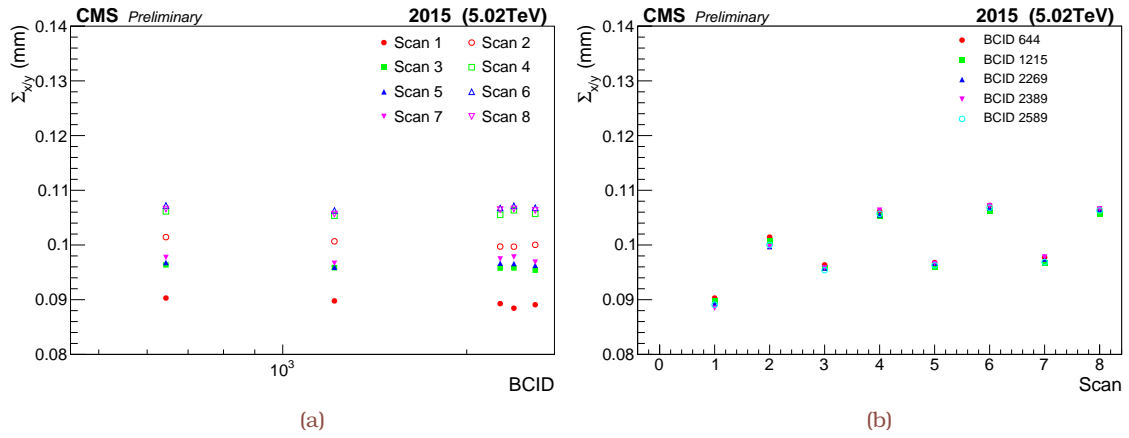


Figure 4.8: The extracted  $\Sigma_{x/y}$  (a,b) from the PCC method as a function of BCID (a) and scan sequence (b) in the November 2015 vdM period [TH3]. The sequential naming convention adheres to the horizontal and vertical scan sequence as illustrated in Fig. 4.4.

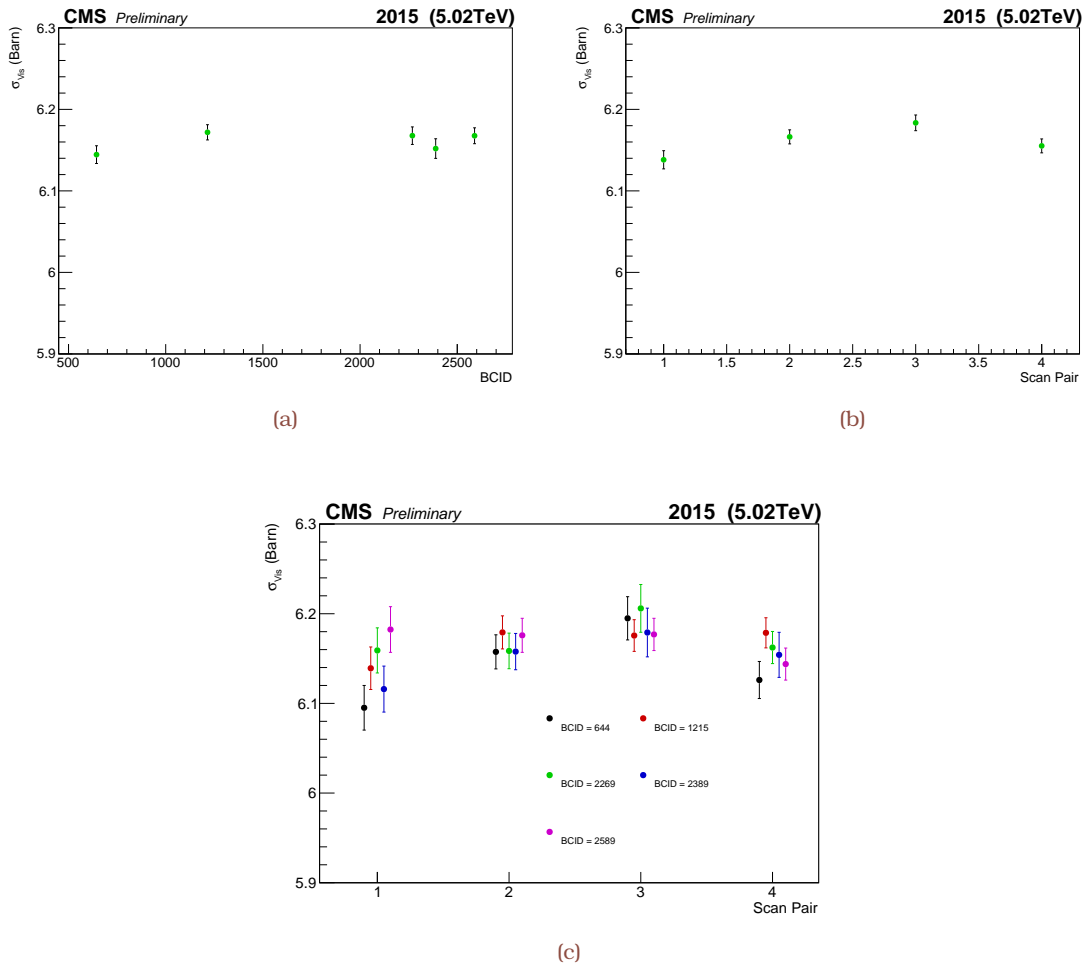
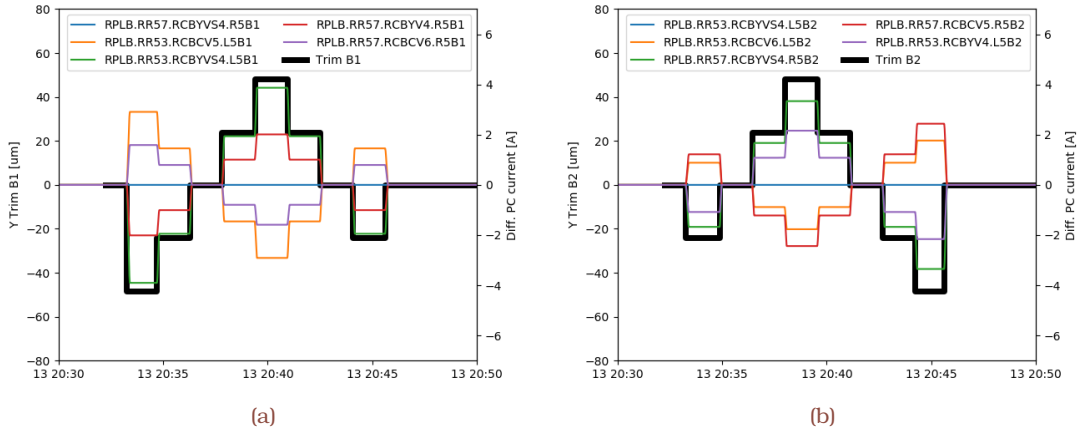


Figure 4.9: The weighted average of PCC  $\sigma_{vis}$  per BCID (a) and scan (b), and separately in each BCID and scan (c) in the November 2015 vdM period [TH3]. Errors are statistical only.



**Figure 4.10:** The beam-separation scans—here a LSC scan of constant separation type—are performed automatically by the operators of LHC moving independently one beam (a) stepwise across the other (b). The nominal displacement of the beams at the IP (black line) is achieved based on a local distortion of the orbit using a pair of steering dipoles located on either side of the IP. The acquisition is performed while the magnets are completely idle, i.e., when the applied magnet currents (colored lines) are stable.

of the beam separation  $\Delta^j$ . These values can in turn be related to  $a_1$  and  $a_2$  via

$$a_1 = \bar{a} - \frac{\epsilon}{2} \quad 4.31$$

and

$$a_2 = \bar{a} + \frac{\epsilon}{2}, \quad 4.32$$

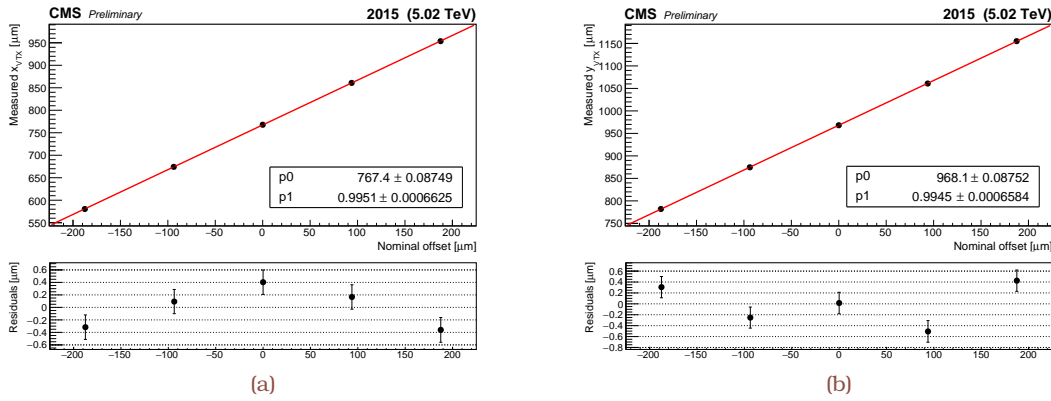
respectively. For all data points for which the nominal offsets of beam 1 and 2 are set to  $-\Delta^j/2$  and  $+\Delta^j/2$ , respectively, the true beam separation is given by

$$\Delta x^j = a_2 \Delta^j - (-a_1 \Delta^j) = \bar{a} \Delta^j. \quad 4.33$$

In other words, the actual beam separation is just the nominal beam separation scaled by the LSC constant  $\bar{a}$ ; the sign of the offset is irrelevant.

The data recorded during the two vertical and horizontal LSC scans, explained in Section 4.2.1, are used. The procedure to derive the LSC calibration constants follows the method already used in previous analyses [229, 230]. The beamspot movement as a function of the nominal offset of the beam centroid is fitted with a first-order polynomial and a calibration constant extracted. This is depicted in Fig. 4.13 for LSC scans in the horizontal and vertical direction.

The fit results for the two LSC scans in both the transverse coordinates are summarized in Table 4.2. These corrections are applied directly to beam-beam separation in the scan curves; their effect found to reduce the measured beam width (horizontally and vertically). The measured visible cross section was reduced by approximately 1% with this correction. A total uncertainty of 0.2% is assigned accounting for the maximum difference between the forward and backward LSC scans (0.1%) and a potential mismodeling owing



**Figure 4.11:** Examples of the measured beamspot positions as a function of the beam centroid offsets are shown for a horizontal (left) and a vertical (right) LSC scans. The points are fitted with a first-order polynomial to derive the LSC calibration constant [TH3].

to missing higher-order terms in the fitting function.

	X	Y
forward LSC scan	$0.9951 \pm 0.0007$	$0.9945 \pm 0.0007$
backward LSC scan	$0.9960 \pm 0.0007$	$0.9954 \pm 0.0006$

**Table 4.2:** Results of the length-scale calibration for the two horizontal and vertical LSC scans in the November 2015 vdM program [TH3]. Errors are of statistical nature. The average of all four calibration constants is found to be 0.993 with negligible uncertainty for the November–December 2016 vdM program [TH2].

### 4.3.2 Orbit drift

Slow orbit drifts can distort the individual scan curves as well as result in slight movement of the beams out of collision between consecutive scans, thereby biasing the measured interaction rate at the peak of the scan. In addition, at each scan step, the actual beam separation may be affected by random jitter of the beam positions from their nominal setting, which in turn induces fluctuations in the luminosity measured at each scan point. No significant orbit drifts of the beams away from their fixed orbit position were observed during the scans that enter into the  $\alpha_{vis}$  calculation. Figure 4.12 shows the difference of the beam orbits from before and after each scan. All horizontal (vertical) drifts, as estimated by the DOROS beam position monitor, are within about  $(-3, +4)$  μm. Assuming that the major part of the horizontal drift happened instantly during the  $X_2$  scan, a decreased effective beam overlap cross section of at most 0.4% was found. Alternative assumptions about the horizontal drift occurring continuously or smoothly over the 25 separation steps were disproved, for systematically producing on average a larger effective beam overlap cross section. This is taken as uncertainty, and no further corrections are applied.

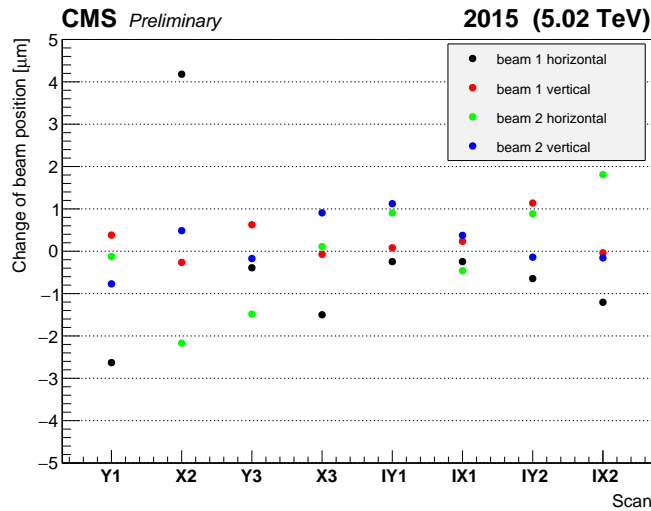


Figure 4.12: Difference between the two vertical and horizontal beam orbits from before and after each scan as measured by the DOROS beam position monitor in  $\mu\text{m}$  is shown [TH3].

### 4.3.3 Impact of nonfactorizable beam shape

The assumption made for the vdM method that the proton bunch densities are factorizable in  $x$  and  $y$ , i.e., that the  $x$  ( $y$ ) shape measured at a working point  $\Delta y_0$  ( $\Delta x_0$ ) does not depend on the working point position, is not valid for the required precision in general and may result in a biased estimate of the beam overlap integral [136]. Short of input from beam-gas imaging measurements [227, 231, 232], a standard technique so far has been the modeling of the transverse density distributions of the two beams by fitting the evolution, during either vdM or the specifically-tailored BI scans, not only of the luminosity itself but also of the position, orientation and shape of its spatial distribution. The latter can be reflected in the distribution of the reconstructed primary vertices. Luminosity profiles can be then generated for simulated vdM or BI scans using these fitted beam parameters, and the impact of nonfactorization in the single-beam distributions is determined from the difference between the “true” luminosity from the simulated overlap integral and the “measured” luminosity from the fits to the two-dimensional simulated luminosity profiles assuming factorization.

To estimate the size of this potential bias in the November 2015 scans, the two-dimensional distributions of reconstructed vertices in the transverse plane recorded during the four BI scans are exploited [233]. Considering a BI scan in  $x$  in consecutive scan steps  $\Delta x$ , the density of the accumulated vertices for a given vertex reconstruction

resolution  $V$  can be expressed based on arbitrary bunch proton densities  $\rho_1$  and  $\rho_2$  as

$$\begin{aligned}
 \sum_n n^{vtx}(x, y; n\Delta x) &\propto \sum_n \rho_1(x, y)\rho_2(x + n\Delta x, y)\Delta x \otimes V \\
 &= \left[ \sum_n \rho_1(x, y)\rho_2(x + n\Delta x, y)\Delta x \right] \otimes V \\
 &\approx \left[ \int_{\Delta x} \rho_1(x, y)\rho_2(x + \Delta x, y) d(\Delta x) \right] \otimes V \\
 &= \rho_1(x, y)(\mathcal{M}_x \rho_2)(y) \otimes V.
 \end{aligned} \tag{4.34}$$

The  $x$  coordinate is thus integrated out and the bunch proton density of the “moving” beam appears marginalized in the direction of the scan, meaning that the two-dimensional vertex distribution constrains the bunch proton density of the beam at rest (“stationary”). Taking four scans of this kind in total, i.e., scanning one beam over the other in  $x$  and  $y$  and vice versa, proton densities are fully constrained.

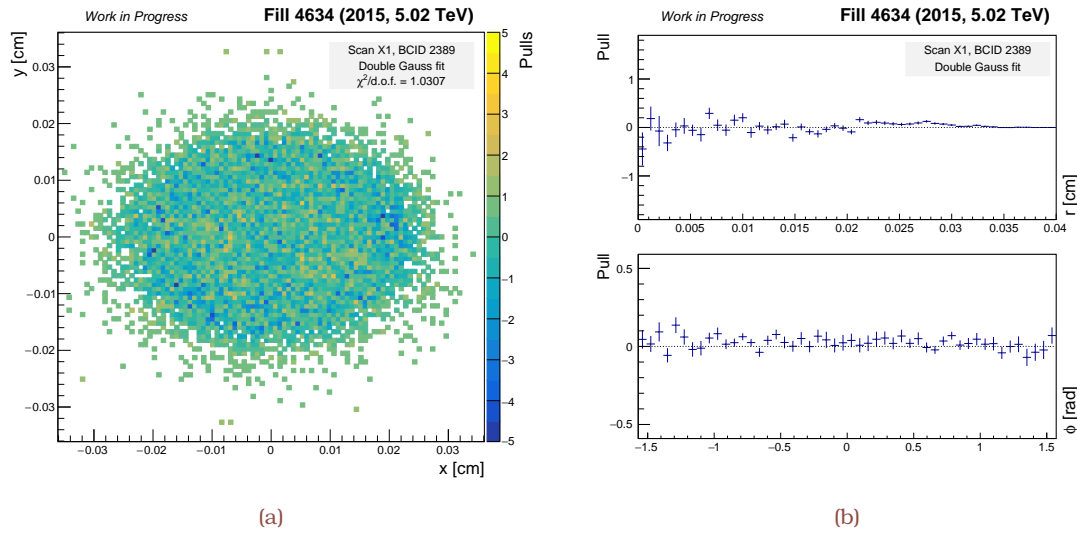
Using the fit model of Eq. (4.34) per BCID, the November 2015 analysis is performed using simultaneously all four vertex distributions accumulated during the BI scans, while the distributions of vertex positions are constructed from the first two vdM scan pairs for the November–December 2016 analysis. The bunch proton or lead densities are approximated to be of double-Gaussian type

$$\rho_i(x, y) = w_i g_W(x, y) + (1 - w_i) g_N(x, y), \tag{4.35}$$

for which the two components  $g_N$  and  $g_W$  incorporate a correlation parameter,  $r$ , and are of the form

$$g_j(x, y) = \frac{1}{2\pi\sigma_{j,x}\sigma_{j,y}\sqrt{1-r_j^2}} \exp\left(-\frac{1}{2(1-r_j^2)}\left[\frac{x^2}{\sigma_{j,x}^2} + \frac{y^2}{\sigma_{j,y}^2} - \frac{2r_jxy}{\sigma_{j,x}\sigma_{j,y}}\right]\right). \tag{4.36}$$

Simulated vdM scans with the extracted bunch densities are performed and compared with the beam overlap integrals obtained taking the nonfactorizability into account. Based on the per-BCID comparison (Table 4.3), a maximum difference of 1.4% in the visible cross section is observed in the November 2015 analysis. This difference is taken as the systematic uncertainty associated with the luminosity estimate independent of  $x$ - $y$  correlations in the proton densities. The resolution in the position of a reconstructed PV strongly depends on the number of tracks used to fit the vertex. A track-splitting method (see Section 3.2.3) is employed for measuring the vertex resolution as a function of the number of tracks associated with the vertex. For the November–December 2016 vdM conditions, the vertex resolution becomes comparable to the expected size of  $\sigma_b$  only for primary vertices using at least 50 vertex-associated tracks (Fig. 4.14). The BCID-averaged correction amounts to 2.3 (2.9)% for the Ppb (pPb) period.



**Figure 4.13:** By measuring the positions of the interaction vertices with the tracker during the dedicated BI scans, the transverse bunch densities of the two colliding beams can be reconstructed. The pull distribution (a) based on two-dimensional proton density models built from a combination of Gaussian pdf, as well as its projection to radial and angular coordinates (b), are shown using the data collected during the IX1 scan for BCID 2389 [TH3]. The pull distributions indicate a good agreement between the fitted model predictions and the recorded data.

#### 4.3.4 Beam–beam effects

When charged-particle bunches collide, the strength of the mutual electromagnetic forces of the two colliding positively-charged bunches changes as a function of separation and, if uncorrected, can bias the results obtained from the vdM scans in two separate ways. The amplitude and the beam-separation dependence of the so-called “beam-dynamical” interaction depend on the beam energy, the number of betatron oscillations and the unperturbed  $\beta$ -functions as well as the bunch intensities and transverse beam sizes. The two beams repel each other (beam–beam deflection), and their electromagnetic repulsion induces a mutual angular kick that distorts their closed orbits by a fraction of a micrometer and modulates the actual transverse separation at the IP. Figure 4.15a shows an example of the beam deflection as a function of the nominal separation, calculated analytically using the procedure from Ref. [234]. A correction to the beam–beam separation in the scan curves is applied per scan and bunch crossing accordingly, resulting in an overall correction of about 1% on the visible cross section. The uncertainty of this calculation is dominated by the  $\beta^*$  uncertainty, which is known with a precision of about 20% [235]. The uncertainty of this beam–beam correction is thus 0.2%.

Apart from the beam–beam deflection, the electromagnetic forces give rise to mutual defocusing (“dynamic- $\beta$ ”) effects of the colliding bunch densities varying as a function of their relative separation. The resulting fractional change in the value of the  $\beta$  function at the considered IP, or equivalently the optical demagnification between the LHC arcs and the collision point, modifies the collision rate at each scan step and hence distorts the

**Table 4.3:** Comparison of the effective beam overlap cross sections from the simulated vdM November 2015 scans—with bunch proton densities extracted using Eq. (4.35) to the beam images (Eq. (4.34))—against the beam overlap integrals incorporating genuine nonfactorizabilities [TH3]. Multiple simulated experiments are performed, and hence the associated statistical uncertainty is negligible.

BCID	$A_{\text{eff}}$ [mm <sup>2</sup> ]	True integral [mm <sup>2</sup> ]	$\frac{(\text{True integral} - A_{\text{eff}})}{A_{\text{eff}}}$ [%]
644	0.0641	0.0646	0.8
1215	0.0637	0.0642	0.8
2269	0.0637	0.0646	1.4
2389	0.0641	0.0649	1.2
2589	0.0639	0.0648	1.4

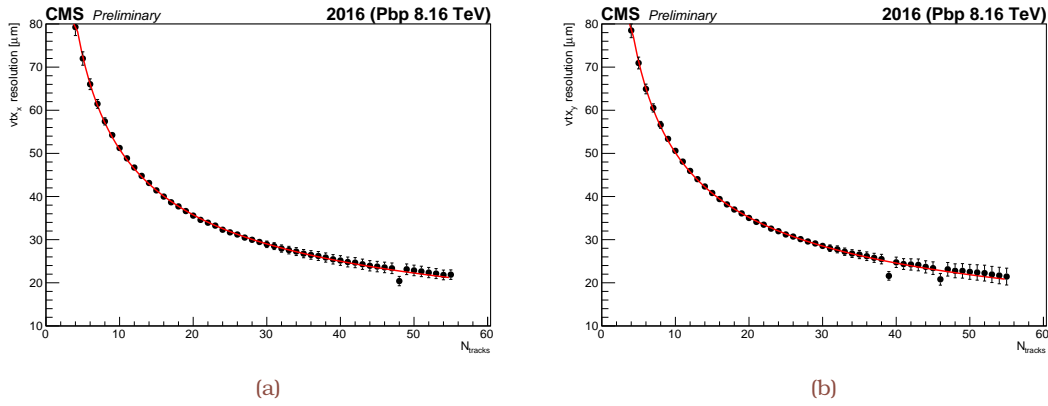
shape of the scan curve. This effect, however, is expected to be small, and is modeled (Fig. 4.15b) using the MAD-X optics code assuming bunch parameters representative of a baseline vdM scan, and then rescaled using the measured intensities and convolved beam sizes per colliding-bunch pair. An uncertainty of 0.5% is assigned to cover the possible impact on the visible cross section measurement. The overall uncertainty due to beam-beam effects is thus 0.6% in the November 2015 vdM analysis.

A conservative uncertainty due to the beam-beam effect is assigned that equals the correction, i.e., 0.3%, in the November–December 2016 vdM analysis. The dynamic- $\beta$  effect is also expected to be small in both Pbp and pPb periods, owing to the low bunch intensities, and a conservative uncertainty of 0.5% is assigned to cover the possible impact on the visible cross section measurement, and hence the overall uncertainty due to beam-beam effects is at the same level in the November–December 2016 as in the November 2015 vdM analysis.

#### 4.3.5 Bunch current corrections and normalization

The LHC beam currents are measured in a three-step process resulting from different capabilities of the available instrumentation. Since the bunch population can significantly vary from one bunch to another, we utilize the FBCT or BQM to measure the single-bunch intensities  $N_{\text{FBCT, BQM}}^j$ . The sum of the individual bunch currents is then normalized such that it matches the total current  $N_{\text{DCCT}}$  measured by the DCCT with an accuracy of 0.3%. Finally, corrections are applied to account for the spurious charge present in a given BCID.

Spurious charges can be present in the LHC, either belonging to noncolliding bunch slots at a level below the FBCT threshold (“ghosts”) or leaking out from the nominally filled RF bucket into a nearby bucket (“satellites”). Figure 4.16 shows an example distribution for the longitudinal beam charge distribution over the RF buckets of 2.5 ns length that are equidistantly distributed over the LHC circumference. Ghost contributions enter in the total beam intensity measurement from DCCT, while satellite charges contribute to the per-bunch-slot granularity current measurement without resulting in a measurable lumi-



**Figure 4.14:** Primary vertex transverse resolution in the horizontal (a) and vertical (b) direction using zero-bias events in November–December 2016 vdM analysis as a function of the track multiplicity ( $N_{\text{tracks}}$ ). The reconstructed tracks associated with primary vertices are selected based on the high-purity requirement (Table 3.1), while vertices within a radial (longitudinal) distance of 24 (2) cm are retained. After the track-splitting, the low event count at higher  $N_{\text{tracks}}$  causes fit instabilities. The curve shows the results of the resolution fitted using a three-parameter function of the form  $A/N_{\text{tracks}}^B + C$  [TH2]. Errors are retained from the covariance matrix of the fit applied after the track-splitting procedure.

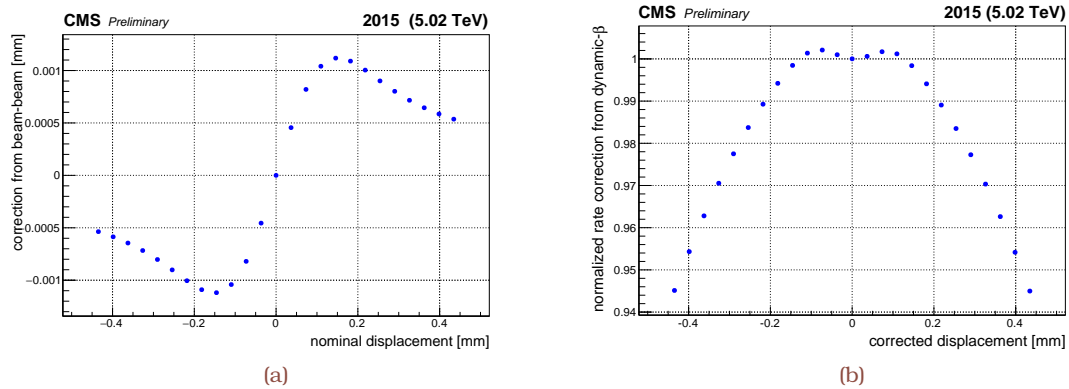
osity [236]. This may introduce an error in the measurement of the bunch populations and, via the bunch population product, in the determination of the luminosity calibration factor. Therefore, the FBCT or BQM measurements of the bunch population is also corrected according to the following formula

$$N_{1,2}^j = N_{\text{FBCT, BQM}}^j (1 - f_{\text{sat}}^j) \frac{N_{\text{DCCT}} (1 - f_{\text{ghost}})}{\sum_j N_{\text{FBCT, BQM}}^j}, \quad 4.37$$

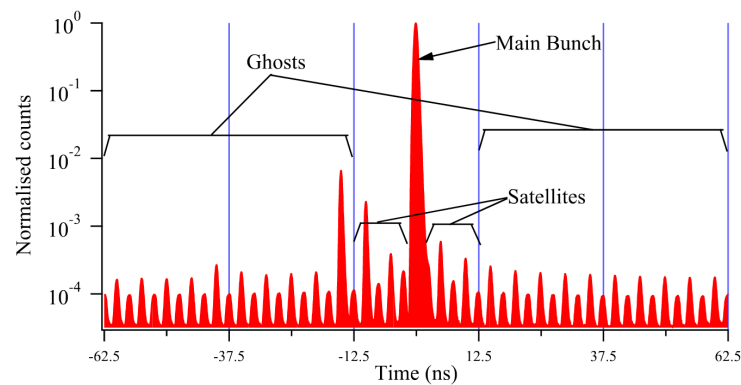
with  $f_{\text{ghost}}$  and  $f_{\text{sat}}^j$  the fraction of ghost charge out of the total beam current and the fraction of satellite charge in bunch  $j$ , respectively. The spurious charge is measured by LDM.

The LDM data for Fill 4634 indicate that the satellite charges were insignificant, below 0.002% for both beams, and have no detectable effect on our measurement. The ghost charges, however, present nonnegligible values and, in particular for beam 1, show a very strong time dependence that has to be taken into account. The ghost contributions are estimated to be constant at  $\sim 0.3\%$  for beam 2, while for beam 1 they increase from 0.8% to 2.4% between the first vdM and the last BI scan. A measurement of ghost charge is provided independently by the LHCb Collaboration, via the rate of beam–gas collisions occurring in nominally empty bunch slots, that confirms the findings from LDM (Fig. 4.17). The ghost charges present nonnegligible values also for November–December 2016 analysis, and, in particular, for the Pb beam, show a very strong time dependence that has to be taken into account. The ghost contributions are estimated to be constant at  $\sim 0.5$  ( $\sim 0.3$ )% for the proton beam in the Pbp (pPb) period, while for the Pb beam they increase from 1.0 to 1.2 (1.0 to 1.3)% between the first and the last vdM scan in Pbp (pPb)





**Figure 4.15:** Example of the intensity-dependent orbit kicks and tune shifts induced by long-range beam–beam interactions distorting (a) the nominal beam displacement and (b) the luminosity scan curves [TH3].



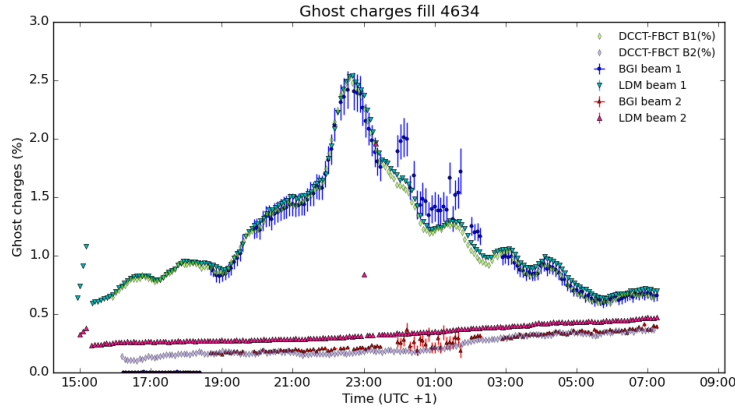
**Figure 4.16:** Longitudinal beam profile monitoring of beam 1 at LHC as seen by LDM [236], which measures synchrotron radiation photons emitted by the beams. The RF-bucket structure is split into nominally filled, satellite, and ghost bins. Conventionally, RF buckets that lie within the 12.5 ns range around the center of a nominally filled bunch are considered as satellite bunches, while buckets accommodated outside this range are lumped altogether in the ghost charge.

period.

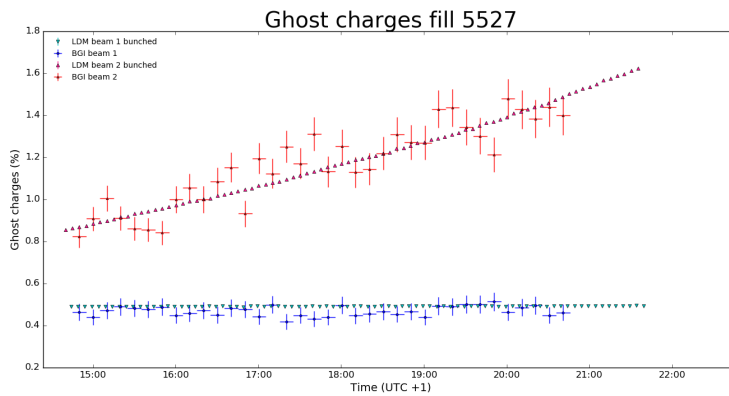
The correction to the overall visible cross section is  $1.8 \pm 0.2\%$  in the November 2015 analysis, and is found to be  $1.3 \pm 0.7$  ( $1.2 \pm 0.6$ )% in the Pbp (pPb) period. The uncertainty safely takes into account the maximum observed difference between the measurements from LDM and LHCb. The DCCT accuracy of 0.3% is taken as an additional systematic uncertainty.

## 4.4 Cross-detector stability

Given that the absolute calibration of the applied algorithms is carried out under specially tailored beam conditions, the comparison of the relative response of several independent luminometers during routine physics conditions reveals possible time- and rate-dependent effects. The stability of the PCC response has been investigated based



(a)



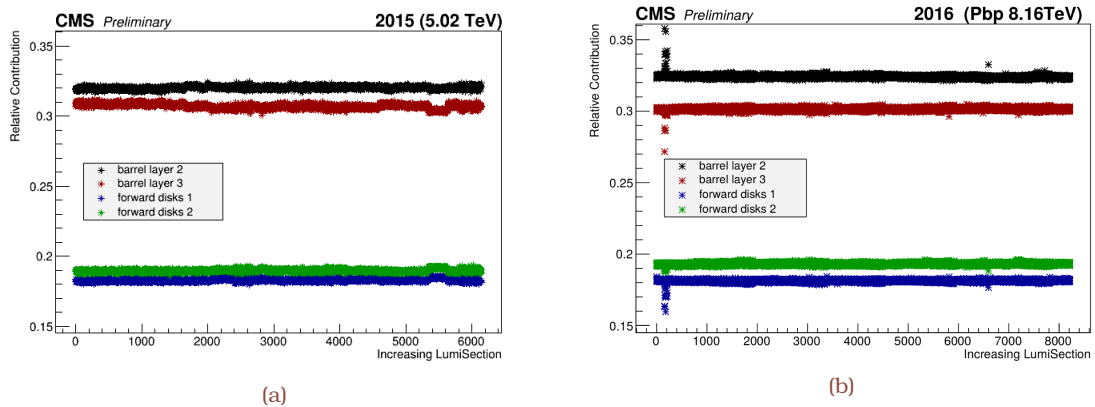
(b)

**Figure 4.17:** Ghost-charge fractions separately for beam 1 and 2 as a function of time for Fills 4634 (a) and 5527 (b) comparing the measurements based on the number of beam-gas interactions (BGI) recorded by LHCb [237] and the longitudinal profile of the synchrotron radiation monitors. The total systematic uncertainty in BGI measurements stems from the trigger efficiency that is fully correlated between the beams. The unbunched population, i.e., the signal level at RF edges, is not subtracted from LDM measurements. The FBCT—normalized to the first valid LHCb ghost-charge value per beam—is subtracted from the DCCT measurement and is superimposed in (a).

on the relative contributions of the different layers of the pixel detector to the total PCC response over the November 2015 data-taking period. Excellent stability and independence to various data-taking conditions of the different layers relative to the total response within 0.5% are found as shown in Fig. 4.18a. The long-term PCC response has been independently monitored and compared to the rate of the DT luminometer over the entire 2015 data-taking period at  $\sqrt{s} = 13$  TeV. The RMS of 1% for the per luminosity section ratio is conservatively assigned as systematic uncertainty due to the stability of the PCC rate in the November 2015 period [217]. If the luminous region moves significantly, then the acceptance of the pixel detector slight changes, and the effect appears as a change in the visible cross section. In 2015, the luminous region has been always within  $\approx \pm 4$  cm from the origin in the longitudinal direction (Fig. 3.13). Within this range, changes in acceptance are negligible [65, 217].

For the November–December 2016 period, the RMS of 0.9 (0.7)% of the PLT over HF

ratios as a function of the single-bunch instantaneous luminosity is considered as the systematic uncertainty related to the stability of the luminometer rate in the Pbp (pPb) period, as displaced in Fig. 4.19. The stability has been independently investigated using the silicon pixel detector; the relative contributions of the different layers of the pixel detector to the total response is shown in Fig. 4.18b. Excellent stability of the layers and independence to various data-taking conditions is indeed found.

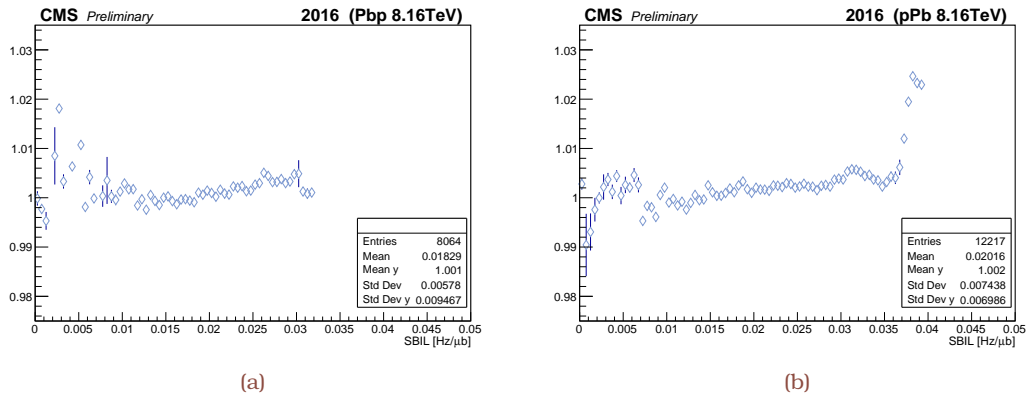


**Figure 4.18:** The relative contributions in percentage of the different pixel layers for the entire (a) November 2015 [TH3] and (b) the first part (Pbp) of November–December 2016 [TH2] data-taking periods. The innermost pixel layer is excluded from the PCC rate measurements. Only periods where the CMS detector is fully operational are considered. The last part of the Pbp period, corresponding to LHC Fill 5538 with active leveling at IP5 and contributing negligibly to the total luminosity, has also been excluded.

## 4.5 Conclusions

Table 4.4 summarizes the derived corrections along with their associated systematic uncertainty. The various sources are grouped into i) “normalization” uncertainty; that is, the luminosity scale calibration from the vdM scan procedure, and ii) “integration” uncertainty; that is, the uncertainty associated with the extrapolation of  $\sigma_{\text{vis}}$  to luminometer rate measurements under usual data-taking conditions. The uncertainties in the integration and normalization are treated as uncorrelated, and are summed in quadrature. An uncertainty of 0.5% associated to the deadtime estimate from the CMS DAQ system is additionally included, affecting the recorded integrated luminosity. The dominant uncertainty contributing to the luminosity scale calibration is associated to the nonfactorizability of the colliding beam bunch densities.

The weighted averages of  $\sigma_{\text{vis}}$  for all BCIDs measured over the considered scans in the Pbp relative to the pPb period, corrected for effects impacting the accuracy of the absolute and relative bunch populations, and for beam–beam interactions, is found to be in excellent agreement for the symmetric luminometers. A combined result is then determined by computing a weighted average from the  $\sigma_{\text{vis}}$  calibrations in the two periods. To avoid underestimating the uncertainty in the combined result it is assumed that most sources of systematic uncertainty are fully correlated except for the uncertainty in bunch-



**Figure 4.19:** Comparison of the PLT over HF luminosity ratio against the single-bunch instantaneous (SBIL) luminosity, the latter estimated based on PLT for the Pbp (a) and pPb (b) periods [TH2]. The last part of the Pbp period, corresponding to LHC Fill 5538 with active levelling at IP5, has been excluded, while the rising trend for SBIL higher than about 0.035 Hz/μB includes less than 2% of the integrated luminosity.

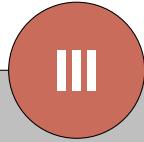
to-bunch and scan-to-scan variations, and cross-detector stability.

In summary, the total uncertainty in the CMS luminosity measurement in November 2015 is 2.3% using proton-proton collisions at  $\sqrt{s} = 5.02$  TeV, and is estimated to be 3.5% for data recorded with proton-nucleus collisions at  $\sqrt{s_{NN}} = 8.16$  TeV in November–December 2016. The results can be compared to those obtained from the ATLAS [238] (proton-proton), and ALICE [239] and LHCb [240] (proton-nucleus) Collaborations.

**Table 4.4:** Summary of the systematic uncertainty in the CMS luminosity measurement using proton-proton and proton-nucleus collisions at  $\sqrt{s} = 5.02$  and  $\sqrt{s_{\text{NN}}} = 8.16$  TeV, respectively. When applicable, the percentage correction is shown.

	Source	Correction (%)			uncertainty (%)		
		pp	Pbp	pPb	pp	Pbp	pPb
Normalization	Beam nonfactorizability	-	-	-	1.4	2.3	2.9
	Bunch-to-bunch variation	-	-	-	-	1.4	1.5
	Scan-to-scan variation	-	-	-	-	0.6	1.0
	Length scale	1.0	-	-	0.2	0.7	0.7
	Ghosts and satellites	1.8	1.3	1.2	0.2	0.7	0.6
	Orbit Drift	-	-	-	0.4	0.5	0.5
	$\Sigma_{x,y}$ compatibility	-	-	-	-	0.5	0.2
	Dynamic- $\beta$ effect	-	-	-	0.5	0.5	0.5
	Beam-beam deflection	1	0.3	0.3	0.2	0.3	0.3
Beam current calibration	-	-	-	0.3	0.3	0.3	
Integration	Cross-detector stability	-	-	-	1	0.9	0.7
	Type 1/2	7	-	-	1	-	-
	DAQdeadtime	-	-	-	0.5	0.5	0.5
	Dynamic inefficiency	-	-	-	0.4	-	-
	Total				2.3	3.2	3.7





Part C





# Chapter 5: Measurement of the $t\bar{t}$ cross section in LHC pp collisions at 5.02 TeV

## Contents

<b>5.1 Filling the gap between Tevatron and LHC energies</b> . . . . .	<b>142</b>
<b>5.2 Data, simulated samples and theoretical cross section</b> . . . . .	<b>143</b>
<b>5.3 Event selection</b> . . . . .	<b>144</b>
<b>5.4 Background estimation</b> . . . . .	<b>147</b>
5.4.1 The $\ell$ +jets final state . . . . .	147
5.4.2 The dilepton final state . . . . .	153
<b>5.5 Systematic uncertainty</b> . . . . .	<b>157</b>
5.5.1 Experimental uncertainty . . . . .	157
5.5.2 Theory uncertainty . . . . .	158
<b>5.6 Measurement of the <math>t\bar{t}</math> cross section</b> . . . . .	<b>161</b>
5.6.1 The $\ell$ +jets final state . . . . .	161
5.6.2 The dilepton final state . . . . .	165
5.6.3 Combination of the $\ell$ +jets and dilepton final states . . . . .	168
<b>5.7 QCD analysis</b> . . . . .	<b>171</b>

This Chapter reports the top quark pair production cross section ( $\sigma_{t\bar{t}}$ ) measured, for the first time, in pp collisions at a center-of-mass energy of 5.02 TeV. The data were collected by the CMS experiment and correspond to an integrated luminosity of  $27.4 \text{ pb}^{-1}$ . The measurement is performed by analyzing events with at least one charged lepton. The measured cross section is  $\sigma_{t\bar{t}} = 69.5 \pm 6.1 \text{ (stat)} \pm 5.6 \text{ (syst)} \pm 1.6 \text{ (lum)} \text{ pb}$ , in agreement with the expectation from the standard model. With a total relative uncertainty of 12% in the final result the impact of the presented measurement on the determination of the gluon distribution function is also investigated at the next-to-next-to-leading order in perturbative quantum chromodynamics.

The original contribution from the present thesis to the material related to the following Chapter and listed in [Scientific output](#) and [Internal notes](#) is:

- The first preliminary measurement in the  $e^+\mu^\mp$  final state alone.
- The updated measurement in the  $\ell$ +jets final state.
- The entire responsibility of all internal stages that the paper had to pass through before it left the Collaboration and after the interaction with the journal.

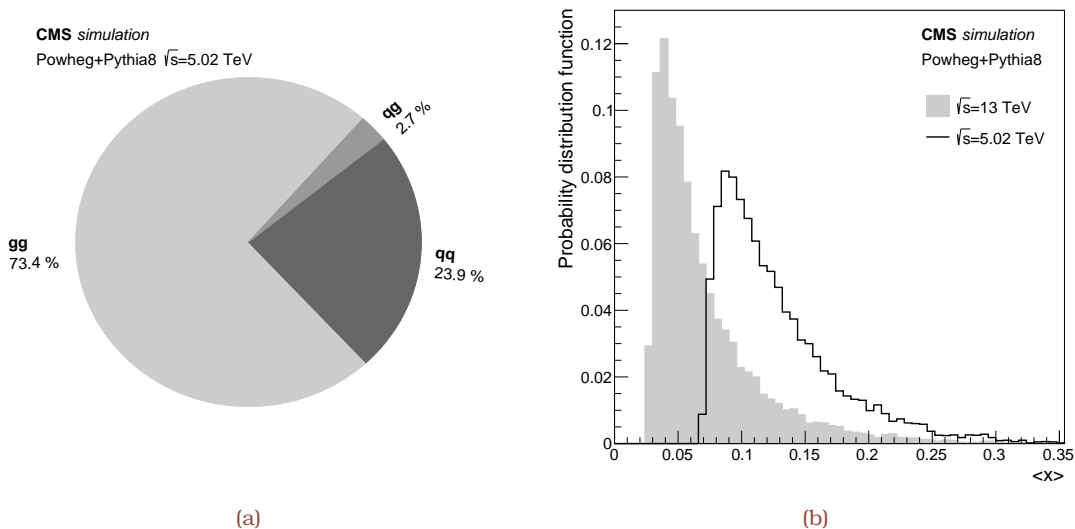
## 5.1 Filling the gap between Tevatron and LHC energies

The pair production cross section as a function of center-of-mass energy is of interest for the extraction of the top quark mass [241], and it has been used to constrain the gluon distribution function [242] at large fractions  $x$  of the proton longitudinal momentum carried by the gluon, where is poorly known and the correlation between the gluon distribution and  $\sigma_{t\bar{t}}$  is maximal [243]. Precise measurements of  $\sigma_{t\bar{t}}$  in pp collisions have been published at  $\sqrt{s}$  values of 7 and 8 [244–247] and 13 TeV [75, 248–250] by the ATLAS and CMS Collaborations at the LHC.

The pp collisions at  $\sqrt{s} = 5.02$  TeV in November 2015 (Table 2.2) offered a unique chance to fill the gap in  $\sqrt{s}$  between Tevatron and LHC. The fraction of  $t\bar{t}$  events initiated by gluon-gluon collisions grows monotonically with  $\sqrt{s}$ . It is around 73% at 5.02 TeV, as calculated with POWHEG (v2) [39, 40] at NLO using the NNPDF3.0 NLO [35] parton distribution functions, and increases to around 86% at 13 TeV (Fig. 5.1), making this new data set partially complementary to the higher-energy samples. Measurements of  $t\bar{t}$  production at various  $\sqrt{s}$  probe different values of  $x$  and thus can provide complementary information on the gluon distribution. In addition, future measurements of  $\sigma_{t\bar{t}}$  in nuclear collisions at the same nucleon-nucleon center-of-mass energy [87, 251] would profit from the availability of a reference measurement in pp collisions at  $\sqrt{s} = 5.02$  TeV, without the need to extrapolate from measurements at different  $\sqrt{s}$ . This has already been demonstrated with the first observation of the  $t\bar{t}$  process using proton-nucleus collisions at a higher nucleon-nucleon center-of-mass energy [TH1].

In the SM, top quarks in pp collisions are mostly produced as  $t\bar{t}$  pairs. Each top quark decays predominantly to a W boson and a bottom quark. The  $t\bar{t}$  events are categorized according to the decay of the two W bosons. In  $t\bar{t}$  events where one W boson decays leptonically and the other hadronically ( $\ell$ +jets final state), the final state presents a typical signature of one isolated lepton, missing transverse momentum, two jets from the W boson hadronic decay, and two jets coming from the hadronization of the b quarks. On the other hand, in  $t\bar{t}$  events where both W bosons decay leptonically (dilepton final state), the final state contains two leptons of opposite electric charge, missing transverse momentum, and at least two b jets. The  $\ell$ +jets final state has a large branching ratio with a moderate amount of background, while a high purity characterizes the dilepton final state, thus compensating for its smaller branching ratio.

This analysis represents the first measurement of  $\sigma_{t\bar{t}}$  in pp collisions at  $\sqrt{s} = 5.02$  TeV using  $t\bar{t}$  candidate events with  $\ell$ +jets, where leptons are either electrons ( $\ell = e$ ) or muons ( $\ell = \mu$ ), and dilepton ( $e^\pm\mu^\mp$  or  $\mu^\pm\mu^\mp$ ) final states. In the former case,  $\sigma_{t\bar{t}}$  is extracted by a fit to the distribution of a kinematic variable for different categories of lepton flavor and jet multiplicity, while in the latter an event counting approach is used. The two results are then combined in the final measurement, which is used as input to a quantum chromodynamics analysis at next-to-next-to-leading order to investigate the impact on the determination of the gluon distribution in the less-explored kinematic range of  $x \gtrsim 0.1$ .



**Figure 5.1:** (a) Representative fractions (in %) of initial partonic states yielding  $t\bar{t}$  pairs in pp collisions at  $\sqrt{s} = 5.02$  TeV simulated at NLO with POWHEG (v2) and the NNPDF3.0 NLO PDF set. (b) Expected average momentum  $\langle x \rangle$  of the incoming partons measured with respect to the proton momentum, for pp collisions at  $\sqrt{s} = 5.02$  and 13 TeV [TH15].

## 5.2 Data, simulated samples and theoretical cross section

This analysis is based on an integrated luminosity of  $27.4 \pm 0.6 \text{ pb}^{-1}$  [TH3]. The presence of multiple proton collisions in the same or nearby bunch crossings results in an average number of overlapping interactions estimated online to be 1.4, assuming a total inelastic cross section of 65 mb.

Several MC event generators are used to simulate signal and background events. The NLO POWHEG (v2) [39, 40, 57] generator is used for  $t\bar{t}$  events, assuming a value of 172.5 GeV for  $m_{\text{top}}$ . These events are passed to PYTHIA (v8.205) [49, 51] to simulate parton showering, hadronization, and the underlying event, using the CUETP8M1 [58, 59] tune for the default  $t\bar{t}$  MC sample. The NNPDF3.0 NLO PDFs with strong coupling  $a_s(M_Z) = 0.118$  at the Z boson mass scale  $M_Z$  are utilized in the MC calculations.

The MADGRAPH5\_aMC@NLO (v2.2.2) generator [38] is used to simulate W boson production with additional jets (W+jets), and high-mass ( $> 50$  GeV) Drell-Yan quark-antiquark annihilation into lepton-antilepton pairs through Z boson or virtual-photon exchange (referred to as “Z/ $\gamma^*$ ”). The simulation includes up to two extra partons at matrix element level, and the FxFx merging scheme [43] is used to interface with PYTHIA. Low-mass Z/ $\gamma^*$  events (20–50 GeV) are simulated with PYTHIA. The normalization of the W+jets and Z/ $\gamma^*$  processes is either derived from data (in the dilepton final state) or estimated based on the NNLO cross sections (in the  $l$ +jets final state) from the FEWZ program (v3.1.b2) [252]. Single top quark plus W boson events (tW) are simulated using POWHEG (v1) [253, 254] interfaced with PYTHIA, and are normalized to the approximate NNLO cross sections [255]. The diboson contributions from WW and WZ production (referred to as “WV”) are simulated with PYTHIA, and are normalized to the NLO cross sections calculated with the MCFM (v8.0) program [256]. All generated events undergo a full GEANT4 [194] simulation of

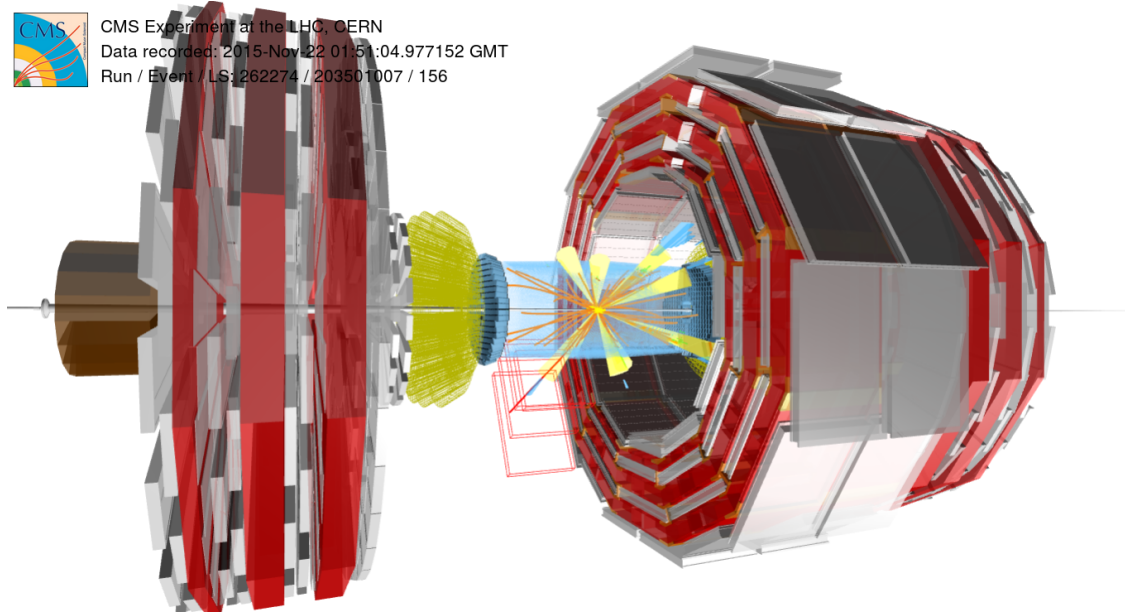
the detector response. A complete overview of all used simulated sample can be found in Table 5.1.

The expected signal yields are normalized to the value of the SM prediction for the  $t\bar{t}$  production cross section:

$$\sigma^{\text{NNLO}} = 68.9^{+1.9}_{-2.3} (\text{scale}) \pm 2.3 (\text{PDF})^{+1.4}_{-1.0} (a_s) \text{ pb}, \quad \text{5.1}$$

as calculated with the TOP++ program [257] at NNLO in perturbative QCD, including soft-gluon resummation at next-to-next-to-leading-logarithmic order [258], using the NNPDF3.0 NNLO PDF set, with  $a_s(M_Z) = 0.118$  and  $m_{\text{top}} = 172.5$  GeV. The systematic uncertainty in the theoretical  $t\bar{t}$  cross section are associated with the choice of the renormalization ( $\mu_R$ ) and factorization ( $\mu_F$ ) scales—nominally set at  $\mu_R = \mu_F = \sqrt{m_{\text{top}}^2 + p_{T,\text{top}}^2}$  with  $p_{T,\text{top}}$  the top quark transverse momentum—as well as with the PDF set and the  $a_s$  value. The uncertainty of 0.1% in the LHC beam energy [259] translates into an additional uncertainty of 0.22 pb in the expected cross section, with negligible impact on the acceptance of any of the final states included in this analysis.

### 5.3 Event selection



**Figure 5.2:** A pp collision featuring a  $t\bar{t} \rightarrow e^\pm \mu^\mp + X$  candidate event in the  $\sqrt{s} = 5.02$  TeV data sample recorded by CMS. Calorimetric energy deposits in ECAL (HCAL) are depicted as green (blue) bars with a height proportional to the magnitude of the deposit. Tracks are represented by orange (red) lines if reconstructed from tracker (combined tracker and muon system) hits. Apart from the beam pipe, the highlighted CMS detectors are the pixel and ECAL barrel, ECAL and HCAL endcaps, HF calorimeter, and part of the muon system along with the matching chambers, CSC and DT segments. The measured kinematics of the reconstructed electron, muon, jets (yellow cones) and  $p_T^{\text{miss}}$  (not highlighted) can be found in Ref. [TH15]. The superimposed model of the CMS detector [69] is imported in the browser-based event display of Ref. [70].

**Table 5.1:** MC event generation employed in the measurement of the inclusive  $\sigma_{\bar{t}t}$  in pp collisions at  $\sqrt{s}=5.02$  TeV [AN3, AN1, AN4]; the samples are generated either inclusively or with a final state restricted to the leptonic mode, including electrons, muons, and  $\tau$  leptons. The absence of reference delineates cross section values calculated from the generator. For the samples restricted to specific decay channels, the branching ratio is included in the quoted cross section. The  $M_{\ell\ell} > 50$  GeV ( $20 < M_{\ell\ell} < 50$  GeV) requirement has been applied on the MC event sample marked with \* (\*\*). The NNPDF3.0 NLO PDFs with strong coupling  $\alpha_s(M_Z) = 0.118$  at the Z boson mass scale  $M_Z$  are utilized in all the MC calculations.

Process	$\sigma$ [pb]	MC event sample indicative description
$\bar{t}t$		TT_TuneCUETP8M1_5020GeV-powheg-pythia8
		TT_TuneCUETP8M1_5020GeV-powheg-scaledown-pythia8
		TT_TuneCUETP8M1_5020GeV-powheg-scaleup-pythia8
	69 (NNLO+NNLL) [257]	TT_TuneEE5C_5020GeV-powheg-herwig++
$tW$		ST_tw_5020GeV_top-powheg-pythia8
	3.04 (NNLL) [255]	ST_tw_5020GeV_antitop-powheg-pythia8
W+jets		WToLNU_XJ{0,1,2}_5020GeV_5f-amcatnLoFFX-pythia8
	21159 (NNLO[QCD]+NLO[EWK]) [252]	WJetsToLNU_TuneCUETP8M1_5020GeV-amcatnLoFFX-pythia8
$Z/\gamma^*$		DYJetsToLL_TuneCUETP8M1_5020GeV-amcatnLoFFX-pythia8
	2010* (NNLO[QCD]+NLO[EWK]) [252] 1506** (NNLO[QCD]+NLO[EWK]) [252]	DYJetsToLL_TuneCUETP8M1_5020GeV-MLM-pythia8
$WW(\rightarrow \ell\nu)$		WWTo2L2Nu_TuneCUETP8M1_5020GeV-powheg-pythia8
	1.77 (NLO) [256]	
	0.441 (NLO) [256]	
ZZ		ZZTo4L_TuneCUETP8M1_5020GeV-powheg-pythia8
WZ		WZTo3LNU_TuneCUETP8M1_5020GeV-powheg-pythia8g
WZ		WZTo3LNU_TuneCUETP8M1_5020GeV-powheg-pythia8g
WZ		WZTo3LNU_TuneCUETP8M1_5020GeV-powheg-pythia8g

The event sample is selected by a loose online trigger (see Section 3.1.3) and further filtered offline to remove noncollision events, such as beam-gas interactions or cosmic rays. Collision events containing one high- $p_T$  electron (muon) candidate are selected online by requiring values of  $E_T$  ( $p_T$ ) greater than 40 (15) GeV and of  $|\eta|$  less than 3.1 (2.5). The measured trigger efficiency for each final state, relative to the final selection, is measured higher than 90%, as shown in Figs. 3.4 and 3.6.

In the  $\ell$ +jets analysis, electron candidates are selected if they have  $p_T > 40$  GeV and  $|\eta| < 2.1$ . Loose identification, including isolation, criteria (Table 3.2) are applied to the electron candidates. Electron candidates in the  $1.44 < |\eta| < 1.57$  region, i.e., in the transition region between the barrel and endcap sections of the ECAL, are excluded. The electron reconstruction in this region is less efficient, and hence the loss of signal acceptance is minimal. Muons are required to be of tight quality type (see Section 3.3.2) and to have  $p_T > 25$  GeV and  $|\eta| < 2.1$ . Additional identification criteria are applied and  $I_{\text{rel}}$  is required to be  $< 15\%$ . Events are rejected if they contain extra muons (electrons) identified using looser (veto) set of identification criteria and have  $p_T > 10$  ( $> 15$ ) GeV

The distinct signature of two b jets, expected in  $t\bar{t}$  decays, is rare in background events, and this is exploited in the  $\ell$ +jets analysis. Backgrounds from W+jets, QCD multijet, and  $Z/\gamma^*$  events are handled counting the number of b-tagged jets in the selected events. In addition, two light-flavor jets are expected to be produced in the decay of one of the W bosons for signal events. The correlation of these light jets carries a distinctive hallmark with respect to the main backgrounds. To that end, jets are selected if they have  $p_T > 30$  GeV and  $|\eta| < 2.4$ . The flavor of the jets is identified using the CSVv2 algorithm with an operating point that yields a b jet identification efficiency of about 70%, and mistag probabilities of about 1% and 15% for light-flavor and c jets, respectively (Table 3.5). The event selection requires at least two non-b-tagged jets, i.e., failing the CSVv2 L working point, to be identified as candidates from the W boson hadronic decay. Additional jets passing the b quark identification criteria are counted and used to classify the selected events in none (0b), exactly one (1b), or at least two ( $\geq 2$ b) tagged jet categories. The efficiency of the b jet identification algorithm, as expected from Fig. 3.27, is measured *in situ*, simultaneously with the signal cross section.

Dilepton events are required to contain at least one muon candidate at L3 trigger level. No requirement on the presence of electron candidates is made at neither L1 nor HLT level owing to the relatively high- $E_T$  threshold (40 GeV) of the lowest unprecaled trigger. Electrons are selected offline if they have  $p_T > 20$  GeV,  $|\eta| < 2.4$ , and  $I_{\text{rel}} < 9$  (or 12)% if in the barrel (or one of the endcaps). As in the  $\ell$ +jets final state, electrons detected in the transition region between the barrel and endcap sections of the ECAL are excluded. Muons are required to have  $p_T > 18$  GeV,  $|\eta| < 2.1$ , and  $I_{\text{rel}} < 15\%$ . At least two jets satisfying the criteria  $p_T > 25$  GeV and  $|\eta| < 3$  are required. Events are subsequently selected if they have a pair of leptons with opposite charge ( $e^\pm\mu^\mp$  or  $\mu^\pm\mu^\mp$ ) passing the requirements listed above, as illustrated in Fig. 5.2. In events with more than one pair of leptons passing the above selection, the two leptons of opposite charge that yield the highest scalar  $p_T$  sum are selected.

Candidate events with dilepton invariant masses of  $M_{\ell^+\ell^-} < 20$  GeV are removed to

suppress events from decays of heavy-flavor resonances and low-mass  $Z/\gamma^*$  processes. Dilepton events with two muons in the final state are still dominated by the  $Z/\gamma^*$  background. In order to suppress this contribution, events in the Z boson mass window of  $76 < M_{\ell^\pm \ell^\mp} < 106$  GeV are vetoed in this final state. To further suppress the  $Z/\gamma^*$  events a requirement on  $p_T^{\text{miss}}$  of  $> 35$  GeV is imposed (Fig. 3.29d).

In both the  $\ell$ +jets and dilepton analyses, events with  $\tau$  leptons are considered as signal if they decay to electrons or muons that satisfy the selection requirements, and are included in the simulation.

Table 6.2 summarizes the basic selection requirements that are applied in the  $\ell$ +jets and dilepton analysis.

**Table 5.2:** Basic selection criteria applied in the three separate final states considered in the first measurement of the inclusive  $\sigma_{t\bar{t}}$  in pp collisions at  $\sqrt{s}=5.02$  TeV. Events in the mass range of  $76 < M_{\ell^\pm \ell^\mp} < 106$  GeV are additionally vetoed in the  $\mu^\pm \mu^\mp$  final state. Jets are reconstructed from the PF candidates using the anti- $k_T$  clustering algorithm with a distance parameter of 0.4.

Filter or physics object	$\ell$ +jets	Dilepton
Trigger	one $\mu$ (e) candidate, $p_T > 15$ ( $E_T > 40$ ) GeV	one $\mu$ candidate, $p_T > 15$ GeV
Leptons	exactly one $\mu$ or e	$e^\pm \mu^\mp$   $\mu^\pm \mu^\mp$
$M_{\ell^\pm \ell^\mp}$	—	$> 20$ GeV
Z veto	—	—   Yes
$p_T^{\text{miss}}$	—	—   $> 35$ GeV
Jet multiplicity		$\geq 2$
b tagging	CSVv2L (Table 3.5) $\Rightarrow$ 1 $\ell$ 2j0b, 1 $\ell$ 2j1b, and 1 $\ell$ 2j2b	

In the following, pre-fit control distributions for different variables in events with 0, 1 or at least 2 b-tagged jets and two non-b-tagged jets in the  $\ell$ +jets final state are shown. The lepton  $p_T$  and  $\eta$  are displayed in Figs. 5.3 and 5.4, respectively. The shapes of the distributions are overall reasonably reproduced as well as the rates prior to any fit. The leading jet  $p_T$  and  $\eta$  distributions are shown in Fig. 5.5 and 5.6, accordingly. The scalar  $p_T$  sum of the selected jets ( $H_T$ ) in the event is shown in Fig. 5.7. A slight trend of the jet  $p_T$  is observed in events without b-tagged jets that are dominated by the W+jets background process.

## 5.4 Background estimation

### 5.4.1 The $\ell$ +jets final state

In the  $\ell$ +jets analysis, the contributions of all background processes are estimated from simulation, except for the QCD multijet background. Owing to its large cross section, there is a nonnegligible contribution from the latter faking a  $t\bar{t}$  event with  $\ell$ +jets in the final state. Both the contribution from hard fragmentation of c and b quarks whose hadrons decay semileptonically, and the contribution from misidentified leptons, such

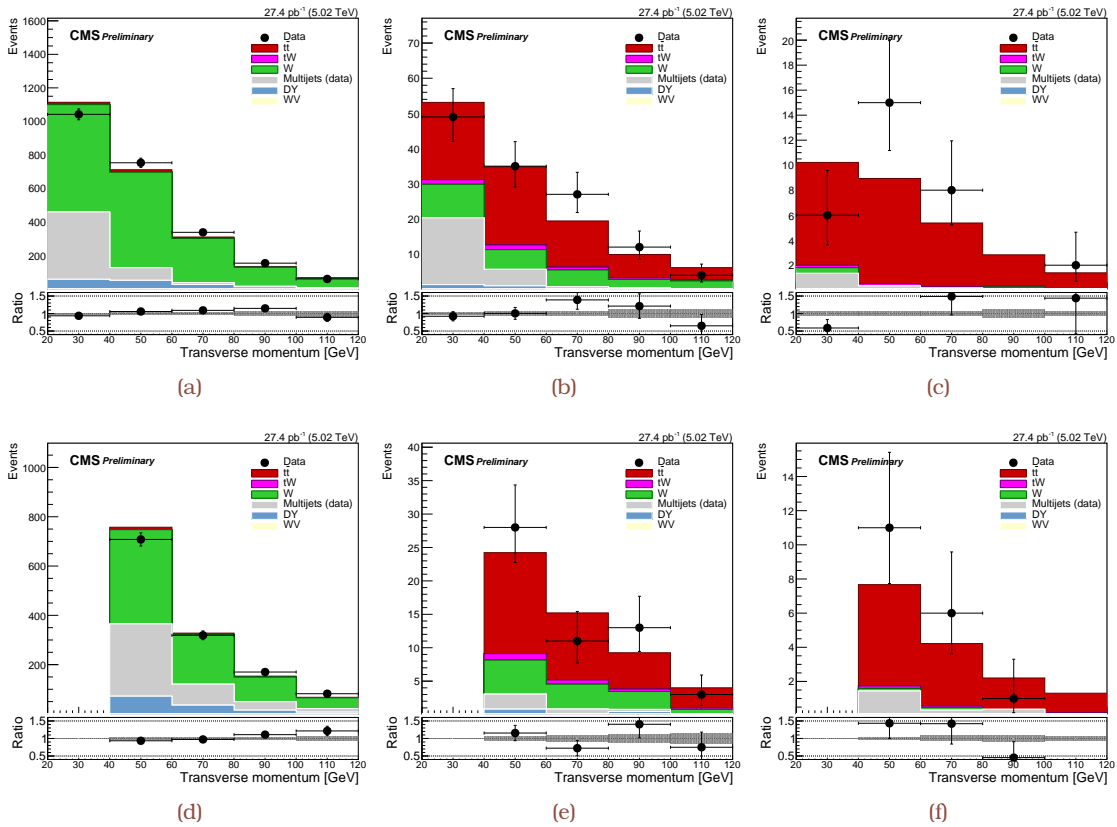
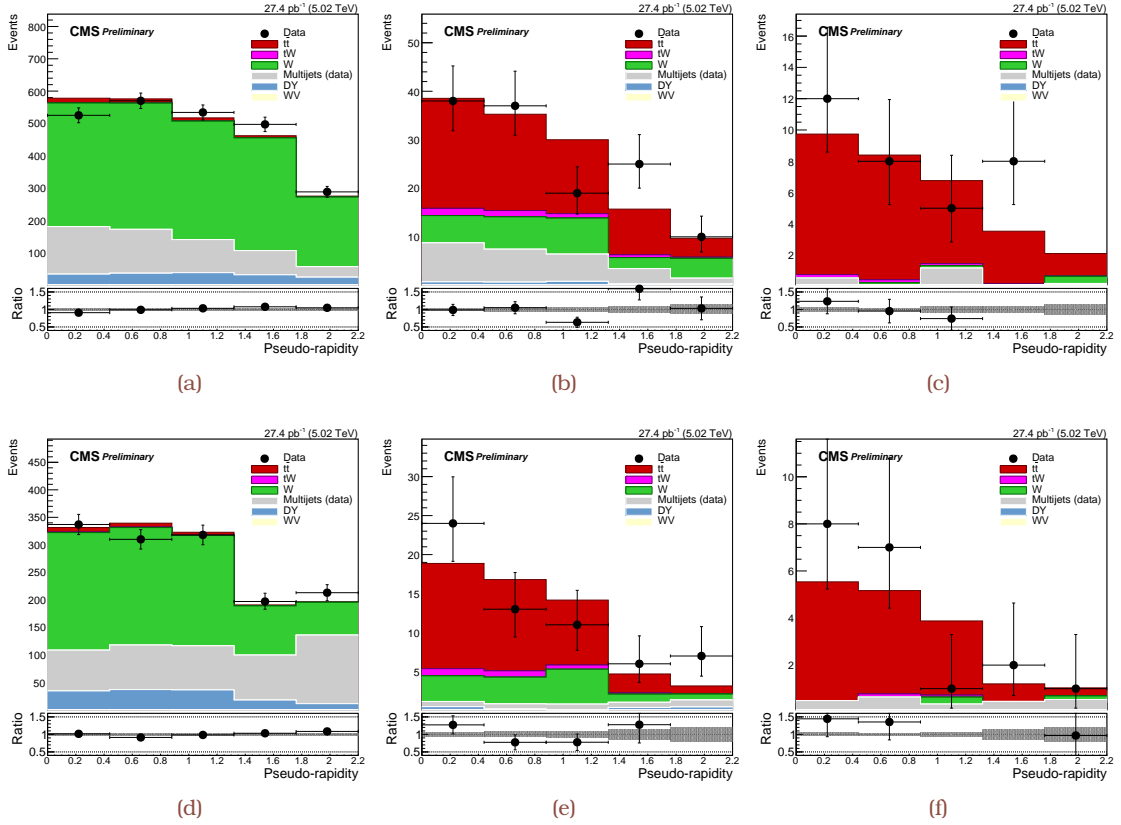
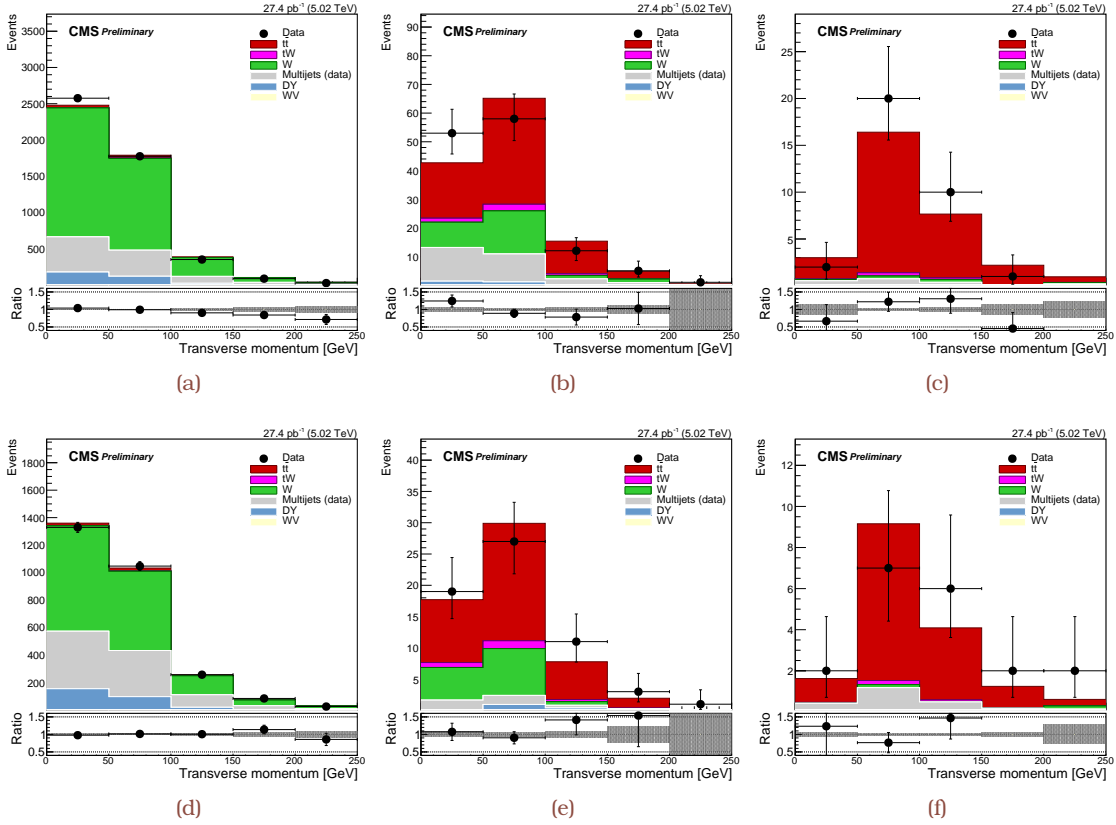


Figure 5.3: Distributions for the lepton  $p_T$  in  $\mu$ +jets (a–c) and  $e$ +jets (d–f) events with 0 (a, d), 1 (b, e) or at least 2 (c, f) b-tagged jets, using pp collisions at  $\sqrt{s} = 5.02$  TeV. The distributions from data (Table B.1) are compared to the sum of the expectations for the signal ( $t\bar{t}$ ) and backgrounds (Table 5.1). The QCD multijet background is estimated from data (see Section 5.4.1). The ratio between data and simulated distributions prior to any fit is shown at the bottom of each figure. The hatched areas indicate the statistical uncertainty due to the finite event count in the simulated samples [AN1].

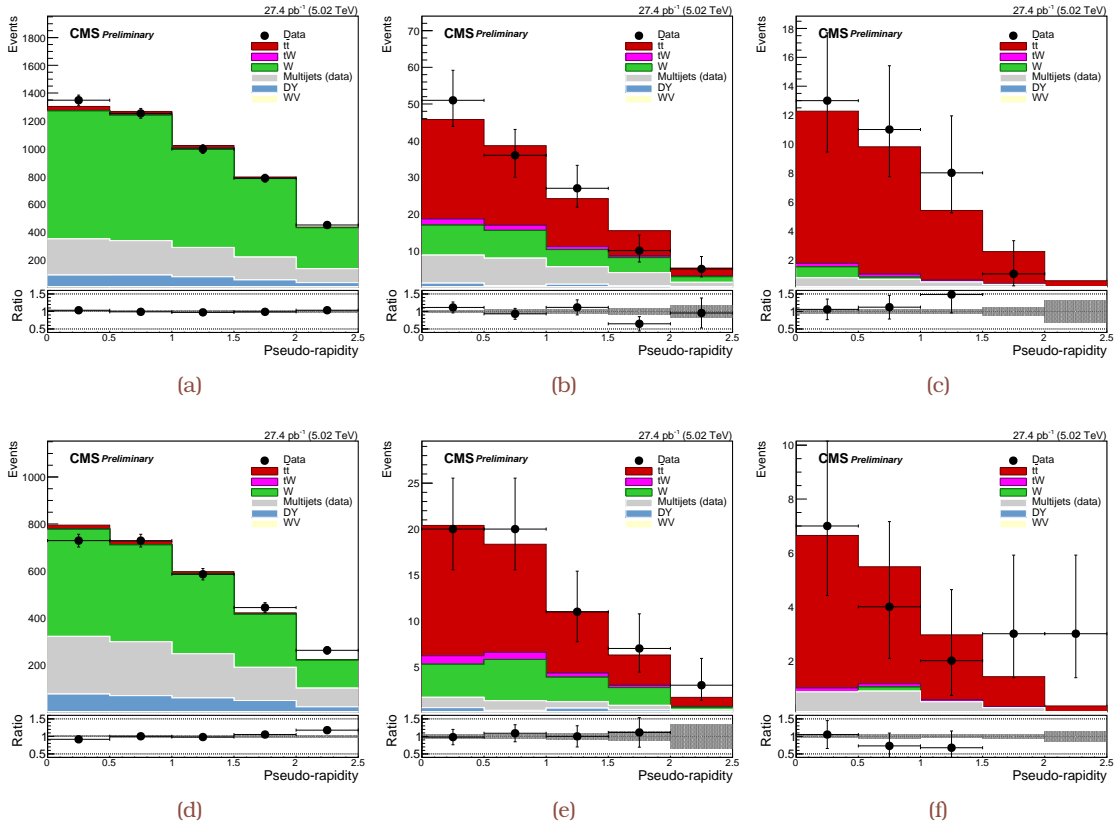




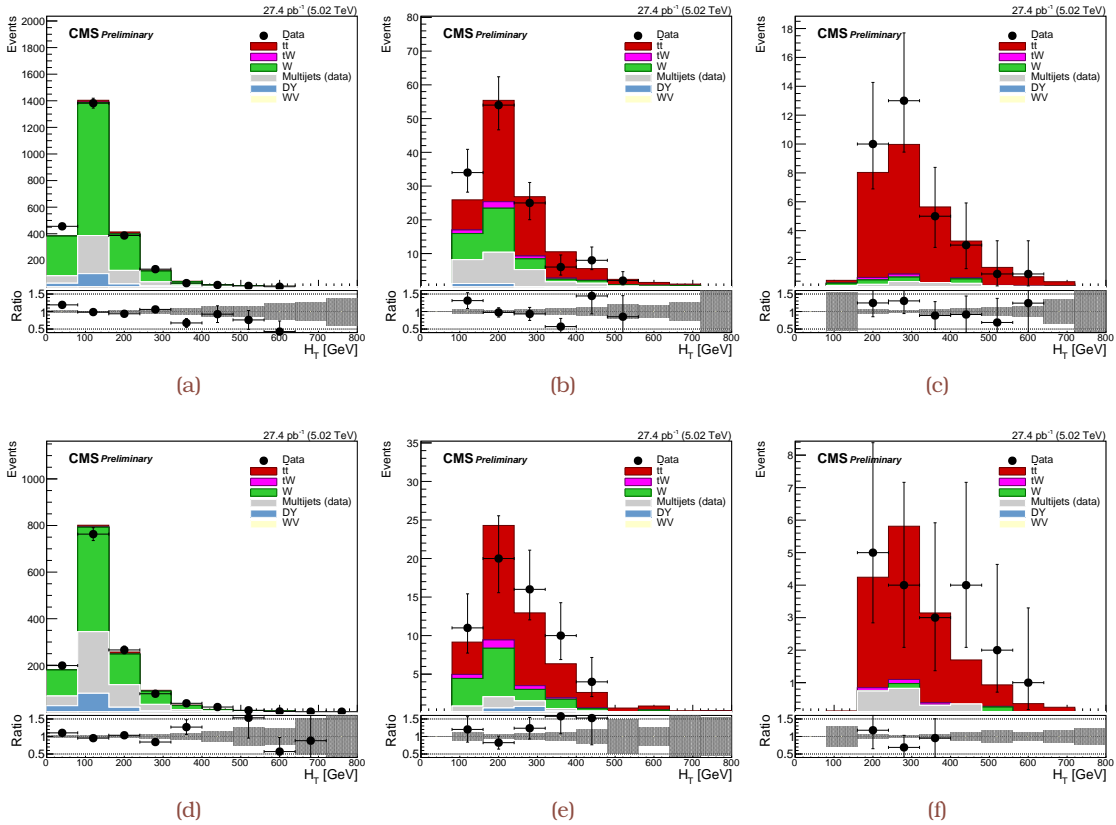
**Figure 5.4:** Distributions for the lepton pseudorapidity in  $\mu$ +jets (a–c) and  $e$ +jets (d–f) events with 0 (a, d), 1 (b, e) or at least 2 (c, f)  $b$ -tagged jets, using pp collisions at  $\sqrt{s} = 5.02$  TeV. The distributions from data (Table B.1) are compared to the sum of the expectations for the signal ( $t\bar{t}$ ) and backgrounds (Table 5.1). The QCD multijet background is estimated from data (see Section 5.4.1). The ratio between data and simulated distributions prior to any fit is shown at the bottom of each figure. The hatched areas indicate the statistical uncertainty due to the finite event count in the simulated samples [AN1].



**Figure 5.5:** Transverse momentum distributions for the two leading in  $p_T$  jets (a, d) or the leading in  $p_T$  b-tagged jet (b, c, e, f) in  $\mu$ +jets (a–c) and  $e$ +jets (d–f) events with 0 (a, d), 1 (b, e) or at least 2 (c, f) b-tagged jets, using pp collisions at  $\sqrt{s} = 5.02$  TeV. The distributions from data (Table B.1) are compared to the sum of the expectations for the signal ( $t\bar{t}$ ) and backgrounds (Table 5.1). The QCD multijet background is estimated from data (see Section 5.4.1). The ratio between data and simulated distributions prior to any fit is shown at the bottom of each figure. The hatched areas indicate the statistical uncertainty due to the finite event count in the simulated samples [AN1].



**Figure 5.6:** Pseudorapidity distributions for the two leading in  $p_T$  jets (a, d) or the leading in  $p_T$  b-tagged jet (b, c, e, f) in  $\mu$ +jets (a-c) and  $e$ +jets (d-f) events with 0 (a, d), 1 (b, e) or at least 2 (c, f) b-tagged jets, using pp collisions at  $\sqrt{s} = 5.02$  TeV. The distributions from data (Table B.1) are compared to the sum of the expectations for the signal ( $t\bar{t}$ ) and backgrounds (Table 5.1). The QCD multijet background is estimated from data (see Section 5.4.1). The ratio between data and simulated distributions prior to any fit is shown at the bottom of each figure. The hatched areas indicate the statistical uncertainty due to the finite event count in the simulated samples [AN1].



**Figure 5.7:** Distributions for the scalar  $p_T$  sum of all jets in  $\mu$ +jets (a–c) and  $e$ +jets (d–f) events with 0 (a, d), 1 (b, e) or at least 2 (c, f) b-tagged jets, using pp collisions at  $\sqrt{s} = 5.02$  TeV. The distributions from data (Table B.1) are compared to the sum of the expectations for the signal ( $t\bar{t}$ ) and backgrounds (Table 5.1). The QCD multijet background is estimated from data (see Section 5.4.1). The ratio between data and simulated distributions prior to any fit is shown at the bottom of each figure. The hatched areas indicate the statistical uncertainty due to the finite event count in the simulated samples [AN1]

as from either punch-through hadrons or collimated jets with a high electromagnetic fraction, can yield  $\ell$ +jets-like topologies.

The estimation of the QCD multijet background is separately performed for the events with 0, 1, or  $\geq 2$  b jets using a control region (CR) where either the muon candidate fails a looser isolation requirement ( $I_{\text{rel}}^\mu < 0.2$ ) or the electron candidate fails the identification criteria. The choice of the QCD multijet control region has been made in such a way as to minimize the contamination attributable to the signal and W+jets events while retaining a large number of events in the sample for the estimation of this type of background.

The initial normalization of the QCD multijet contribution in the signal region (SR) is derived from the “reduced-signal region” (RSR) of  $p_{\text{T}}^{\text{miss}} < 20$  GeV. Events in both the reduced-signal and control regions fulfilling this requirement are counted. After subtracting the expected contributions from non-QCD processes, the ratio between the numbers of events observed in the reduced-signal region and in the control region,

$$N_{\text{SR}}(\text{QCD}) = (N_{\text{CR}}(\text{obs}) - N_{\text{CR}}(\text{non-QCD})) \frac{N_{\text{RSR}}^{p_{\text{T}}^{\text{miss}} < 20}(\text{obs}) - N_{\text{RSR}}^{p_{\text{T}}^{\text{miss}} < 20}(\text{non-QCD})}{N_{\text{CR}}^{p_{\text{T}}^{\text{miss}} < 20}(\text{obs}) - N_{\text{CR}}^{p_{\text{T}}^{\text{miss}} < 20}(\text{non-QCD})}, \quad 5.2$$

is used as a transfer factor to normalize the QCD multijet background estimate. In both the electron and muon final states, a 30% uncertainty is assigned to the estimate of the expected contribution from non-QCD processes, estimated after varying the QCD scales in the W+jets simulation. This uncertainty propagates as both normalization and shape uncertainty in the predicted distributions for the QCD multijet process. The variations are applied independently in the reduced-signal and control regions in order to determine an uncertainty envelope. A more accurate normalization for the QCD multijet contribution is obtained by the fit performed to extract the final cross section, described in Section 5.6.1.

Figures 5.8 shows the  $p_{\text{T}}^{\text{miss}}$  distributions in SR for  $mu$ +jets and  $e$ +jets final states, respectively. For validation the transverse mass ( $m_{\text{T}}$ ) of the lepton- $\bar{p}_{\text{T}}^{\text{miss}}$  system,

$$m_{\text{T}} = \sqrt{2p_{\text{T}}^{\text{miss}} p_{\text{T}}^{\text{lepton}} (1 - \cos\Delta\phi)}, \quad 5.3$$

is displayed in Fig. 5.9 for the same event categories and final states. In both cases, the distributions in these exclusive events, i.e., requiring two non-b-tagged jets plus 0, 1 or at least 2 b-tagged jets, a fair agreement is observed in both variables. For higher number of b-tagged jets, the expected contamination from non-QCD processes tends to be larger.

### 5.4.2 The dilepton final state

Final states with two genuine leptons can originate from background processes, primarily from  $Z/\gamma^* \rightarrow \tau^+ \tau^-$  (where the  $\tau$  leptonic decays can yield  $e^\pm \mu^\mp$  or  $\mu^\pm \mu^\mp$  plus  $p_{\text{T}}^{\text{miss}}$  due to the neutrinos), tW, and WV events. Other background sources, such as W+jets events or  $t\bar{t}$  production in the  $\ell$ +jets final state, can contaminate the signal sample if a jet is misidentified as a lepton, or if an event contains a lepton from the decay of b or c hadrons. These are included in the “non-W/Z” category since genuine leptons are defined as originating from decays of W or Z bosons. The yields from tW and WV events are

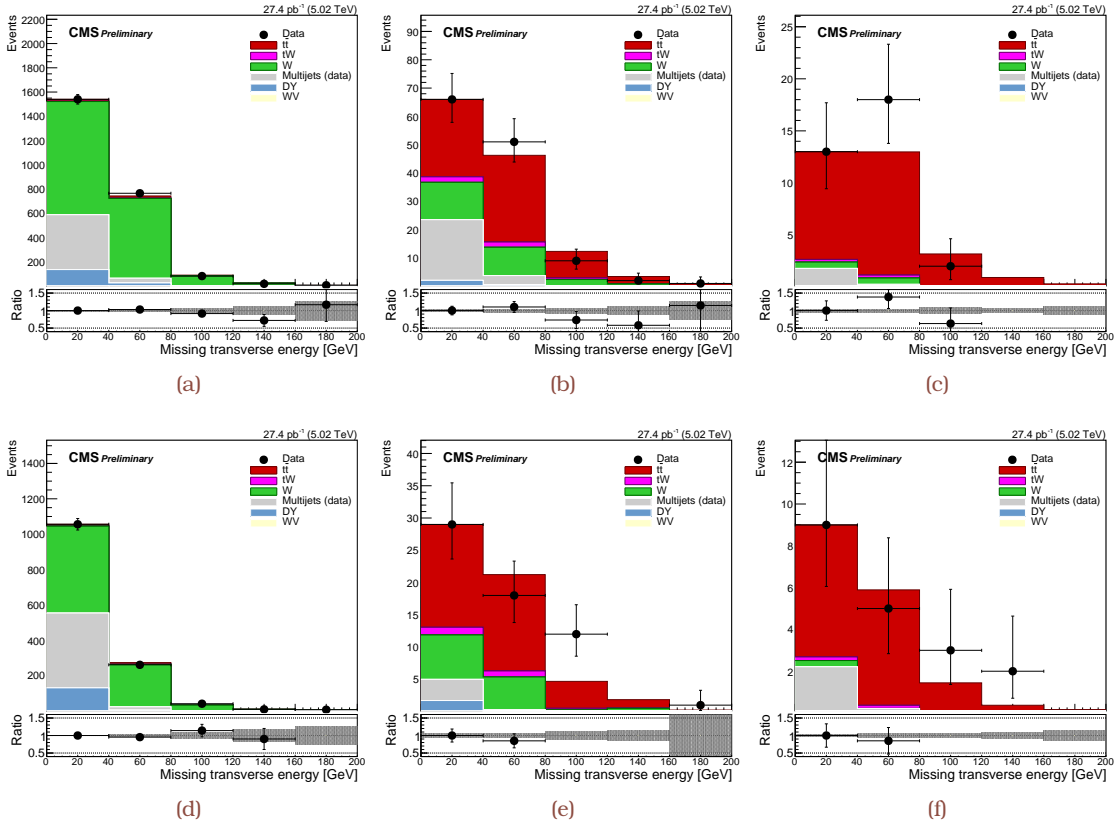


Figure 5.8: Distributions of  $p_T^{\text{miss}}$  in SR for  $\mu$ +jets (a–c) and  $e$ +jets (d–f) final states. From left to right, events with 0, 1 or at least 2 b-tagged jets are shown. The ratio between data (Table B.1) and simulated distributions (Table 5.1) prior to any fit is shown at the bottom of each figure. The hatched areas indicate the statistical uncertainty due to the finite event count in the simulated samples [AN1].

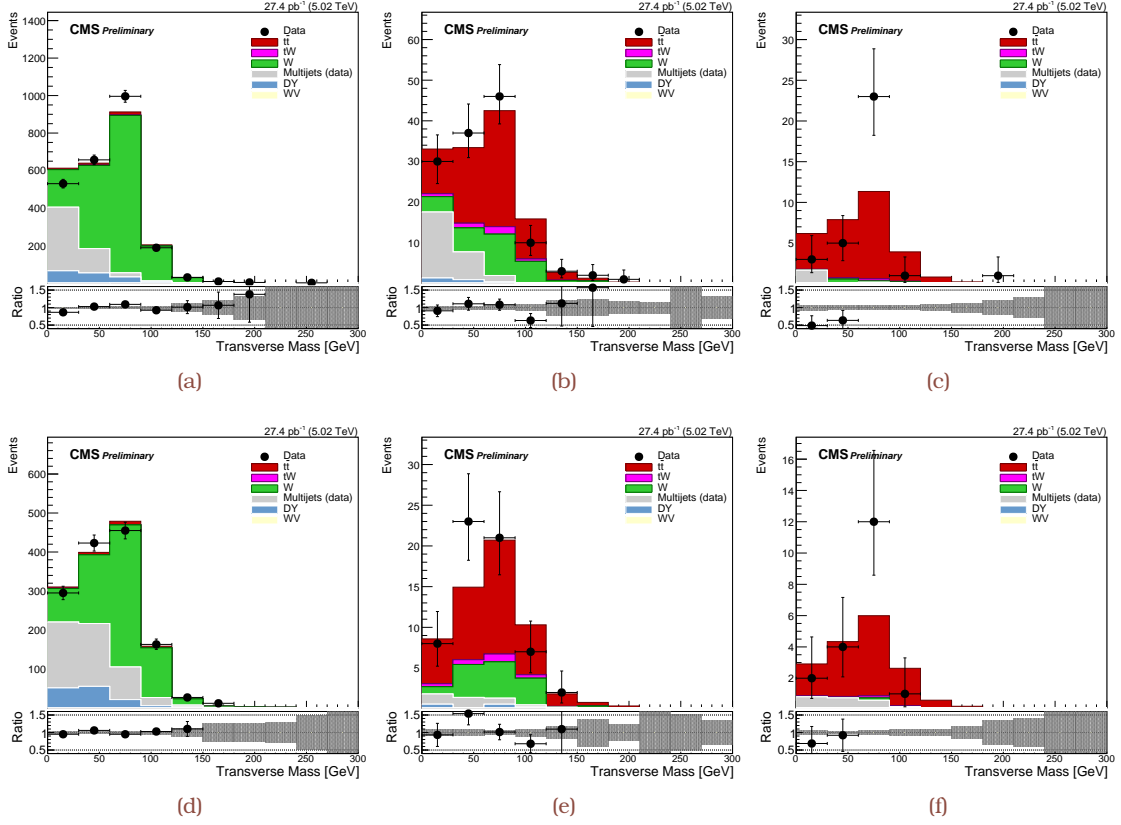
estimated from simulation, while the contribution of the  $Z/\gamma^*$  background is evaluated using control samples in data. The rate of non- $W/Z$  backgrounds is extracted from control samples in data for the  $e^\pm\mu^\mp$  final state and is estimated from the simulation in the  $\mu^\pm\mu^\mp$  final state.

A scale factor for the  $Z/\gamma^*$  background normalization is estimated from the number of events within the  $Z$  boson mass window in data, which is extrapolated to the number of events outside the window. This ratio,  $R_{\text{out/in}}$ , is estimated from MC simulation, and the contamination from processes other than the  $Z/\gamma^*$  contribution in the  $Z$  boson mass window is considered to be negligible. The number of extrapolated events can be inferred from data as

$$N_{\text{out}}^{\ell^\pm\ell^\mp, \text{obs}} = R_{\text{out/in}}^{\ell^\pm\ell^\mp} (N_{\text{in}}^{\ell^\pm\ell^\mp} - 0.5N_{\text{in}}^{e^\pm\mu^\mp} k_{\ell^\pm\ell^\mp}). \quad 5.4$$

$R_{\text{out/in}}$  is the ratio of the number of events outside (“out”) and inside (“in”) the  $Z$  mass region, estimated from the  $Z/\gamma^*$  MC simulation event samples (Table 5.1)

$$R_{\text{out/in}} = \frac{N_{Z/\gamma^*, \text{MC}}^{\text{out}}}{N_{Z/\gamma^*, \text{MC}}^{\text{in}}}, \quad 5.5$$



**Figure 5.9:** Distributions of the  $m_T$  variable in SR for  $\mu$ +jets (a-c) and  $e$ +jets (d-f) final states. From left to right, events with 0, 1 or at least 2 b-tagged jets are shown. The ratio between data (Table B.1) and simulated distributions (Table 5.1) prior to any fit is shown at the bottom of each figure. The hatched areas indicate the statistical uncertainty due to the finite event count in the simulated samples [AN1].

with  $k$  a corrector factor that must be applied to take into account the differences between electron and muon reconstruction. It is calculated using the events in the  $Z$  mass region that fulfill the standard dilepton selection, and can be written as

$$k_{\ell^\pm \ell'^\mp} = \sqrt{\frac{N_{\text{in}}^{\ell^\pm \ell'^\mp}}{N_{\text{in}}^{\ell'^\pm \ell^\mp}}}, \quad 5.6$$

with  $\ell = e$  and  $\ell' = \mu$  or  $\ell = \mu$  and  $\ell' = e$ .

Tables 5.3 and 5.4 summarize the input obtained from data and simulation at different selection levels in the  $e^\pm \mu^\mp$  and  $\mu^\pm \mu^\mp$  final states, respectively. The data-based estimation is then compared to the MC simulation estimation to derive scale factors. In the  $e^\pm \mu^\mp$  final state, the scale factor is defined as

$$\text{SF}_{e^\pm \mu^\mp} = \sqrt{\text{SF}_{e^\pm e^\mp} \text{SF}_{\mu^\pm \mu^\mp}}. \quad 5.7$$

A scale factor of  $0.96 \pm 0.78$  (stat) is obtained in the  $\mu^\pm \mu^\mp$  final state, and  $0.91 \pm 0.14$  (stat) in the  $e^\pm \mu^\mp$  final state. The estimation is performed using events with at least two jets, and the dependence on different jet multiplicities is discussed in Section 5.5.

**Table 5.3:**  $Z/\gamma^*$  background estimation in the  $e^\pm\mu^\mp$  final state for events with at least two reconstructed jets. Uncertainties are of statistical nature. The breakdown of the results obtained for the different dilepton final states is given [AN3,AN4].

	$e^\pm e^\mp$	$\mu^\pm\mu^\mp$	$e^\pm\mu^\mp$
$N_{\text{in}}$ (MC)	$226.0 \pm 5.5$	$250.3 \pm 5.6$	
$N_{\text{out}}$ (MC)	$18.2 \pm 2.4$	$32.9 \pm 3.0$	
$R_{\text{out/in}}$ (MC)	$0.081 \pm 0.013$	$0.132 \pm 0.015$	
$k_{\ell^\pm\ell^\mp}$	$0.950 \pm 0.022$	$1.052 \pm 0.025$	
$N_{\text{in}}$ (Obs)	$212 \pm 14.6$	$226 \pm 15.0$	$5.0 \pm 2.2$
$N_{\text{out}}$	$16.9 \pm 0.8$	$29.4 \pm 1.0$	
SF (Obs/MC)	$0.93 \pm 0.17$	$0.89 \pm 0.11$	$0.91 \pm 0.14$
$Z/\gamma^*$ (MC)	$18.6 \pm 2.3$	$33.6 \pm 3.1$	$1.8 \pm 0.4$
$Z/\gamma^*$ (Obs)	$17.3 \pm 2.1$	$30.0 \pm 2.8$	$1.6 \pm 0.4$

**Table 5.4:**  $Z/\gamma^*$  background estimation in the  $\mu^\pm\mu^\mp$  final state for different levels in the event selection. Uncertainties are of statistical nature [AN3,AN4].

	$\mu^\pm\mu^\mp$	$p_{\text{T}}^{\text{miss}} > 35 \text{ GeV}$	$\geq 2 \text{ jets}$
$N_{\text{in}}$ ( $\mu^\pm\mu^\mp$ , MC)	$6955.2 \pm 21.4$	$7.3 \pm 1.0$	$4.2 \pm 0.8$
$N_{\text{out}}$ ( $\mu^\pm\mu^\mp$ , MC)	$799.7 \pm 12.0$	$3.0 \pm 0.7$	$1.2 \pm 0.5$
$R_{\text{out/in}}$ (MC)	$0.115 \pm 0.002$	$0.407 \pm 0.151$	$0.300 \pm 0.173$
$N_{\text{in}}$ ( $e^\pm e^\mp$ , MC)	$6408.3 \pm 21.2$	$7.5 \pm 1.0$	$4.1 \pm 0.7$
$k_{\ell^\pm\ell^\mp}$	$1.042 \pm 0.003$	$0.985 \pm 0.129$	$1.010 \pm 0.185$
$N_{\text{in}}$ ( $\mu^\pm\mu^\mp$ , Obs)	$6609 \pm 81.3$	$10 \pm 3.2$	$6.0 \pm 2.4$
$N_{\text{in}}$ ( $e^\pm\mu^\mp$ , Obs)	$12.0 \pm 3.5$	$5.0 \pm 2.2$	$4.0 \pm 2.0$
$N_{\text{out}}$	$763.9 \pm 2.2$	$3.1 \pm 0.7$	$1.2 \pm 0.5$
SF (Obs/MC)	$0.949 \pm 0.017$	$1.036 \pm 0.481$	$0.957 \pm 0.788$

The non-W/Z background in the  $e^\pm\mu^\mp$  final state is estimated using an extrapolation from a control region of same-sign (SS) dilepton events to the signal region of opposite-sign (OS) lepton pairs. The SS control region is defined using the same criteria as for the nominal signal region, except requiring dilepton pairs of the same charge. The muon isolation requirement is relaxed in order to enhance the number of events. The SS dilepton events predominantly contain at least one misidentified lepton. Other SM processes produce genuine SS or charge-misidentified dilepton events with significantly smaller rates; these are estimated using simulation and subtracted from the observed number of events in data.

The scaling from the SS control to the signal region in data is performed using an extrapolation factor extracted from MC simulation, given by the ratio of the number of OS events with misidentified leptons to the number of SS events with misidentified (“misID”) leptons. Taking these into account, the fake lepton contribution  $N_{\text{data}}^{\text{OS fakes}}$  to the measurement is estimated as

$$N_{\text{data}}^{\text{OS fakes}} = \text{SF}_{\text{fakes}} N_{\text{MC}}^{\text{OS fakes}},$$

5.8



where  $SF_{\text{fakes}}$  is defined to be

$$SF_{\text{fakes}} = \frac{N_{\text{data}}^{\text{SS}} - N_{\text{misID}}^{\text{SS}} - N_{\text{rare}}^{\text{SS}}}{N_{\text{MC}}^{\text{SS fakes}}} = \frac{N'_{\text{data}}{}^{\text{SS}}}{N_{\text{MC}}^{\text{SS fakes}}} \quad 5.9$$

The contribution from  $t\bar{t}$ ,  $Z/\gamma^*$  and  $tW$  processes are subtracted from  $N_{\text{data}}^{\text{SS}}$ , giving rise to the numerator ( $N'_{\text{data}}{}^{\text{SS}}$ ) on the right-hand side of Eq. (5.8). The resulting estimate for the non-W/Z background is thus  $1.0 \pm 0.9$  (stat) events in the  $e^\pm\mu^\mp$  final state, where the central value comes from the estimation using events with at least two reconstructed jets. No particular dependence of this scale factor is observed for different jet multiplicities within the sizeable statistical uncertainty.

**Table 5.5:** The non-W/Z background estimation in the  $e^\pm\mu^\mp$  final state using events with at least two reconstructed jets. The uncertainties are of statistical nature [AN3, AN4].

Source	$N_{\text{jets} \geq 2}$
Prompt SS (MC)	$1.4 \pm 0.2$
Fake SS (MC)	$5.2 \pm 0.7$
Fake OS (MC)	$0.5 \pm 0.3$
R = fake OS/fake SS	$0.10 \pm 0.08$
$N_{\text{data}}^{\text{SS}}$	11
$N_{\text{data}}^{\text{OS}}$	$1.0 \pm 0.9$

## 5.5 Systematic uncertainty

### 5.5.1 Experimental uncertainty

The following sources of experimental uncertainty are considered in the analysis:

**Integrated luminosity** - The integrated luminosity has been estimated offline using a pixel-cluster counting method (see Section 4.5). The estimation takes into account the normalization uncertainty and the uncertainty related to the different conditions during typical physics periods relative to the specially tailored beam-separation scans, adding up to a total uncertainty of  $\pm 2.3\%$  [TH3].

**Trigger and selection efficiency** - The uncertainty in the electron trigger efficiency (1.5%) and the identification and isolation efficiency (2.5%) are estimated by changing the values of the data-to-simulation scale factors within the uncertainty, as obtained from a tag-and-probe method (see Appendix B.0.2). The uncertainty in the muon identification and isolation efficiency, including the trigger efficiency (Fig. 3.6), is 3% and covers one standard deviation of the scale factor from unity (see Appendix B.0.3).

**Jet energy scale and resolution** - The impact of the uncertainty in the jet energy scale is estimated by changing the  $p_T$ - and  $\eta$ -dependent JES corrections by a constant 2.8, and the uncertainty in jet energy resolution is estimated through  $\eta$ -dependent changes in the JER corrections to the simulation (Fig. 3.22). The uncertainty arising

from the use of  $p_T^{\text{miss}}$  in the  $\mu^\pm\mu^\mp$  final state is dominated by the unclustered energy contribution to  $p_T^{\text{miss}}$  (Table 3.6).

**Jet misidentification efficiencies** - An uncertainty of 30% is assigned to the b jet misidentification probability in the  $\ell$ +jets analysis, independently of the jet  $p_T$  and  $\eta$ . Owing to the fact that no dedicated measurement of this quantity has been performed for the considered data set, the misidentification rate is conservatively taken as twice the value expected at  $\sqrt{s} = 13$  TeV conditions (Table 3.5).

**QCD multijet estimation** - The uncertainty assigned to the QCD multijet background includes the statistical uncertainty in the data, and the uncertainty from the non-QCD multijet contributions subtracted from the control region, as described in Section 5.4.1, and an additional 30%-100% normalization uncertainty. The latter depends on the event category and stems from the measured difference with respect to the alternative estimate of the QCD multijet normalization based on the transverse mass of the lepton and  $p_T^{\text{miss}}$ .

**Limited event count in the W+jets MC simulation** - Owing to the finite event count in the W+jets simulated sample, an additional bin-by-bin uncertainty is assigned by generating an alternative shape to fit, where the bin prediction is varied by  $\pm 1$  standard deviation while keeping all the other bins at their nominal expectation.

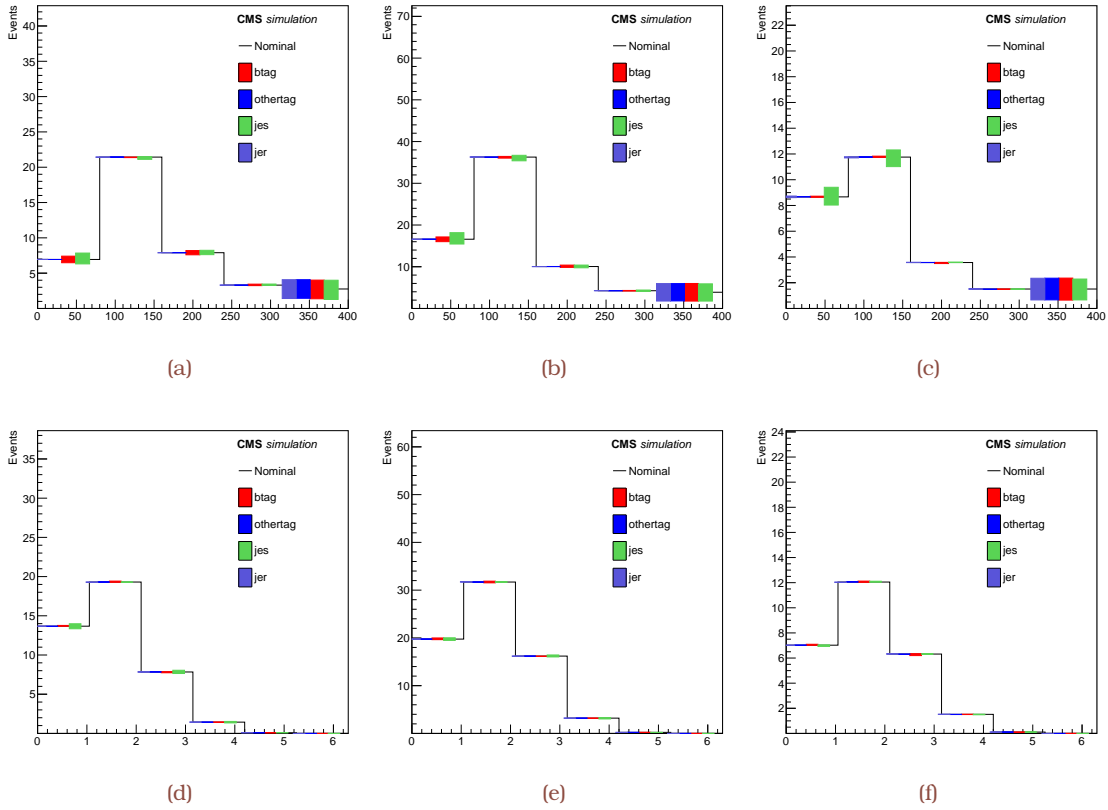
Figure 5.10 shows the signal predictions and the effect of varying the main sources of experimental uncertainty, according to the prescriptions described above.

Two variables, which are sensitive to the resonant behavior of the light jets produced from the W boson hadronic decay in a  $t\bar{t}$  event, were independently considered for the fit. Given that these light jets, here denoted by  $j$  and  $j'$ , are correlated during production, they are also expected to be closer in phase space when compared to pairs of other jets in the event. The angular distance  $\Delta R$  can thus be used as a metric to rank all pairs of non-b-tagged jets in the event, maximizing the probability of selecting those from the W boson hadronic decay in cases where more than two non-b-tagged jets are found. The invariant mass  $M(j, j')$  of jets  $j$  and  $j'$  also has a distinctive peaking feature for the signal in contrast with a smooth background continuum. From simulation, we expect that the minimum angular distance  $\Delta R$  between all pairs of jets  $j$  and  $j'$ ,  $\Delta R_{\min}(j, j')$ , is robust against signal modeling uncertainty such as the choice of the  $\mu_R, \mu_F$  scales and jet energy scale and resolution, while the  $M(j, j')$  variable tends to be more affected by this kind of uncertainty. Owing to being less affected by the systematic uncertainty while maintaining good signal-to-background discrimination power, the  $\Delta R_{\min}(j, j')$  variable is used to extract the  $t\bar{t}$  cross section.

### 5.5.2 Theory uncertainty

The following sources of theory uncertainty are considered in the analysis:

**QCD scale choices** - The uncertainty in the simulation of  $t\bar{t}$  production cause a systematic bias related to the missing higher-order diagrams in POWHEG, which is estimated through studies of the signal modeling by modifying the  $\mu_R, \mu_F$  scales within a factor of two with respect to their nominal value. In the  $\ell$ +jets analysis, the impact of the



**Figure 5.10:** Distributions for signal  $M(j,j')$  (a–c) and  $\Delta R_{\min}(j,j')$  (d–f) shapes with the main sources of experimental uncertainty superimposed for  $t\bar{t}$  events with 0 (a, d), 1 (b, e) or at least 2 (c, f) b-tagged jets [AN1]. The b tagging efficiency is treated as a parameter of interest in the fit.

$\mu_R, \mu_F$  variations are examined independently, while in the dilepton analysis they are varied simultaneously. In both analyses, these variations are applied independently at the matrix element (ME) and parton shower (PS) levels.

**Hadronizer choice** - The uncertainty arising from the hadronization model mainly affects the JES and the fragmentation of jets. The hadronization uncertainty is determined by comparing samples of events generated with POWHEG, where the hadronization is either modeled with PYTHIA or HERWIG++ (v2.7.1) [52]. This also accounts for differences in the PS model and the underlying event.

**PDFs** - The uncertainty from the choice of PDF is determined by reweighting the sample of simulated  $t\bar{t}$  events according to the RMS variation of the 100 NNPDF3.0 replica sets. Two extra variations of  $a_s$  are added in quadrature to determine the total PDF uncertainty.

**Non-QCD background estimate** - In the  $\ell$ +jets analysis, the uncertainty in the choice of the  $\mu_R, \mu_F$  scales in the W+jets simulation is taken into account by considering alternative shapes and yields after varying independently the  $\mu_R, \mu_F$  scales, following a similar procedure to that described above for the signal. A 30% normalization uncertainty in the theoretical  $tW, Z/\gamma^*,$  and  $WV$  background cross sections is assigned [246], given the previously unexplored  $\sqrt{s}$  value and that the final states con-

tain several jets. In the dilepton final state, an uncertainty of 30% is assumed [246] for the cross sections of the  $tW$  and  $WV$  backgrounds to cover the theoretical uncertainties and the effect of finite simulated samples. The uncertainty in the  $Z/\gamma^*$  estimation is calculated by combining in quadrature the statistical uncertainty and an additional 30% from the variation of the scale factor in the different levels of selection, resulting in uncertainties of about 30 and 80% in the  $e^\pm\mu^\mp$  and  $\mu^\pm\mu^\mp$  final states, respectively. The systematic uncertainty in the non-W/Z background is estimated to be 90% in the  $e^\pm\mu^\mp$  final state and is dominated by the statistical uncertainty in the method. Owing to the limited sample size in the data, the method cannot be applied in the  $\mu^\pm\mu^\mp$  final state. The estimation is therefore based on MC simulation, and an uncertainty of 100% is conservatively assigned.

Figure 5.11 summarizes the expected effect on the signal shapes from the choice of the QCD/PS-scales and hadronizer. For the QCD scale choices at matrix element level, the effect on the normalization is externalized externalized and assigned separately as an extrapolation uncertainty.

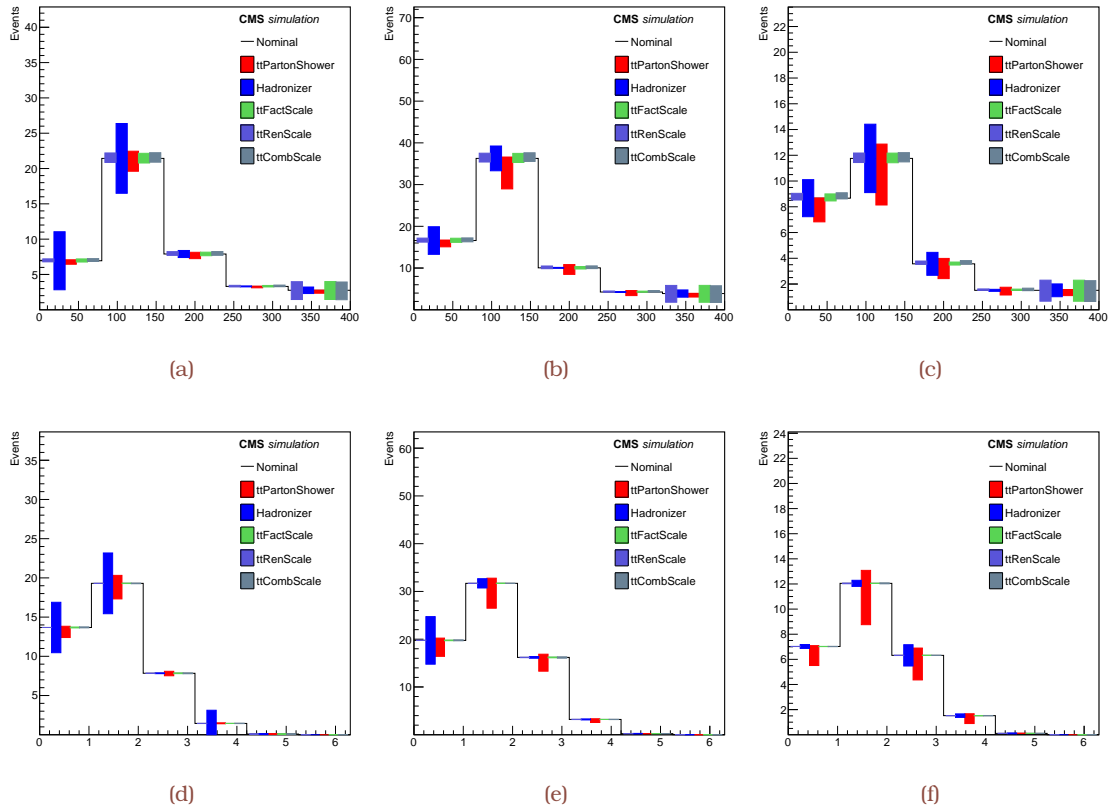


Figure 5.11: Distributions for signal  $M(j,j')$  (a–c) and  $\Delta R_{\min}(j,j')$  (d–f) shapes with the sources of theoretical uncertainty related to  $\mu_R$ ,  $\mu_F$  scales at the matrix element and parton shower levels and hadronizer choice superimposed for  $t\bar{t}$  events with 0 (a, d), 1 (b, e) or at least 2 (c, f) b-tagged jets [AN1].

**Table 5.6:** The number of expected background and signal events and the observed event yields in the different b tag categories for the e+jets and  $\mu$ +jets analyses, prior to the fit. With the exception of the QCD multijet estimate, for which the total uncertainty is reported, the uncertainties reflect the statistical uncertainty in the simulated samples [TH4].

Source	b tag category					
	0b		1b		$\geq 2b$	
	e+jets	$\mu$ +jets	e+jets	$\mu$ +jets	e+jets	$\mu$ +jets
tW	3.03±0.02	5.6±0.03	2.49±0.02	4.5±0.03	0.39±0.01	0.67±0.01
W+jets	776±17	1704±26	13±2	26±3	0.2±0.3	0.8±0.6
Z/ $\gamma^*$	136±4	162±5	1.7±0.5	2.8±0.6	0.1±0.1	0.1±0.1
WV	0.52±0.01	1.01±0.02	<0.01	< 0.02	< 0.01	< 0.01
QCD multijet	440±130	490±150	3.6±1.1	28±8	2.5±0.8	2.0±0.8
$t\bar{t}$ signal	22.8±0.3	42.3±0.4	36.9±0.4	71.1±0.5	13.8±0.2	27.0±0.3
Total	1380±130	2410±150	57.7±2.4	131±9	16.8±0.9	31±1
Observed data	1375	2406	61	129	19	33

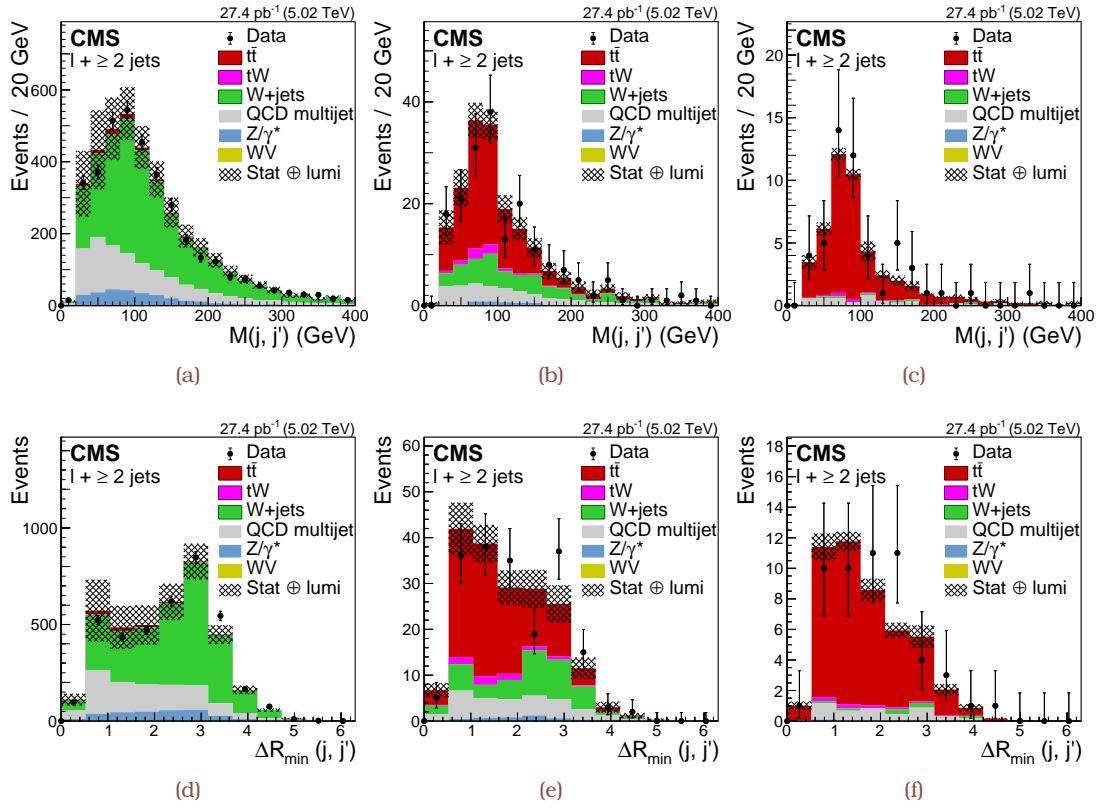
## 5.6 Measurement of the $t\bar{t}$ cross section

### 5.6.1 The $\ell$ +jets final state

In the  $\ell$ +jets analysis, the  $t\bar{t}$  cross section is measured in a fiducial phase space through a fit. In order to maximize the sensitivity of the analysis, the  $\Delta R_{\min}(j, j')$  distributions are categorized according to the number of jets—in addition to the ones assigned to the W boson hadronic decay—passing the b quark identification criteria. In total, 6 categories are used, corresponding to electron or muon events with 0, 1, or  $\geq 2$  b jets. The expected number of signal and background events in each category prior to the fit and the observed yields are given in Table 5.6. Good agreement is observed between data and expectations.

The  $M(j, j')$  and  $\Delta R_{\min}(j, j')$  distributions are shown in Fig. 5.12. The distributions have been combined for the e+jets and  $\mu$ +jets final states to maximize the statistical precision and are shown for events with different b-tagged jet multiplicities. From simulation, we expect that the signal peaks at low  $\Delta R$ , while the background is uniformly distributed up to  $\Delta R \approx 3$ . Above that value, fewer events are expected, and background processes are predicted to dominate. A fair agreement is observed between data and pre-fit expectations.

A profile likelihood ratio (PLR) method, similar to the one employed in Ref. [75], is used to perform the fit the number of events counted in the different categories. The likelihood function takes into account the expectations from background processes as well as signal. These expectations depend on: (i) the simulation- or data-based expectations ( $\hat{S}$  or  $\hat{B}$  for signal and background, respectively), and (ii) nuisance parameters ( $\partial_i$ ) that parameterize the uninteresting variables used to control the effect of the systematic variations, as described in Section 5.5. The effect of each source of uncertainty is separated in a rate- and shape-changing nuisance parameter. In the fit, the nuisance parameters are assumed to be distributed according to log-normal probability distribution functions, if affecting the rate, or Gaussian PDFs if affecting the shapes. The signal expectation is also



**Figure 5.12:** The predicted and observed distributions of the (a–c)  $M(j, j')$  and (d–f)  $\Delta R_{\min}(j, j')$  variables for  $\ell$ +jets events in the 0b (a,d), 1b (b,e), and  $\geq 2b$  (c,f) tagged jet categories. The distributions from data are compared to the sum of the expectations for the signal and backgrounds prior to any fit. The QCD multijet background is estimated from data (see Section 5.4.1). The cross-hatched band represents the statistical and the integrated luminosity uncertainties in the expected signal and background yields added in quadrature. The vertical bars on the data points represent the statistical uncertainties [TH4].

modulated by a multiplicative factor, i.e., the ratio of the observed  $t\bar{t}$  cross section to the expectation from theory, the so-called signal strength  $\mu = \sigma/\sigma_{\text{th}}$  for  $m_{\text{top}} = 172.5 \text{ GeV}$ . For each category (k), the total number of expected events can be written as

$$\hat{N}_k(\mu, \bar{\Theta}) = \mu \hat{S}_k \prod_i (1 + \delta_i^S \partial_i) + \hat{B}_k \prod_i (1 + \delta_i^B \partial_i), \quad 5.10$$

where  $\bar{\Theta}$  is the set of all nuisance parameters, the index k runs over the bins of the distributions (or the yields for the cross-check event counting analysis), and  $\delta_i^S$  and  $\delta_i^B$  are changes in yields induced through one-standard-deviation changes in the  $i$ th source of uncertainty in the signal and background, respectively. The likelihood function is then defined as

$$\mathcal{L}(\mu, \bar{\Theta}) = \prod_k \mathcal{P}[N_k | \hat{N}_k(\mu, \partial_i)] \prod_i \rho(\partial_i), \quad 5.11$$

where  $\mathcal{P}$  is a Poisson distribution,  $N_k$  is the number of events observed in the  $k$ th category, and  $\rho(\partial_i)$  corresponds to the PDF associated with a nuisance parameter. The  $t\bar{t}$  cross

section is measured maximizing the PLR test statistic, which is asymptotically distributed as a  $\chi^2$  distribution,

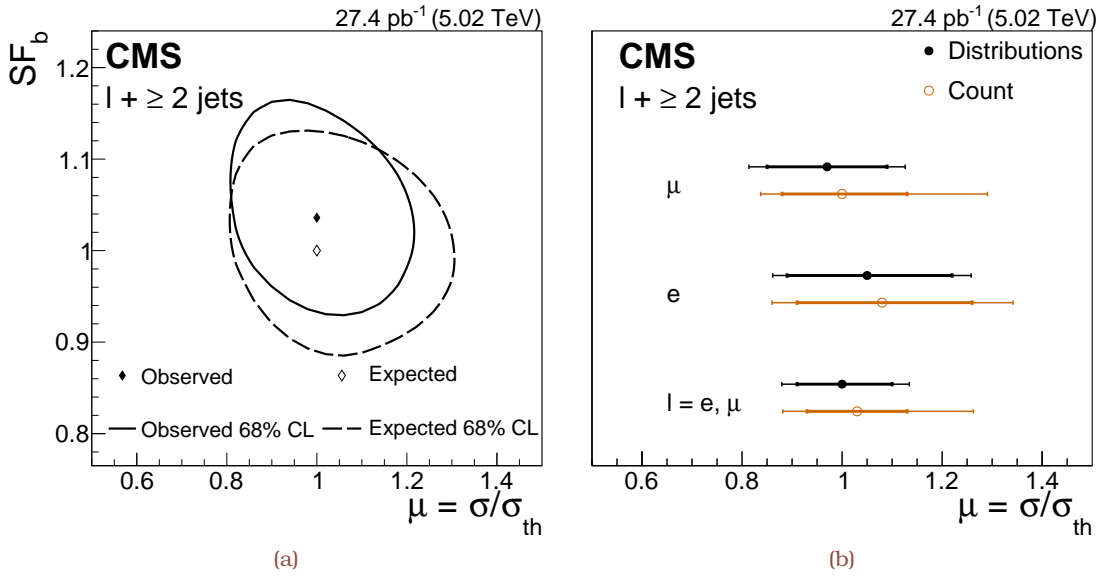
$$\hat{\rho}(\mu, \text{SF}_b) = \frac{\mathcal{L}(\mu, \text{SF}_b, \hat{\Theta})}{\mathcal{L}(\hat{\mu}, \hat{\text{SF}}_b, \hat{\Theta})}, \quad 5.12$$

where  $\mu$  and  $\Theta$  are the signal strength and the set of nuisance parameters, respectively, and  $\text{SF}_b$  is an additional parameter of interest meant for the  $b$  tagging efficiency. The quantities  $\hat{\Theta}$  correspond to the values of the nuisance parameters that maximize the likelihood for the specified signal strength and  $b$  tagging efficiency (conditional likelihood), and  $\hat{\mu}$ ,  $\hat{\text{SF}}_b$ ,  $\hat{\Theta}$  are, respectively, the values of the signal strength,  $b$  tagging efficiency, and nuisance parameters that maximize the likelihood. In the presence of nuisance parameters, the resulting PLR as a function of  $\mu$  and  $\text{SF}_b$  tends to be broader relative to the one obtained when the values are well known and fixed. This reflects the loss of information because of the systematic uncertainty [260]. The uncertainty interval ( $\pm 1$  standard deviation) corresponds to an increase in the parabolic shape of the PLR from the minimum obtained with the best-fit signal strength  $\hat{\mu}$  by a factor of 1. In the case of a two-dimensional likelihood contour, the factor to derive the uncertainty is 2.3 instead.

Figure 5.13 (left) shows the two-dimensional contours at the 68% confidence level (CL) obtained from the scan of  $-2 \ln(\hat{\rho})$ , as functions of  $\mu$  and  $\text{SF}_b$ . The expected results, obtained using the Asimov data set [260], are compared to the observed results and found to agree well within one standard deviation. The signal strength is obtained after profiling  $\text{SF}_b$ , and the result is  $\mu = 1.00^{+0.10}_{-0.09}$  (stat)  $^{+0.09}_{-0.08}$  (syst).

As a cross-check, the signal strength is also extracted by fitting only the total number of events observed in each of the six categories. The observed value  $\mu = 1.03^{+0.10}_{-0.10}$  (stat)  $^{+0.21}_{-0.11}$  (syst) is in agreement with the analysis using the  $\Delta R_{\min}(j, j')$  distributions. Figure 5.13 (right) summarizes the results obtained for the signal strength fit in each final state separately from the analysis of the distributions and event counting. In both cases, a substantial contribution to the uncertainty is systematic in nature, although the statistical component is still significant. In the  $l+jets$  combination, the  $\mu+jets$  final state is expected and observed to carry the largest weight.

To estimate the impact of the experimental systematic uncertainty in the measured signal strength, the fit is repeated after fixing one nuisance parameter at a time at its post-fit uncertainty ( $\pm 1$  standard deviation) values. The impact on the signal strength fit is then evaluated from the difference induced in the final result from this procedure. By repeating the fits, the effect of some nuisance parameters being fixed may be reabsorbed by a variation of the ones being profiled, owing to correlations. As such, the individual sources of experimental uncertainty obtained and summarized in Table 5.7 and Appendix A, respectively, can only be interpreted as the observed post-fit values, and not as an absolute, orthogonalized breakdown of the uncertainty. Compared to the event counting (Fig. A.1), the analysis of the distributions is less prone to the uncertainty in the QCD multijet background and jet energy resolution. In both cases, the signal modeling uncertainty and the  $b$  tagging efficiency are among the most significant sources of uncertainty. The post-fit constraints are mostly observed in the shape analysis (Fig. A.2),



**Figure 5.13:** a: The 68% CL contour obtained from the scan of the likelihood in  $l$ +jets analysis, as a function of  $\mu$  and  $SF_b$  in the  $l$ +jets analysis. The solid (dashed) contour refers to the result from data (expectation from simulation). The solid (hollow) diamond represents the observed fit result (SM expectation) [TH4]. b: Summary of the signal strengths separately obtained in the  $e$ +jets and  $\mu$ +jets final states, and after their combination in the  $l$ +jets final state. The results of the analysis from the distributions are compared to those from the cross-check analysis with event counting (Count). The inner (outer) bars correspond to the statistical (total) uncertainty in the signal strengths [TH4].

and mainly in nuisance parameters that could lead to significantly different expectations (Figs. 5.10 and 5.11) for the  $\Delta R_{\min}(j, j')$  variable.

The fiducial cross section is measured in events with one electron (muon) in the range  $p_T > 35$  (25) GeV and  $|\eta| < 2.1$  (including the transition region for electrons), and at least two jets with  $p_T > 25$  GeV and  $|\eta| < 2.4$ . After multiplying the signal strength by the theoretical expectations (Eq. (5.1)), we find

$$\sigma_{\text{fid}} = 20.8 \pm 2.0 (\text{stat}) \pm 1.8 (\text{syst}) \pm 0.5 (\text{lum}) \text{ pb}.$$

The combined acceptance in the  $e$ +jets and  $\mu$ +jets final states is estimated using the NLO POWHEG simulation to be  $\mathcal{A} = 0.301 \pm 0.007$ , with the uncertainty being dominated by the variation of the  $\mu_R, \mu_F$  scales at ME and PS levels and the hadronization model used for the  $t\bar{t}$  signal. The uncertainty due to the PDFs is included but verified to be less critical. Taking into account the acceptance of the analysis and its uncertainty, the inclusive  $t\bar{t}$  cross section is determined to be

$$\sigma_{t\bar{t}} = 68.9 \pm 6.5 (\text{stat}) \pm 6.1 (\text{syst}) \pm 1.6 (\text{lum}) \text{ pb},$$

in agreement with the SM prediction and attaining a 13% total relative uncertainty.



**Table 5.7:** The estimated impact of each source of uncertainty in the value of  $\mu$  extracted from the analysis of distributions, and in the cross-check from event counting. The “Other background” component includes the contributions from  $Z/\gamma^*$ ,  $tW$ , and  $WV$  events. The total uncertainty is obtained by adding in quadrature the statistical, experimental systematic, and theoretical uncertainties. The individual experimental uncertainties are obtained by repeating the fit after fixing one nuisance parameter at a time at its post-fit uncertainty ( $\pm 1$  standard deviation) value. The values quoted have been symmetrized [TH4].

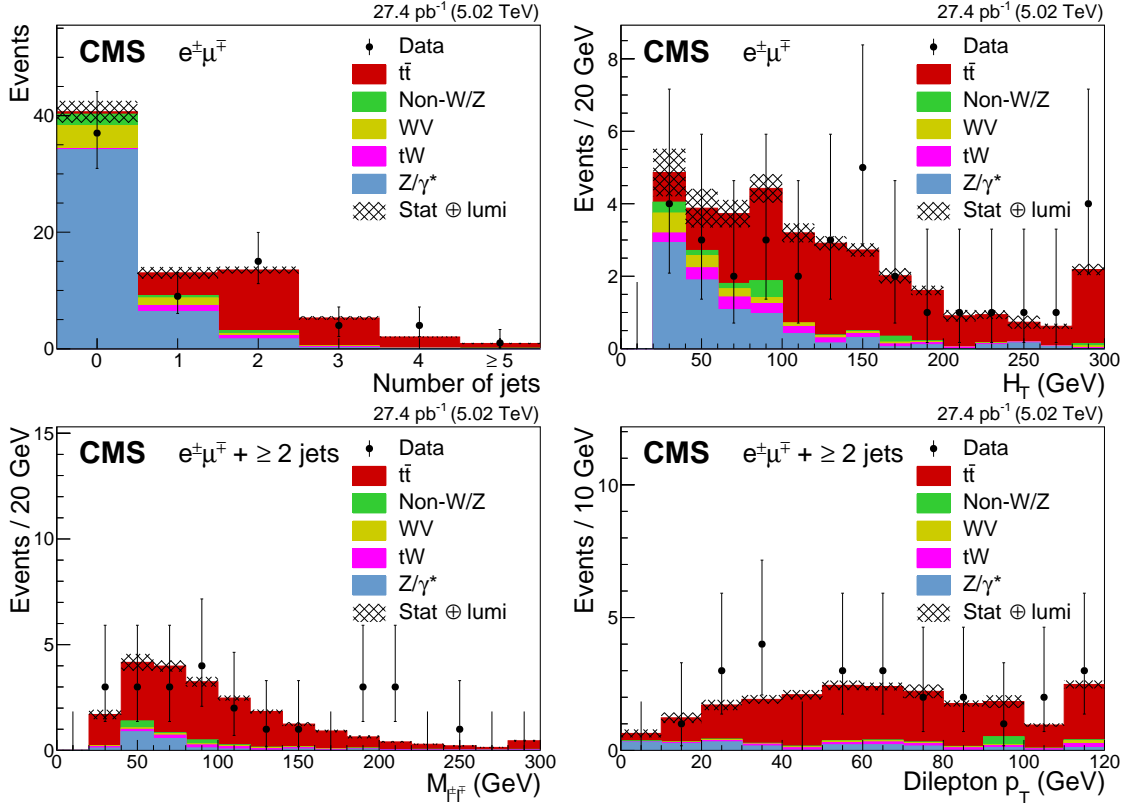
Source	$\Delta\mu/\mu$	
	Distr.	Count
Statistical uncertainty	0.095	0.100
Experimental systematic uncertainty	0.085	0.160
<i>Sources of experimental uncertainty</i>		
W+jets background	0.035	0.025
QCD multijet background	0.024	0.044
Other background	0.013	0.013
Jet energy scale	0.030	0.031
Jet energy resolution	0.006	0.023
b tagging	0.034	0.045
Electron efficiency	0.011	0.028
Muon efficiency	0.017	0.022
<i>Sources of theoretical uncertainty</i>		
Hadronization model of $t\bar{t}$ signal	0.028	0.069
$\mu_R, \mu_F$ scales of $t\bar{t}$ signal (PS)	0.044	0.115
$\mu_R, \mu_F$ scales of $t\bar{t}$ signal (ME)	<0.010	<0.010
Total uncertainty	0.127	0.189

### 5.6.2 The dilepton final state

In the dilepton analysis, the  $t\bar{t}$  cross section is extracted from an event counting measurement. Figure 5.14 shows the distributions of the jet multiplicity and the scalar  $p_T$  sum of all jets ( $H_T$ ), for events passing the dilepton criteria in the  $e^\pm\mu^\mp$  final state. In addition, it displays the lepton-pair invariant mass and  $p_T$  distributions, after requiring at least two jets in the event in the  $e^\pm\mu^\mp$  final state. Figure 5.15 shows the  $p_T^{\text{miss}}$  and the lepton-pair invariant mass distributions in the  $\mu^\pm\mu^\mp$  final state for events passing the dilepton criteria, and the  $Z$  boson veto with the  $p_T^{\text{miss}} > 35$  GeV requirement, in the second case. The predicted distributions take into account the efficiency corrections described in Section 5.3 and the background estimations discussed in Section 5.4.2. Good agreement is observed between the data and predictions for both signal and background.

The fiducial  $t\bar{t}$  production cross section is measured by counting events in the visible phase space (defined by the same  $p_T$ ,  $|\eta|$ , and multiplicity requirements for leptons and jets as in Section 5.3, but including the transition region for electrons) and is denoted by  $\sigma_{\text{fid}}$ . It is extrapolated to the full phase space in order to determine the inclusive  $t\bar{t}$  cross section using the expression

$$\sigma_{t\bar{t}} = \frac{N - N_B}{\varepsilon \times \mathcal{A} \times \mathcal{L}} = \frac{\sigma_{\text{fid}}}{\mathcal{A}}, \quad 5.13$$



**Figure 5.14:** Predicted and observed distributions of the (upper row) jet multiplicity and scalar  $p_T$  sum of all jets ( $H_T$ ) for events passing the dilepton criteria, and of the (lower row) invariant mass and  $p_T$  of the lepton pair after requiring at least two jets, in the  $e^\pm\mu^\mp$  final state. The  $Z/\gamma^*$  and non-W/Z backgrounds are determined from data (see Section 5.4.2). The cross-hatched band represents the statistical and integrated luminosity uncertainties in the expected signal and background yields added in quadrature. The vertical bars on the data points represent the statistical uncertainty. The last bin of the distributions contains the overflow events [TH4].

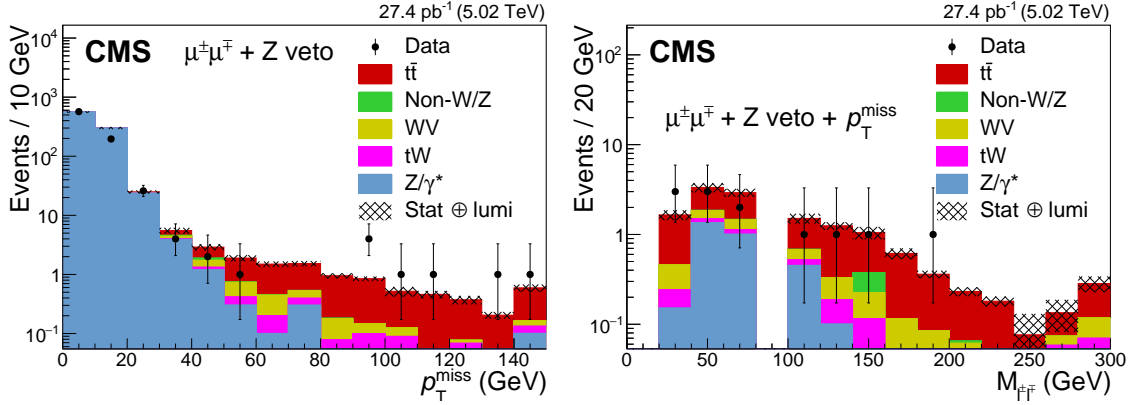
where  $N$  is the total number of dilepton events observed in data,  $N_B$  the number of estimated background events,  $\varepsilon$  the selection efficiency,  $\mathcal{A}$  the acceptance, and  $\mathcal{L}$  the integrated luminosity. Table 5.8 gives the total number of events observed in data, together with the total number of signal and background events expected from simulation or estimated from data, after the full set of selection criteria. The total detector, trigger, and reconstruction efficiency is estimated from data to be  $\varepsilon = 0.55 \pm 0.02$  ( $0.57 \pm 0.04$ ) in the  $e^\pm\mu^\mp$  ( $\mu^\pm\mu^\mp$ ) final state. Using the definitions above, the yields from Table 5.8, and the systematic uncertainty from Table 5.9, the measured fiducial cross section for  $t\bar{t}$  production is

$$\sigma_{\text{fid}} = 41 \pm 10 (\text{stat}) \pm 2 (\text{syst}) \pm 1 (\text{lum}) \text{ pb},$$

in the  $e^\pm\mu^\mp$  final state, and

$$\sigma_{\text{fid}} = 22 \pm 11 (\text{stat}) \pm 4 (\text{syst}) \pm 1 (\text{lum}) \text{ pb},$$

in the  $\mu^\pm\mu^\mp$  final state.



**Figure 5.15:** Predicted and observed distributions of the (left)  $p_T^{\text{miss}}$  in events passing the dilepton criteria and Z boson veto, and of the (right) invariant mass of the lepton pair after the  $p_T^{\text{miss}} > 35 \text{ GeV}$  requirement in the  $\mu^\pm\mu^\mp$  final state. The cross-hatched band represents the statistical and integrated luminosity uncertainties in the expected signal and background yields added in quadrature. The vertical bars on the data points represent the statistical uncertainty. The last bin of the distributions contains the overflow events [TH4].

**Table 5.8:** The predicted and observed numbers of dilepton events obtained after applying the full selection. The values are given for the individual sources of background,  $t\bar{t}$  signal, and data. The uncertainties are of statistical nature [TH4].

Source	$e^\pm\mu^\mp$	$\mu^\pm\mu^\mp$
tW	$0.92 \pm 0.02$	$0.29 \pm 0.01$
Non-W/Z leptons	$1.0 \pm 0.9$	$0.04 \pm 0.01$
Z/ $\gamma^*$	$1.6 \pm 0.2$	$1.1 \pm 0.8$
WW	$0.44 \pm 0.02$	$0.15 \pm 0.01$
$t\bar{t}$ signal	$18.0 \pm 0.3$	$6.4 \pm 0.2$
Total	$22.0 \pm 0.9$	$7.9 \pm 0.8$
Observed data	24	7

The acceptance, as estimated from MC simulation, is found to be  $\mathcal{A} = 0.53 \pm 0.01$  ( $0.37 \pm 0.01$ ) in the  $e^\pm\mu^\mp$  ( $\mu^\pm\mu^\mp$ ) final state. The statistical uncertainty (from MC simulation) is included in the uncertainty in  $\mathcal{A}$ . By extrapolating to the full phase space, the inclusive  $t\bar{t}$  cross section is measured to be

$$\sigma_{t\bar{t}} = 77 \pm 19 \text{ (stat)} \pm 4 \text{ (syst)} \pm 2 \text{ (lum)} \text{ pb},$$

in the  $e^\pm\mu^\mp$  final state, and

$$\sigma_{t\bar{t}} = 59 \pm 29 \text{ (stat)} \pm 11 \text{ (syst)} \pm 1 \text{ (lum)} \text{ pb},$$

in the  $\mu^\pm\mu^\mp$  final state. Table 5.9 summarizes the relative and absolute statistical and systematic uncertainties from different sources contributing to  $\sigma_{t\bar{t}}$ . The separate total systematic uncertainty without the uncertainty in the integrated luminosity, the part

**Table 5.9:** Summary of the individual contributions to the systematic uncertainty in the  $\sigma_{t\bar{t}}$  measurements for the dilepton final states. The relative uncertainty  $\Delta\sigma_{t\bar{t}}/\sigma_{t\bar{t}}$  (in %), as well as the absolute uncertainty in  $\sigma_{t\bar{t}}$ ,  $\Delta\sigma_{t\bar{t}}$  (in pb), are presented. The statistical and total uncertainties are also given, where the latter are the quadrature sum of the statistical and systematic uncertainties [TH4].

Source	$e^{\pm}\mu^{\mp}$		$\mu^{\pm}\mu^{\mp}$	
	$\Delta\sigma_{t\bar{t}}/\sigma_{t\bar{t}}$ (%)	$\Delta\sigma_{t\bar{t}}$ (pb)	$\Delta\sigma_{t\bar{t}}/\sigma_{t\bar{t}}$ (%)	$\Delta\sigma_{t\bar{t}}$ (pb)
Electron efficiency	1.4	1.0	—	—
Muon efficiency	3.0	2.3	6.1	3.6
Jet energy scale	1.3	1.0	1.3	0.7
Jet energy resolution	< 0.1	< 0.1	< 0.1	< 0.1
Missing transverse momentum	—	—	0.7	0.4
$\mu_R, \mu_F$ scales of $t\bar{t}$ signal (PS)	1.2	0.9	1.7	1.0
$\mu_R, \mu_F$ scales of $t\bar{t}$ signal (ME)	0.2	0.1	1.1	0.6
Hadronization model of $t\bar{t}$ signal	1.2	0.9	5.2	3.1
PDF	0.5	0.4	0.4	0.2
MC sample size	1.4	1.1	2.4	1.4
tW background	1.4	1.1	1.6	0.9
WV background	0.7	0.5	0.9	0.5
Z/ $\gamma^*$ background	2.7	2.1	15	9.1
Non-W/Z background	2.5	1.9	0.7	0.4
Total systematic uncertainty (w/o integrated luminosity)	5.8	4.4	18	11
Integrated luminosity	2.3	1.8	2.3	1.4
Statistical uncertainty	25	19	48	29
Total uncertainty	25	19	52	31

attributed to the integrated luminosity, and the statistical contribution are added in quadrature to obtain the total uncertainty. The cross sections, measured with a relative uncertainty of 25 and 52%, are in agreement with the SM prediction (Eq. (5.1)) within the large uncertainty in the measurements.

### 5.6.3 Combination of the $l$ +jets and dilepton final states

The three individual  $\sigma_{t\bar{t}}$  measurements are combined using the BLUE method [261, 262] to determine the overall  $t\bar{t}$  cross section. All sources of systematic uncertainty are considered as fully correlated across all final states, with the following exceptions: the uncertainty associated with the finite event size of the simulated samples is taken as uncorrelated; the electron identification is not relevant for the  $\mu^{\pm}\mu^{\mp}$  final state; and the b tagging and QCD multijet background uncertainties are only considered for the  $l$ +jets final state. In the  $l$ +jets final state, the WV and Z/ $\gamma^*$  backgrounds are not considered separately but as part of the “Other background” component, which is dominated by tW events. The uncertainty associated with this category is therefore treated as fully correlated with the tW uncertainty in the dileptonic final states and uncorrelated with the

WV and  $Z/\gamma^*$  uncertainties. The individual results used as input to the combination are summarized in Table 5.10, including the correlations assumed between the individual sources of uncertainty.

**Table 5.10:** Inputs to the combination of the  $l$ +jets and dilepton final states, and the QCD analysis (see Section 5.7). The relative importance of statistical and systematic uncertainties in the combined result is attained by implementing the BLUE algorithm with all sources of systematic uncertainty set to zero; the impact of systematic variations is then calculated by quadratically subtracting the statistical and luminosity from the total uncertainty. The post-fit correlation matrix of the nuisance parameters in the  $l$ +jets analysis indicated correlations between the  $\mu_R, \mu_F$  scales in the W+jets background, and between JES uncertainty in signal and the  $\mu_R$  scale in the W+jets background. Additional cross-checks to evaluate whether the assumed correlations significantly affected the outcome of the combination thus involved setting to zero the correlation for the W+jets and JES uncertainty across the three final states, once at a time, while preserving all other correlations [AN9].

Final state	$e^\pm\mu^\mp$	$\mu^\pm\mu^\mp$	$l$ +jets	
Central value (pb)	76.5	59.2	68.9	
Sources of uncertainties (%)				Correlation
b tagging efficiency	-	-	3.4	-
Electron efficiency	1.4	-	1.1	1
Muon efficiency	3.0	6.1	1.7	1
Jet energy scale	1.3	1.3	3.0	1
Jet energy resolution	<0.1	<0.1	0.6	1
$p_T^{\text{miss}}$	-	0.7	-	-
$\mu_R, \mu_F$ scales of $t\bar{t}$ signal (PS)	1.2	1.7	4.4	1
$\mu_R, \mu_F$ scales of $t\bar{t}$ signal (ME)	<0.1	1.1	<0.1	1
Hadronization model of $t\bar{t}$ signal	1.2	5.2	3.7	1
PDF	0.5	0.4	<0.1	1
MC sample event count	1.4	2.4	0.1	0
tW background	1.4	1.6	1.3	1
WV background	0.7	0.9	-	1
$Z/\gamma^*$ background	2.7	15.4	-	1
W+jets background	2.5	0.7	3.5	1
QCD multijet background	-	-	2.4	-
Data sample event count	24.5	51.7	9.5	0
Integrated luminosity	2.3	2.3	2.3	1

The combined inclusive  $t\bar{t}$  cross section is measured to be

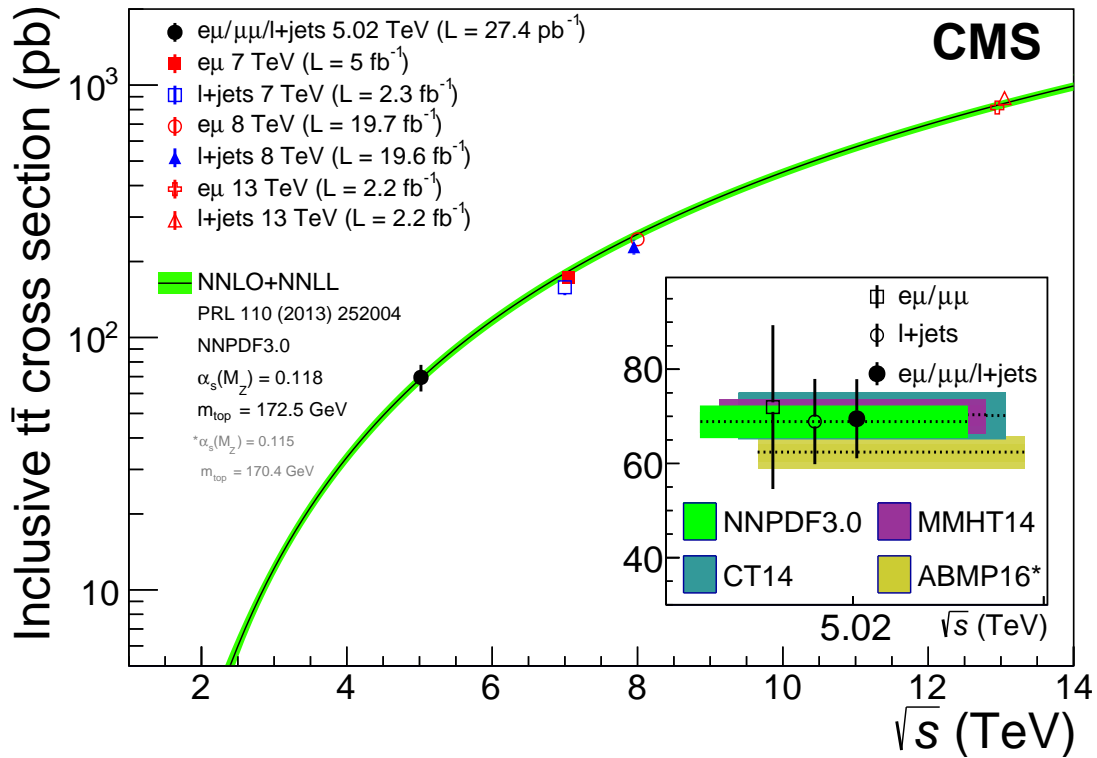
$$\sigma_{t\bar{t}} = 69.5 \pm 6.1 \text{ (stat)} \pm 5.6 \text{ (syst)} \pm 1.6 \text{ (lum)} \text{ pb} = 69.5 \pm 8.4 \text{ (total)} \text{ pb},$$

where the total uncertainty is the sum in quadrature of the individual sources of uncertainty. The weights of the individual measurements, to be understood in the sense of Ref. [262], are 81.8% for  $l$ +jets, 13.5% for  $e^\pm\mu^\mp$ , and 4.7% for  $\mu^\pm\mu^\mp$  final states.

The combined result is found to be robust by performing an iterative variant of the

BLUE method [263]—at each iteration the uncertainty is rescaled and the weights are recalculated—and varying some assumptions on the correlations of different combinations of systematic uncertainty. Also, the post-fit correlations between the nuisance parameters in the  $l$ +jets final state have been checked, and found to have negligible impact by retrieving the combination result as the input  $l$ +jets measurement in the limit of very large uncertainty in the dilepton final states.

Figure 5.16 presents a summary of CMS measurements [75, 246, 247, 250] of  $\sigma_{t\bar{t}}$  in pp collisions at different  $\sqrt{s}$  in the  $l$ +jets and dilepton final states, compared to the NNLO+NNLL prediction using the NNPDF3.0 PDF set with  $\alpha_s(M_Z) = 0.118$  and  $m_{\text{top}} = 172.5$  GeV. In the inset, the results from this analysis at  $\sqrt{s} = 5.02$  TeV are also compared to the predictions from the MMHT14 [34], CT14 [33], and ABMP16 [32] PDF sets, with the latter using  $\alpha_s(M_Z) = 0.115$  and  $m_{\text{top}} = 170.4$  GeV. Theoretical predictions using different PDF sets have comparable values and uncertainties, once consistent values of  $\alpha_s$  and  $m_{\text{top}}$  are associated with the respective PDF set.



**Figure 5.16:** Inclusive  $\sigma_{t\bar{t}}$  in pp collisions as a function of the center-of-mass energy; previous CMS measurements at  $\sqrt{s} = 7, 8$  [246, 247], and 13 [75, 250] TeV in the separate  $l$ +jets and dilepton final states are displayed, along with the combined measurement at 5.02 TeV from this analysis [TH4]. The NNLO+NNLL theoretical prediction [258] using the NNPDF3.0 [35] PDF set with  $\alpha_s(M_Z) = 0.118$  and  $m_{\text{top}} = 172.5$  GeV is shown in the main plot. In the inset, additional predictions at  $\sqrt{s} = 5.02$  TeV using the MMHT14 [34], CT14 [33], and ABMP16 [32] PDF sets, the latter with  $\alpha_s(M_Z) = 0.115$  and  $m_{\text{top}} = 170.4$  GeV, are compared, along with the NNPDF3.0 prediction, to the individual and combined results from this analysis. The vertical bars and bands represent the total uncertainty in the data and in the predictions, respectively.

## 5.7 QCD analysis

To illustrate the impact of the  $\sigma_{\bar{t}t}$  measurements at  $\sqrt{s} = 5.02$  TeV on the knowledge of the proton PDFs the results are used in a QCD analysis at NNLO, together with the combined measurements of neutral- and charged-current cross sections for deep inelastic electron- and positron-proton scattering (DIS) at HERA [36], and the CMS measurement [264] of the muon charge asymmetry in W boson production at  $\sqrt{s} = 8$  TeV. The precise HERA DIS data are directly sensitive to the valence and sea quark distributions and probe the gluon distribution through scaling violations. The latter data set is used to check the feasibility of improving the constraints on the light-quark distributions.

Version 2.0.0 of xFITTER [265, 266], an open-source QCD-analysis framework for PDF determination, is employed, with the partons evolved using the Dokshitzer–Gribov–Lipatov–Altarelli–Parisi equations [267–272] at NNLO, as implemented in the QCDNUM17 – 01/13 program [273]. The treatment and the choices for the central values and variations of the c and b quark masses, the strong coupling, and the strange-quark content fraction of the proton follow that of earlier CMS analyses, e.g., Ref. [264]. The  $\mu_R, \mu_F$  scales are set to the four-momentum transfer in the case of the DIS data (restricted to  $Q_{\min}^2 > 3.5 \text{ GeV}^2$ ), the W boson mass for the muon charge asymmetry results, and the top quark mass in the case of  $\sigma_{\bar{t}t}$ .

The systematic uncertainty in all three measurements of  $\sigma_{\bar{t}t}$  and their correlations are treated the same way as in the combination described in Section 5.6.3. The theoretical predictions for  $\sigma_{\bar{t}t}$  are obtained at NNLO using the HATHOR (v2.1) calculation [274], assuming  $m_{\text{top}} = 172.5 \text{ GeV}$ . The bin-to-bin correlations of the experimental uncertainty in the muon charge asymmetry and DIS measurements are taken into account. The theoretical predictions for the muon charge asymmetry are obtained as described in Ref. [264].

The procedure for the determination of the PDFs follows the approach of HERAPDF2.0 as used in the QCD analysis of Ref. [264]. The parametrized PDFs are the gluon distribution,  $xg$ , the valence quark distributions,  $xu_v$ ,  $xd_v$ , and the u-type and d-type antiquark distributions,  $x\bar{U}$ ,  $x\bar{D}$ . The relations  $x\bar{U} = x\bar{u}$  and  $x\bar{D} = x\bar{d} + x\bar{s}$  are assumed at the initial scale of the QCD evolution  $Q_0^2 = 1.9 \text{ GeV}^2$ . At this scale, the parametrizations are of the form:

$$\begin{aligned}
 xg(x) &= A_g x^{B_g} \times (1-x)^{C_g} \times (1 + D_g x) , \\
 xu_v(x) &= A_{u_v} x^{B_{u_v}} \times (1-x)^{C_{u_v}} \times (1 + D_{u_v} x + E_{u_v} x^2) , \\
 xd_v(x) &= A_{d_v} x^{B_{d_v}} \times (1-x)^{C_{d_v}} , \\
 x\bar{U}(x) &= A_{\bar{U}} x^{B_{\bar{U}}} \times (1-x)^{C_{\bar{U}}} \times (1 + E_{\bar{U}} x^2) , \\
 x\bar{D}(x) &= A_{\bar{D}} x^{B_{\bar{D}}} \times (1-x)^{C_{\bar{D}}} .
 \end{aligned}
 \tag{5.14}$$

The normalization parameters  $A_{u_v}$ ,  $A_{d_v}$ , and  $A_g$  are determined by the QCD sum rules, the  $B$  parameters are responsible for the small- $x$  behavior of the PDFs, and the  $C$  parameters describe the shape of the distribution as  $x \rightarrow 1$ . Additional constraints  $B_{\bar{U}} = B_{\bar{D}}$  and  $A_{\bar{U}} = A_{\bar{D}}(1 - f_s)$  are imposed, with  $f_s$  being the strangeness fraction,  $\bar{s}/(\bar{d} + \bar{s})$ , which is set to  $0.31 \pm 0.08$  as in Ref. [275], consistent with the value obtained using the CMS

**Table 5.11:** Partial  $\chi^2$  per number of data points,  $n_{\text{dp}}$ , and the global  $\chi^2$  per degrees of freedom,  $n_{\text{dof}}$ , as obtained in the QCD analysis of DIS data, the CMS muon charge asymmetry measurements, and the  $\sigma_{t\bar{t}}$  results at  $\sqrt{s} = 5.02$  TeV from this analysis. For the HERA measurements, the energy of the proton beam ( $E_p$ ) is listed for each data set, with the electron/positron energy of 27.5 GeV. The correlated part of the global  $\chi^2$  value is also given. The sources of correlated systematic uncertainty in the  $\sigma_{t\bar{t}}$  measurement are treated as nuisance parameters; for each parameter, a penalty term is added to the  $\chi^2$  [TH4].

Data sets	Partial $\chi^2/n_{\text{dp}}$
HERA neutral current, $e^+p$ , $E_p = 920$ GeV	449/377
HERA neutral current, $e^+p$ , $E_p = 820$ GeV	71/70
HERA neutral current, $e^+p$ , $E_p = 575$ GeV	224/254
HERA neutral current, $e^+p$ , $E_p = 460$ GeV	218/204
HERA neutral current, $e^-p$ , $E_p = 920$ GeV	218/159
HERA charged current, $e^+p$ , $E_p = 920$ GeV	43/39
HERA charged current, $e^-p$ , $E_p = 920$ GeV	53/42
CMS $W^\pm$ muon charge asymmetry	2.4/11
CMS $\sigma_{t\bar{t}}$ , $e^\pm\mu^\mp$ , 5.02 TeV	1.03/1
CMS $\sigma_{t\bar{t}}$ , $\mu^\pm\mu^\mp$ , 5.02 TeV	0.01/1
CMS $\sigma_{t\bar{t}}$ , $l$ +jets, 5.02 TeV	0.70/1
Correlated $\chi^2$	100
Global $\chi^2/n_{\text{dof}}$	1387/1145

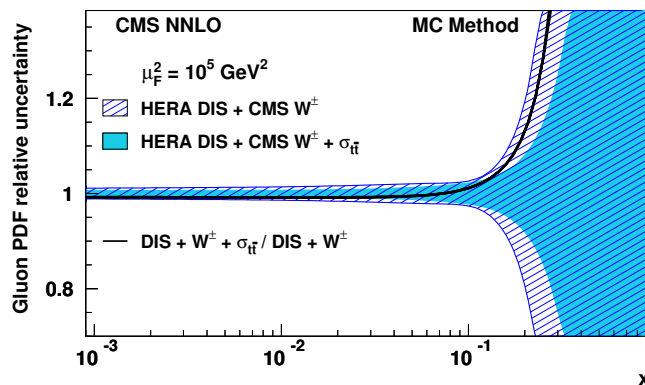
measurements of  $W+c$  production [276].

The predicted and measured cross sections for all the data sets, together with their corresponding uncertainties, are used to build a global  $\chi^2$ , minimized to determine the PDF parameters [265, 266]. The parameters are selected by first fitting with all  $D$  and  $E$  parameters set to zero, and then including them independently one at a time in the fit. The improvement in the global  $\chi^2$  of the fit is monitored, and the procedure is terminated when no further improvement is found. Using the measured values for  $\sigma_{t\bar{t}}$  allows the addition of a new free parameter,  $D_{u_v}$ , in Eq. (5.14), as compared to the analysis in Ref. [264], leading to a 14-parameter fit. The results of the fit are given in Table 5.11.

The quality of the overall fit can be judged based on the global  $\chi^2$  divided by the number of degrees of freedom,  $n_{\text{dof}}$ . For each data set included in the fit, the partial  $\chi^2$  divided by the number of the measurements (data points),  $n_{\text{dp}}$ , is also provided. The correlated part of  $\chi^2$ , also given in Table 5.11, quantifies the influence of the correlated systematic uncertainty in the fit. The global and partial  $\chi^2$  values indicate a general agreement among all the data sets. The low  $\chi^2/n_{\text{dp}}$  value in the  $\mu^\pm\mu^\mp$  final state reflects the large uncertainty in the measurement. The somewhat high  $\chi^2/n_{\text{dp}}$  values for the combined DIS data are very similar to those observed in Ref. [36], where they are investigated in detail.

The experimental uncertainty in the measurements are propagated to the extracted





**Figure 5.17:** The relative uncertainty in the gluon distribution function of the proton as a function of  $x$  at  $\mu_F^2 = 10^5 \text{ GeV}^2$  from a QCD analysis using the HERA DIS and CMS muon charge asymmetry measurements (hatched area), and also including the CMS  $\sigma_{ii}$  results at  $\sqrt{s} = 5.02 \text{ TeV}$  (solid area). The relative uncertainty is found after the two gluon distributions have been normalized to unity. The solid line shows the ratio of the gluon distribution function found from the fit with the CMS  $\sigma_{ii}$  measurements included to that found without [TH4].

QCD fit parameters using the MC method [277, 278]. In this method, 400 replicas of pseudo-data are generated with the measured values for  $\sigma_{ii}$  allowed to vary within the statistical and systematic uncertainties, and taking into account their correlations (Table 5.10). For each replica, the PDF fit is performed, and the uncertainty is estimated as the RMS around the central value. In Fig. 5.17, the ratio and the relative uncertainty in the gluon distributions, as obtained in the QCD analyses with and without the measured values for  $\sigma_{ii}$  at  $\mu_F^2 = 10^5 \text{ GeV}^2$ , are shown. A moderate reduction of the uncertainty in the gluon distribution at  $x \gtrsim 0.1$  is observed, once the measured values for  $\sigma_{ii}$  are included in the fit. The uncertainty in the valence quark distributions remains unaffected (Fig. 5.18). All changes in the central values of the PDFs are well within the fit uncertainty. The latter is determined using the tolerance criterion of  $\Delta\chi^2 = 1$ .

Possible effects from varying the model input parameters and the initial PDF parametrization are investigated in the same way as in the similar analysis of Ref. [264]. The modeling uncertainty arises from the variations in the values assumed for the  $c$  quark mass, the strangeness fraction, and the value of  $Q_{\text{min}}^2$  imposed on the HERA data. The parametrization uncertainty is estimated by varying the functional form of the PDFs with the parameters  $D$  and  $E$  added or removed one at a time, and the value of  $Q_0^2$ . The two cases when the measured values for  $\sigma_{ii}$  are included or excluded from the fit are considered, resulting in the same associated model and parametrization uncertainties.

In conclusion, the  $\sigma_{ii}$  measurements at  $\sqrt{s} = 5.02 \text{ TeV}$  provide improved uncertainty in the gluon PDF at high  $x$ , though the impact is small, owing to the large experimental uncertainty.

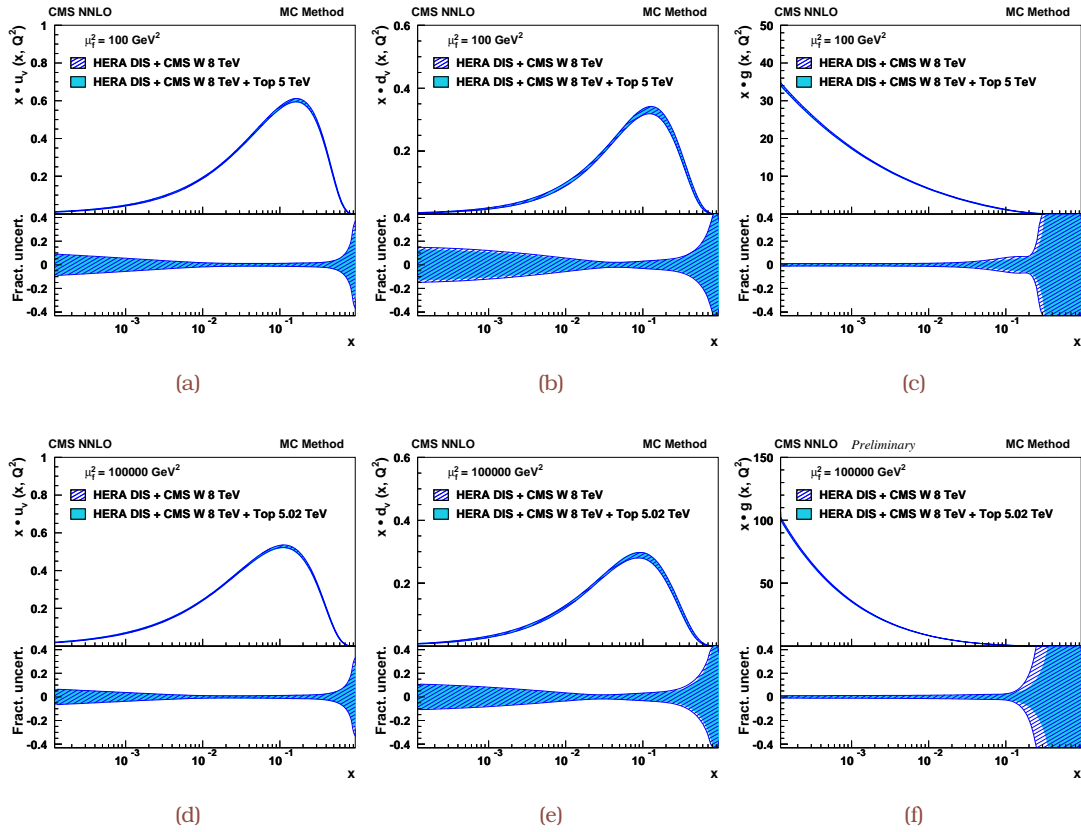


Figure 5.18: The valence quark (a,b,d,e) and gluon (c,f) distribution functions of the proton and their relative uncertainties (bottom panels) as a function of  $x$  at  $\mu_F^2 = 10^2$  and  $10^5 \text{ GeV}^2$  from a QCD analysis using the HERA DIS and CMS muon charge asymmetry measurements (hatched area), and also including the CMS  $\sigma_{t\bar{t}}$  results at  $\sqrt{s} = 5.02 \text{ TeV}$  (solid area) [AN9].

## Chapter 6: Observation of top quark production in proton-nucleus collisions

### Contents

<b>6.1 Breaking the barrier of low <math>\sqrt{s_{NN}}</math> and event count</b>	<b>176</b>
<b>6.2 Event sample description</b>	<b>176</b>
<b>6.3 Event sample selection and physics objects reconstruction</b>	<b>177</b>
<b>6.4 Signal and background parameterization</b>	<b>186</b>
6.4.1 Signal parameterization	187
6.4.2 Background parameterization	188
<b>6.5 Systematic uncertainty</b>	<b>193</b>
6.5.1 Experimental uncertainty	194
6.5.2 Theory uncertainty	195
<b>6.6 Measurement of the inclusive <math>t\bar{t}</math> cross section</b>	<b>196</b>
6.6.1 Fit model	196
6.6.2 Extraction of the inclusive $t\bar{t}$ cross section	196
6.6.3 Impact of the statistical and systematic uncertainties	198
6.6.4 Statistical significance of the signal	200

The first observation of top quark production in proton-nucleus collisions is reported in this Chapter, using proton-lead data collected by the CMS experiment at the CERN LHC at a nucleon-nucleon center-of-mass energy of  $\sqrt{s_{NN}} = 8.16$  TeV. The measurement is performed using events with exactly one isolated electron or muon candidate and at least four jets. The data sample corresponds to an integrated luminosity of  $174 \text{ nb}^{-1}$ . The significance of the  $t\bar{t}$  signal against the background-only hypothesis is above five standard deviations. The measured cross section is  $\sigma_{t\bar{t}} = 45 \pm 8 \text{ nb}$ , consistent with predictions from perturbative quantum chromodynamics.

The original contribution from the present thesis to the material related to the following Chapter and listed in [Scientific output](#) and [Internal notes](#) is:

- The data analysis that resulted in this measurement.
- The entire responsibility of the signal and background event generation. For that purpose, the default configuration of the used NLO generator had to be further tuned to include nuclear modifications on the gluon PDF.
- The partial responsibility of the internal stages that the paper had to pass through before it left the Collaboration and after the interaction with the journal to ensure that it communicated physics with clarity.

## 6.1 Breaking the barrier of low $\sqrt{s_{\text{NN}}}$ and event count

Top quark cross sections at LHC are dominated by pair production via gluon-gluon fusion processes ( $gg \rightarrow t\bar{t} + X$ ), and can be computed with great accuracy in perturbative quantum chromodynamics [257, 258]. In proton-nucleus collisions, the top quark is a novel and theoretically precise probe of the nuclear gluon density at high virtualities  $Q^2 \approx m_t^2$  (where  $m_t$  is the top quark mass) in the unexplored high Bjorken- $x$  region ( $x \gtrsim 2m_t/\sqrt{s_{\text{NN}}} \approx 0.05$ ) [87, 279]. In this region, “antishadowing” and “EMC” effects [85] are expected to modify the gluon density with respect to that in the free-proton case [86, 88]. The production of top quarks thus provides information on the nuclear parton distribution functions (nPDF) that is complementary to that obtained through studies of electroweak boson production. In comparison to the W and Z cases [204, 280, 281], top-pair cross sections are more sensitive to gluon (rather than quark) densities, at Bjorken- $x$  values about twice as large. Novel studies of parton energy loss using top quarks in the quark-gluon plasma formed in nucleus-nucleus collisions have also been proposed [87, 93, 282]. A good understanding of top quark production in proton-nucleus collisions is crucial as a baseline for these studies.

Since the  $l$ +jets final state features a large branching fraction ( $\approx 34\%$  for the combined  $e$ +jets and  $\mu$ +jets channels including events from the  $t \rightarrow W \rightarrow \tau \rightarrow e, \mu$  decay chain) and moderate background contamination it provides favorable conditions for the detection of  $t\bar{t}$  production in proton-nucleus collisions. This Chapter describes the first observation of top quark production in nuclear collisions that is carried out with pPb collisions collected by the CMS experiment at  $\sqrt{s_{\text{NN}}} = 8.16 \text{ TeV}$ , using  $t\bar{t}$  candidates. The event topology comprises exactly one isolated charged lepton, either a muon or an electron, accompanied by the presence of at least four jets, stemming from the hadronization of the b quarks, and of the light quarks from the decay of one of the W bosons. The  $t\bar{t}$  cross section is then extracted from a combined maximum likelihood fit of the invariant mass of the two light-quark jets, in different categories of events with zero, one, or at least two b-tagged jets.

## 6.2 Event sample description

The event sample of proton-nucleus collisions collected by the CMS detector in 2016 corresponds to an integrated luminosity of  $174 \pm 6 \text{ nb}^{-1}$  [TH2]. The lead nuclei and protons had beam energies of 2.56 and 6.5 TeV per nucleon, respectively, corresponding to a nucleon-nucleon centre-of-mass energy of  $\sqrt{s_{\text{NN}}} = 8.16 \text{ TeV}$ . The direction of the proton beam was initially clockwise (“Pbp” configuration) and was then reversed (“pPb” configuration). Owing to the energy difference between the proton-lead colliding beams, the nucleon-nucleon center-of-mass (CM) frame is not at rest with respect to the laboratory (lab) frame. Massless particles emitted at a pseudorapidity  $\eta_{\text{CM}}$  in the CM frame

experience a longitudinal boost according to.

$$|\Delta\eta_{\text{CM}}| = \frac{1}{2} \times \left| \ln \left( \frac{Z_{82}^{208\text{Pb}^+} \times A_p}{Z_p \times A_{82}^{208\text{Pb}^+}} \right) \right| = \frac{1}{2} \times \ln \left( \frac{208}{82} \right) = 0.465 \quad 6.1$$

The pseudorapidity  $\eta_{\text{lab}}$  is defined such as to have positive value in the direction of motion of the proton in both Pbp and pPb data samples. The average number of collisions per bunch crossing is unity in the combined data set, and assuming a pp inelastic cross section of 69.2 mb multiplied by  $A = 208$ .

The  $pN \rightarrow t\bar{t} + X$  process ( $N = p, n$ ) is simulated using the LO PYTHIA (v6.424 [49], tune Z2\* [58, 177]) and the NLO POWHEG (v2 [39, 40], tune CUETP8M1 [58, 59]) generators with a mixture of pp and pn interactions corresponding to their ratio in pPb collisions. The nuclear modification of the up- and down-type valence quark and gluon distribution functions are rendered using the EPPS16 [88] nuclear PDFs for the  ${}_{82}^{208\text{Pb}^+}$  ions. The quark densities are scaled according to the isospin symmetry

$$\begin{aligned} f_{82}^{u, 208\text{Pb}^+}(A, Z) &= \frac{Z}{A} \left( R_s^u f_p^{\bar{u}} + R_v^u (f_p^u - f_p^{\bar{u}}) \right) + \frac{A-Z}{A} \left( R_s^d f_p^{\bar{d}} + R_v^d (f_p^d - f_p^{\bar{d}}) \right) \\ f_{82}^{d, 208\text{Pb}^+}(A, Z) &= \frac{Z}{A} \left( R_s^d f_p^{\bar{d}} + R_v^d (f_p^d - f_p^{\bar{d}}) \right) + \frac{A-Z}{A} \left( R_s^u f_p^{\bar{u}} + R_v^u (f_p^u - f_p^{\bar{u}}) \right) \end{aligned} \quad 6.2$$

where  $f_{82}^{u, 208\text{Pb}^+}$  and  $f_{82}^{d, 208\text{Pb}^+}$  represent the up- and down-type valence quark PDF inside the  ${}_{82}^{208\text{Pb}^+}$  ion, and  $R_v$  ( $R_s$ ) parameterize the nuclear modification factor for valence (sea) quarks. The value of  $m_t$  used in all simulated samples is 172.5 GeV.

Simulated samples of W+jets and Drell-Yan production of charged-lepton pairs with an invariant mass larger than 30 GeV are generated using PYTHIA 6. The MC is used solely for efficiency measurements and validation of the functional forms used for the background distributions since the latter is determined *in situ* from the data. All signal and background samples are embedded (see Section 3.3.1) into pPb events generated with EPOS-LHC [192] (v.3400), tuned to reproduce the global pPb event properties experimentally measured, and reconstructed with the same analysis code as used for the data. The kinematics of all MC-generated events are boosted to account for the different energies of the proton and lead beam. Simulated samples include an emulation of the full detector response, based on GEANT4 [194], with simulated alignment and calibration conditions tuned on data, and a realistic description of the luminous region (see Section 3.2.4) produced by the collisions.

## 6.3 Event sample selection and physics objects reconstruction

### Event selection

This analysis is restricted to events that fired trigger paths requiring the presence of at least one muon (electron) candidate with transverse momentum (energy)  $p_T > 12$  GeV ( $E_T > 20$  GeV). Looser online identification criteria are applied as compared to the offline selection, and no requirement on additional analysis objects is imposed at this level.

**Table 6.1:** MC data samples for the inclusive  $\sigma_t$  measurement using ppb collisions at  $\sqrt{s_{NN}} = 8.16 \text{ TeV}$  [AN2]. The samples are generated either inclusively or with a final state restricted to the leptonic mode, including either electrons or muons. The absence of reference delineates cross section values calculated from the generator. For the samples restricted to specific decay channels, the branching ratio is included in the quoted cross section value. The samples with "Emb" in the name are embedded using the EPOS-LHC (v.3400) generator (see Section 3.3.1).

Process	$\sigma[\text{nb}]$	MC event sample indicative description
$t\bar{t}$	59.0 (NNLO+NNLL) [256]	TTbar_pPb-EmbEP05_8160GeV_pythia6 TTbar_Pbp-EmbEP05_8160GeV_pythia6 TT_TuneCUETP8M1_pPb-EmbEP05_8160GeV_powheg_pythia8 TT_TuneCUETP8M1_Pbp-EmbEP05_8160GeV_powheg_pythia8
W+jets	1970	WJetstollNu_TuneZ2_pPb-EmbEP05_8160GeV_pythia6 WJetstollNu_TuneZ2_Pbp-EmbEP05_8160GeV_pythia6
$Z/\gamma^*(\rightarrow \ell^\pm \ell^\mp) + \text{jets}$	229	DYJetstoll_TuneZ2_M-30_pPb-EmbEP05_8160GeV_pythia6 DYJetstoll_TuneZ2_M-30_Pbp-EmbEP05_8160GeV_pythia6

Particle candidates are reconstructed offline with the CMS PF algorithm, which identifies and provides a list of particles using an optimized combination of information from the various elements of the CMS detector. Events are required to contain exactly one tight muon (see Section 3.3.2) or medium electron (see Section 3.3.3) candidate, with  $p_T > 30$  GeV and  $|\eta| < 2.1$ , excluding in the electron case the transition region  $1.444 < |\eta| < 1.566$  between the ECAL barrel and endcap, where the reconstruction of electron objects is less efficient. The muon and electron candidates are required to be isolated from nearby hadronic activity within a cone of  $\Delta R = 0.3$  around the direction of the track at the primary event vertex, where  $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ , and  $\Delta\eta$  and  $\Delta\phi$  are the separations in pseudorapidity and azimuthal angle. A charged lepton is selected if its relative isolation discriminant value satisfies  $I_{\text{rel}} < 0.15$  (muon), 0.07 (electron in the barrel), or 0.08 (electron in one of the endcaps). These thresholds have been optimized to reduce the contamination from nonprompt leptons. To remove the Drell-Yan background, events are rejected from the analysis if they contain extra electrons (muons) that are reconstructed using a looser set of identification criteria and have  $p_T > 20$  (15) GeV within  $|\eta| < 2.5$  (2.4). The efficiency of the lepton selection is measured using a tag-and-probe method in events enriched with Z boson candidates and selected by the same trigger requirements as the signal candidate events. The combined reconstruction, lepton identification, and trigger efficiency is determined as a function of lepton  $p_T$  and  $\eta$ .

Events are required to have at least four reconstructed jets with  $p_T > 25$  GeV and  $|\eta| < 2.5$ , that are separated by at least  $\Delta R = 0.3$  from the selected muon or electron. Jets are reconstructed from the PF candidates using the anti- $k_T$  clustering algorithm with a distance parameter of 0.4. Jet energy corrections extracted from the full detector simulation are applied as functions of jet  $p_T$  and  $\eta$  [187, 198] to both data and simulated samples. A residual correction to the data is applied to account for a small data-MC discrepancy in the jet energy response. Jets from b quarks are tagged based on the presence of a secondary vertex from B-hadron decays, identified using a multivariate algorithm combining tracking information.

The distinct  $t\bar{t}$  signature of two b jets in the event, which rarely occurs in background processes such as W+jets and QCD multijet (collectively labeled as “non-top” background), is used to extract the signal. The number of jets passing a threshold on the b-jet identification discriminant, corresponding to a b tagging efficiency of approximately 70% with a misidentification rate of less than 0.1% for light-flavor jets, as estimated in simulated pp and cross-checked with pPb events (Table 3.5), is used to classify the selected events into no (0b), exactly one (1b), or at least two ( $\geq 2$ b) tagged-jet categories, described in the following. All three event categories are exploited in a maximum-likelihood fit in order to extract the signal cross section, and simultaneously constrain the background contamination and determine the efficiency of the b jet identification.

Table 6.2 summarizes the selection requirements that are applied in the analysis to select a high purity  $t\bar{t} \rightarrow \ell + \text{jets}$ ,  $\ell = e$  or  $\mu$ , sample.

**Table 6.2:** Selection criteria applied for the inclusive  $t\bar{t}$  cross section measurement in the  $\mu$ +jets and  $e$ +jets final states using pPb collisions at  $\sqrt{s_{NN}} = 8.16$  TeV.

Filter or physics object	Selection
Trigger	one $\mu$ ( $e$ ) candidate, $p_T > 12$ ( $E_T > 20$ ) GeV
Electrons	medium ID, $p_T > 30$ GeV, $ \eta  < 2.1$ , $I_{\text{rel}}^{\rho, \text{corr}} < 0.08$
Muons	tight ID, $p_T > 30$ GeV, $ \eta  < 2.1$ , $I_{\text{rel}}^{\delta\beta, \text{corr}} < 0.15$
Jets	loose ID, $p_T > 25$ GeV, $ \eta  < 2.5$
Jet multiplicity	$\geq 4$
b tagging	CSVv2M (Table 3.5) $\Rightarrow$ 1 $\ell$ 4j0b, 1 $\ell$ 4j1b, and 1 $\ell$ 4j2b exclusive categories

### Event categorization

As we observed in the measurement of the inclusive  $t\bar{t}$  cross section in the  $\ell$ +jets final state using pp collisions at  $\sqrt{s} = 5.02$  TeV [TH4], it is expected the main contamination from background processes to originate from QCD multijet and W+jets events mainly. Given the characteristics of these processes, categorization of events, according to the number of b-tagged jets, separates signal from background. The reconstruction of the kinematics of the W boson and top quark decays should further distinguish between the resonant nature of the signal and the continuum from the background.

Figure 6.1 summarizes the event categories of the analysis that are attained based on identification and isolation requirements of the lepton candidates, and the heavy-flavor content of the jets, i.e., identified as b or light quark jets. The signal dominates the event categories with at least one lepton and four jets, two of which are b tagged (the “1 $\ell$ 4j2b” category). If only one jet passes the b-tagging requirement the event is classified to the “1 $\ell$ 4j1b” category, while in case none of the jets satisfies the b-tagging requirement the event falls in the “1 $\ell$ 4j0b” category. To model the QCD multijet background an additional category is considered where the lepton fails identification or isolation requirements, and to further reduce the contamination from W+jets and  $t\bar{t}$  events in the QCD multijet control region no jet should fulfill the b-tagging requirement. The category is referred to as “1f4j0b”, with “f” standing for fake.

### Reconstruction of the W boson leptonic and hadronic decays

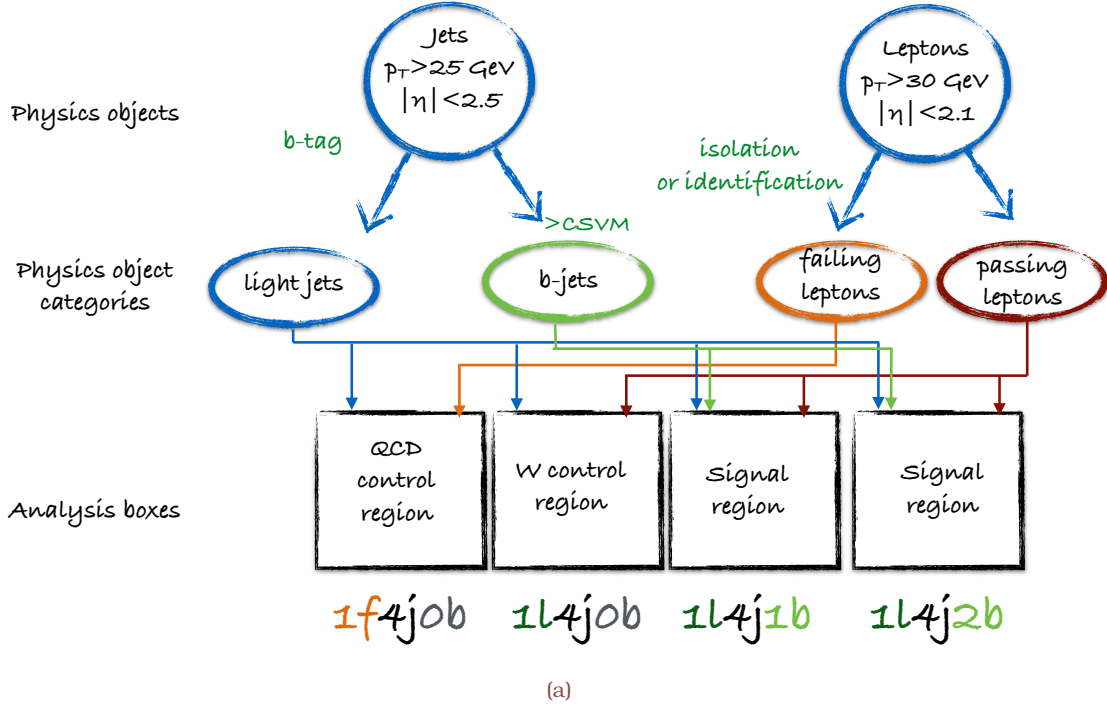
In the case of W boson leptonic decay in a charged lepton and associated neutrino the reconstructed  $p_T^{\text{miss}}$  (see Section 3.3.6) is used to infer the  $p_T$  of the neutrino. The longitudinal component of the neutrino momentum,  $p_{v,z}$ , is computed by using the energy-momentum conservation at the  $W_{\ell\nu}$  vertex and constraining the W boson mass to  $m_W = 80.4$  GeV [16]. This leads to a quadratic equation in  $p_{v,z}$  with solutions as

$$p_{z,v} = \frac{\Lambda p_{z,\ell}}{p_{T,\ell}^2} \pm \frac{1}{p_{T,\ell}^2} \sqrt{\Lambda^2 p_{z,\ell}^2 - p_{T,\ell}^2 (E_\ell^2 p_T^{\text{miss},2} - \Lambda^2)}, \quad 6.3$$

where

$$\Lambda = \frac{m_W^2}{2} + \vec{p}_{T,\ell} \times \vec{p}_T^{\text{miss}}, \quad 6.4$$



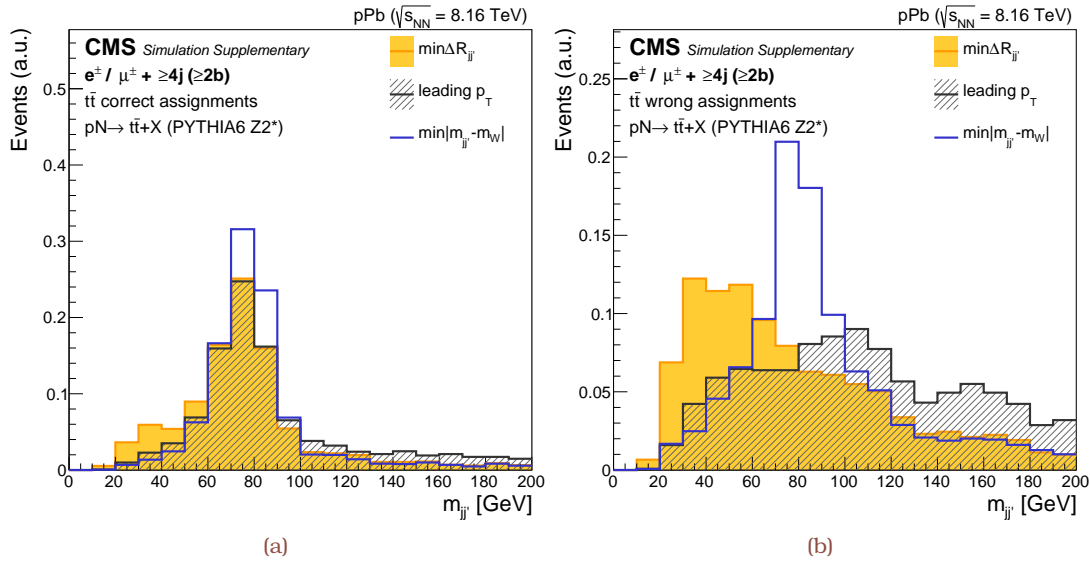


**Figure 6.1:** Representation of the physics objects and event categories (analysis boxes) for the inclusive  $t\bar{t}$  cross section in the  $l$ +jets final state using pPb collisions at  $\sqrt{s_{NN}} = 8.16 \text{ TeV}$  [AN2].

and  $E_l^2 = p_{T,l}^2 + p_{z,l}^2$  denotes the energy of the charged lepton. In most of the cases this leads to two real solutions for  $p_{z,v}$  and the solution that minimizes  $|p_{v,z} - p_{l,z}|$  is chosen [TH13]. For some events the discriminant in Eq. (6.3) becomes negative leading to complex solutions for  $p_{z,v}$ . In that case, the imaginary component is eliminated by the modification of  $\vec{p}_T^{\text{miss}}$  so that  $m_T = m_W$ , while still respecting the  $m_W$  constraint. This is achieved by imposing that the determinant, and thus the square-root term in Eq. (6.3), is null. This condition gives a quadratic relation between  $p_{x,v}$  and  $p_{y,v}$  with two possible solutions, and one remaining degree of freedom. The solution is chosen by finding the neutrino transverse momentum  $\vec{p}_{T,v}$  that has the minimum vectorial distance from the  $\vec{p}_T^{\text{miss}}$  in the  $p_x$ - $p_y$  plane. With the kinematics of the neutrino fully specified, the kinematics of the leptonically decaying W boson can be then computed as  $P_W = P_l + P_\nu$ , where  $P_i$  is the four-vector of particle species  $i$ .

In case of W boson hadronic decays in two light quarks,  $q$  and  $\bar{q}'$ , all selected jets,  $j$  and  $j'$ , are used after removing up to two b-tagged candidates using the CSVv2M working point (Table 3.5). When more than two jets are left a sorting algorithm based on the proximity of the jets by  $\Delta R$  is applied. The  $\Delta R_{\min}$  criterion preserves a high efficiency with respect to the number of events in which both jets from the  $W \rightarrow q\bar{q}'$  decay are available while reducing the bias, i.e., any mimicking effects from the background. A criterion based on the closeness of  $m_{jj'}$  to the W boson mass maximizes efficiency but at the same time biases considerably the background. To that end, in the analysis the  $\Delta R_{\min}(j, j')$  sorting algorithm is used, similar to the what has been found to be optimal in [TH4].

A comparison of the sorting algorithms based on simulation can be found in Fig. 6.2



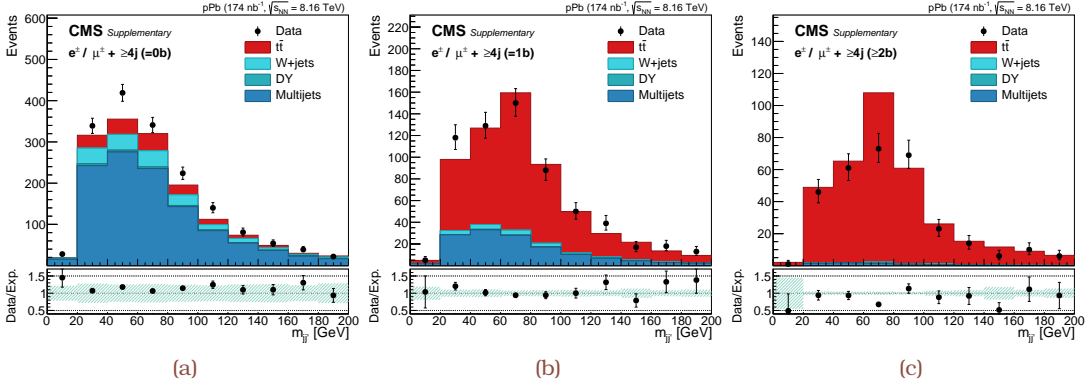
**Figure 6.2:** Dijet invariant mass ( $m_{jj}$ ) spectrum in the  $1l4j2b$  event category calculated using different algorithms in the pairing of the jets in each event: distance-based ( $\Delta R_{\min}(j, j')$ ),  $p_T$ -based (leading  $p_T$ ), and minimizing the distance to the W boson mass  $m_W$  ( $\min|m_{jj} - m_W|$ ) [TH1]. Pairs fully matched to  $W \rightarrow q\bar{q}'$  decays are shown in (a), while pairs with at least one reconstructed jet not matched at parton level are shown on panel (b). The results are based on PYTHIA (v6.424 [49])  $t\bar{t}$  simulation, and using the Z2\* [58, 177] tune.

separately for “correct” and “wrong” assignments. Successful combinations of  $j$  and  $j'$  for which the algorithms are able to reconstruct both jets geometrically matched to the corresponding generator level quarks from the W or t decays are shown in Fig. 6.2a. Wrong or unmatched combinations of  $j$  and  $j'$  for which the algorithms either select jets that cannot be geometrically matched to generator level information or mix these physics objects from different decays are displayed in Fig. 6.2b.

### Top pair decay reconstruction

Once the W bosons have been reconstructed they should be optimally combined with the b-jet candidates to retrieve the  $t \rightarrow Wb$  decay chains. In cases where jets do not fulfill the b-tagging threshold, the two jets with the highest CSVv2 discriminator value are used as the b-jet candidates. In the rest of the cases, the pairing of each b-jet candidate to W boson candidates can be tested using different ranking algorithms. On the one hand, the proximity-based metric  $\Delta R_{\min}$  is expected to outperform in high- $p_T$  events, and top-quark-mass-based criteria may induce a bias on the background. On the other hand, a less biased mass requirement consists of minimizing the difference between the mass of the hadronic (all jets,  $t_{\text{had}}$ ) and leptonic ( $b\ell\nu$ ,  $t_{\text{lep}}$ )

Table 6.3 compares the expected efficiency for the different sorting algorithms, and further comparisons of the distributions for both correctly and wrongly assigned objects in the reconstruction of top quarks in simulated W+jets events are given in the following. We conclude that the strategy based on minimizing the  $t_{\text{had}}$  and  $t_{\text{lep}}$  mass difference is expected

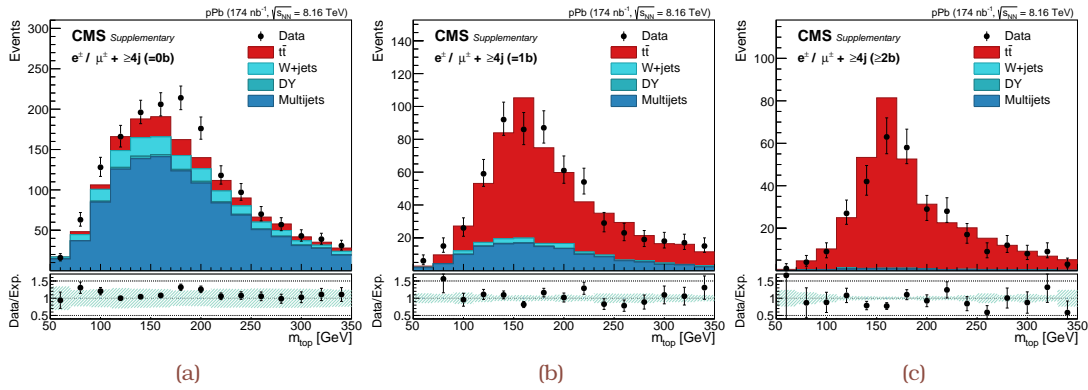


**Figure 6.3:** Dijet invariant mass ( $m_{jj}$ ) distributions in the  $1l4j0b$  (a),  $1l4j1b$  (b), and  $1l4j2b$  (c) event categories after the complete event selection. On the upper panels, the reconstructed data are compared to the stacked expected contributions from the signal and the main background processes [TH1]. The  $t\bar{t}$ , W+jets, and Drell-Yan (DY) processes are simulated with PYTHIA (v6.424 [49], tune Z2\* [58, 177]) and normalized to the theoretical cross sections (Table 6.1). The QCD multijet (Multijets) contribution is estimated from data using a control region. The bottom panels display the ratio between the data and the expectations. The shaded band represents the relative uncertainty due to the limited statistics in the simulated samples and in the estimate of the normalization of the QCD multijet background.

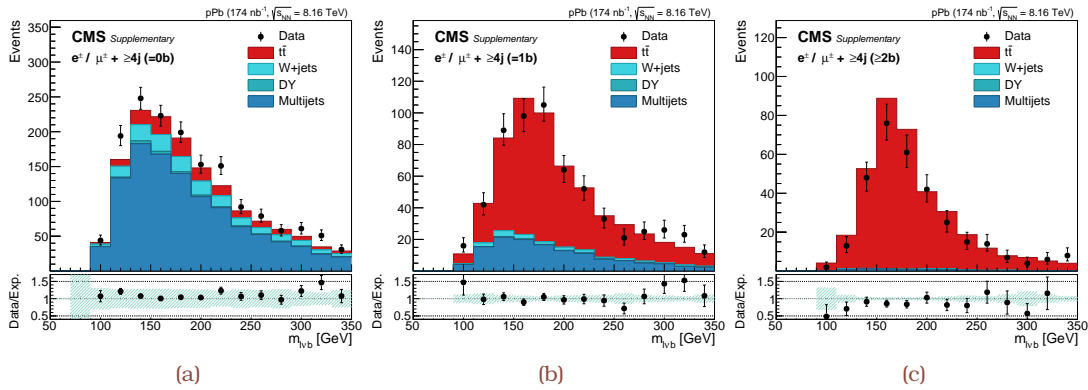
**Table 6.3:** Fraction of events in which two b-tagged jets are successfully matched to parton level b quarks from top decays (2b) and are subsequently correctly matched to reconstruct the  $t_{\text{had}}$  and  $t_{\text{lep}}$  decay chains. The pairing was tested using three metrics, i.e., a proximity-based ( $\Delta R$ ) and two mass-based metrics that minimize the mass difference of either  $t_{\text{had}}$  and  $t_{\text{lep}}$  quarks ( $\min |m_{t_{\text{had}}} - m_{t_{\text{lep}}}|$ ) or the mass difference of  $t_{\text{had}}$  quark to the world average [16]  $m_t$  value ( $\min |m_{t_{\text{had}}} - m_t|$ ) [AN2].

Category	2b	$t_{\text{had}}$	$t_{\text{lep}}$	Algorithm
$1l4j0b$	$0.170 \pm 0.004$	$0.098 \pm 0.003$	$0.307 \pm 0.005$	$\min \Delta R$
		$0.108 \pm 0.003$	$0.242 \pm 0.005$	$\min  m_{t_{\text{had}}} - m_t $
		$0.089 \pm 0.003$	$0.289 \pm 0.005$	$\min  m_{t_{\text{had}}} - m_{t_{\text{lep}}} $
$1l4j1b$	$0.544 \pm 0.008$	$0.198 \pm 0.006$	$0.468 \pm 0.008$	$\min \Delta R$
		$0.200 \pm 0.006$	$0.458 \pm 0.008$	$\min  m_{t_{\text{had}}} - m_t $
		$0.194 \pm 0.006$	$0.454 \pm 0.008$	$\min  m_{t_{\text{had}}} - m_{t_{\text{lep}}} $
$1l4j2b$	$0.791 \pm 0.005$	$0.265 \pm 0.006$	$0.540 \pm 0.006$	$\min \Delta R$
		$0.269 \pm 0.006$	$0.535 \pm 0.006$	$\min  m_{t_{\text{had}}} - m_t $
		$0.266 \pm 0.006$	$0.541 \pm 0.006$	$\min  m_{t_{\text{had}}} - m_{t_{\text{lep}}} $

to have comparable efficiency to the other sorting algorithms. The resolution is expected to be improved with respect to a  $\Delta R_{\text{min}}$  ranking while avoiding to induce significant bias from the background, i.e., a peak close to the top mass in the reconstructed distribution. Therefore in the analysis the b-jets are combined with the W boson candidates based on the  $\min |m_{t_{\text{had}}} - m_{t_{\text{lep}}}|$  criterion. The output prior to any fit are given in Figs. 6.4 and 6.5.



**Figure 6.4:** Hadronic top mass ( $m_{\text{top}}$ ) distributions in the 0 (a), 1 (b), and 2 (c) b-tagged jet categories after all selections. On the upper panels, the reconstructed data are compared to the stacked expected contributions from signal and the main background processes [TH1]. The  $t\bar{t}$ , W+jets, and Drell-Yan (DY) processes are simulated with PYTHIA (v6.424 [49], tune Z2\* [58, 177]) and normalized to the expected cross sections and integrated luminosity. The QCD multijet (Multijets) contribution is estimated from data using a control region. The bottom panels display the ratio between the data and the expectations. The shaded band represents the relative uncertainty due to the limited statistics in the simulated samples and in the estimate of the normalization of the QCD multijet background.

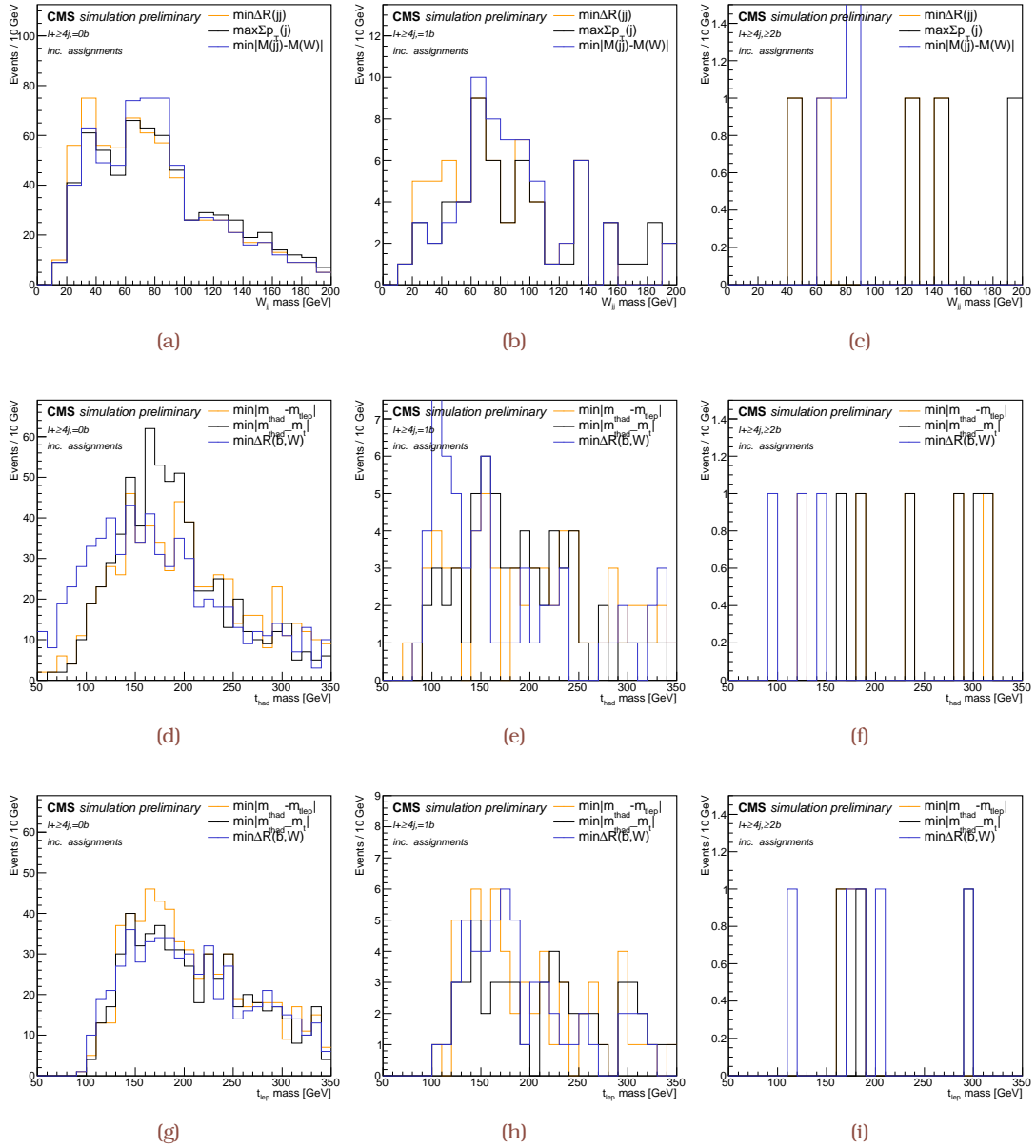


**Figure 6.5:** Leptonic top mass ( $m_{\text{top}}$ ) distributions in the 0 (a), 1 (b), and 2 (c) b-tagged jet categories after all selections. On the upper panels, the reconstructed data are compared to the stacked expected contributions from signal and the main background processes [TH1]. The  $t\bar{t}$ , W+jets, and Drell-Yan (DY) processes are simulated with PYTHIA (v6.424 [49], tune Z2\* [58, 177]) and normalized to the expected cross sections and integrated luminosity. The QCD multijet (Multijets) contribution is estimated from data using a control region. The bottom panels display the ratio between the data and the expectations. The shaded band represents the relative uncertainty due to the limited statistics in the simulated samples and in the estimate of the normalization of the QCD multijet background.

### W boson and top pair-like reconstruction in W+jets background samples

The expected effect of the reconstruction and sorting algorithms on the W+jets background is evaluated using MC event simulation. Because of the low event count in the

default pPb simulation, a larger inclusive W+jets simulation is also investigated, using pp collisions at 8 TeV and based on the LO MADGRAPH5 [176] generator. The results are shown in Figs. 6.6 and 6.7 for the pPb and pp MC event samples, respectively. Although the statistical precision is low in both samples, and especially in the 1 $l$ 4j2b category, It can be seen that for the chosen sorting algorithm  $\min|m_{t_{\text{had}}} - m_t|$  no bias is expected either in the  $m_{jj}$  or the hadronic and leptonic top quark mass distributions, irrespective of the event category within statistical uncertainties. An additional algorithm for reconstructing the  $t_{\text{had}}$  candidates based on their closeness to the world average  $m_t$  [16] reveals top mass distributions that are severely biased for the W+jets background.



**Figure 6.6:** Dijet invariant mass (a–c), and hadronic (d–f) and leptonic (g–i) top mass distributions in the 0 (a, d, g), 1 (b, e, h), and 2 (c, f, i) b-tagged jet categories after all selections [AN2], using the simulated W+jets MC event sample of pPb collisions (Table 6.1). The curves correspond to the different sorting algorithms as described in Table 6.3.

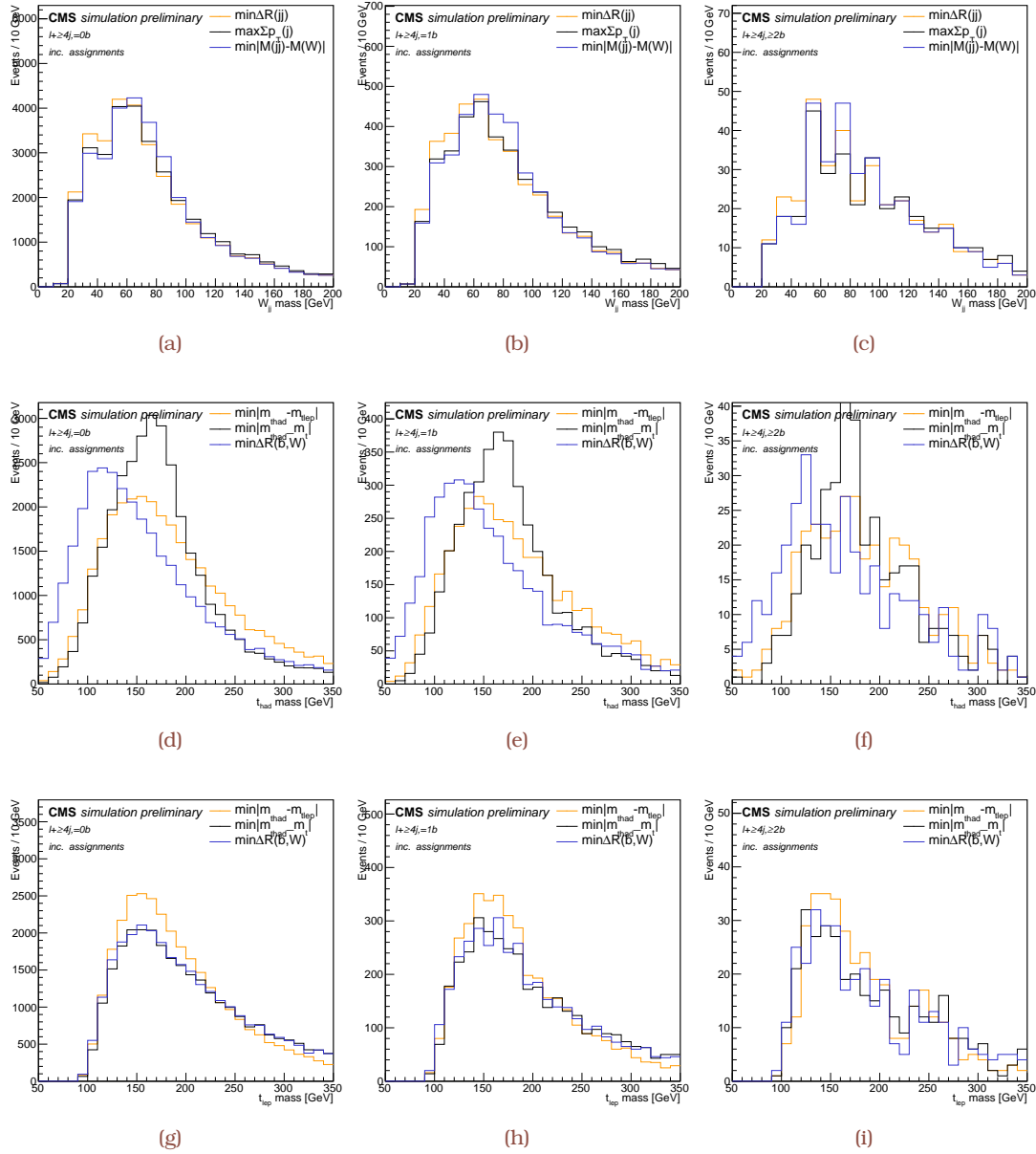
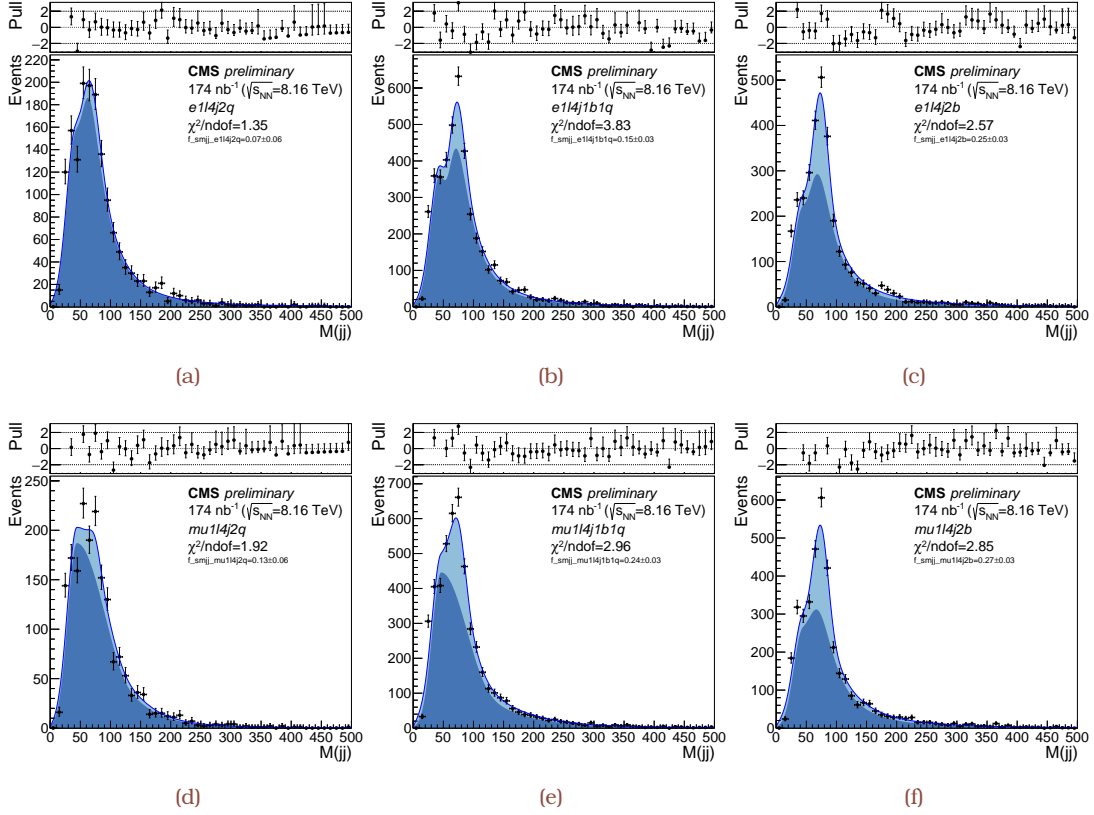


Figure 6.7: Dijet invariant mass (a-c), and hadronic (d-f) and leptonic (g-i) top mass distributions in the 0 (a, d, g), 1 (b, e, h), and 2 (c, f, i) b-tagged jet categories after all selections, using a higher event count W+jets MC sample of pp collisions at 8 TeV [AN2]. The curves correspond to the different sorting algorithms as described in Table 6.3.

## 6.4 Signal and background parameterization

In this section, the fit model applied to extract the inclusive  $\sigma_{\bar{t}t}$  is discussed. The  $\bar{t}t$  MC simulation is used to guide the parameterization of the signal, while the shape of the background is estimated *in-situ* from the data. The normalization of both signal and background is left to float freely. The efficiency of the b-identification algorithm is measured simultaneously with the signal cross section, and hence it is also data-based estimated. The background from QCD multijet and W+jets events are controlled by counting the number of jets identified as originating from b quarks.

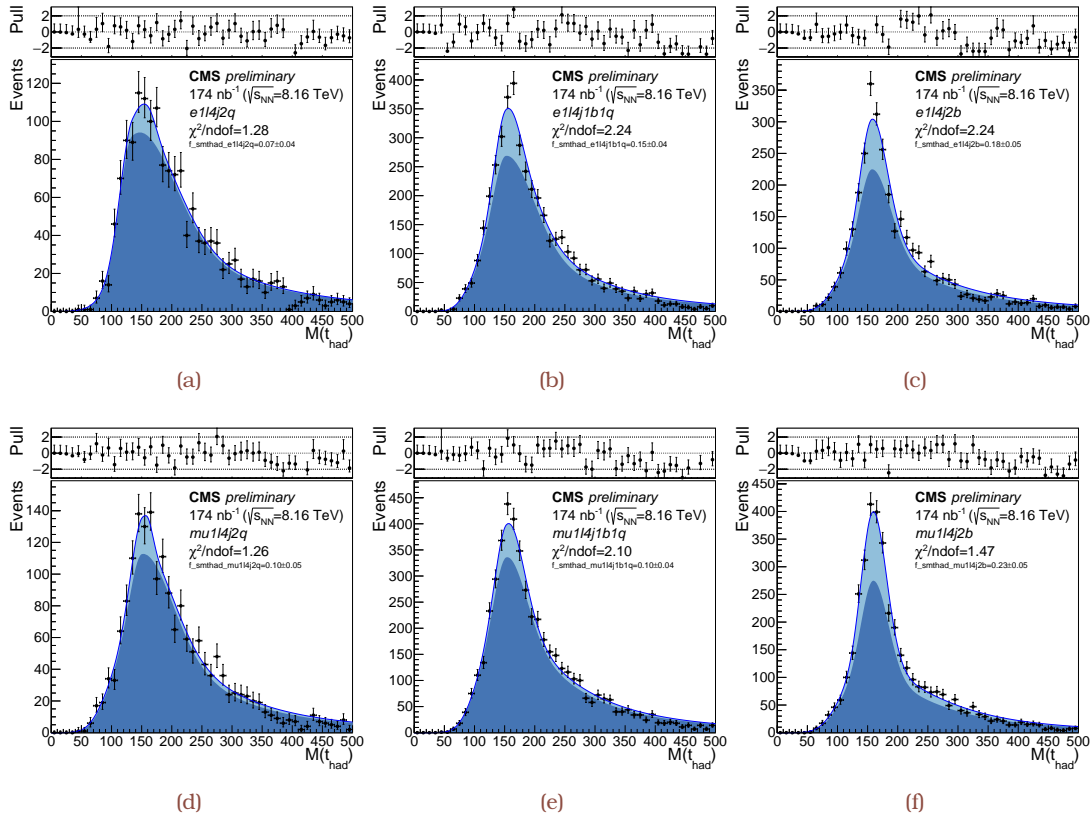
## 6.4.1 Signal parameterization



**Figure 6.8:** Parametrized signal spectra in  $e+jets$  (a-c) and  $\mu+jets$  (d-f) final states as a function of  $m_{jj}$  in the  $1l4j0b$  (a, d),  $1l4j1b$  (b, e), and  $1l4j2b$  (c, f) event categories, using the PYTHIA MC simulation (Table 6.1). On the bottom panels, the estimated contributions from correct (wrong) assignments are shown in light (dark) blue. The parameterizations of correct and wrong assignments are fitted to the MC simulated events (points) and their difference, divided by the statistical uncertainty in the simulation, is represented on the top panels [AN2].

For each mass variable, i.e.,  $m_{jj}$ ,  $m_{\text{had}}$  and  $m_{\text{lep}}$ , the  $\bar{t}t$  MC simulation is used to estimate the expected contributions from correctly or wrongly assigned pairs. Independently of the variable and event category it is seen that a Crystal-Ball function [283] summed with a Gamma function is able to model the spectrum of correct assigned pairs precisely, whereas wrongly assigned pairs are described with a bifurcated Gaussian, i.e., a Gaussian with different widths to the left and right of its mean summed with a Landau function. The definition of the pdfs and the routines used to adjust the parameters are based on RooFit [284].

Figures 6.8, 6.9 and 6.10 show the results of the parameterizations obtained for the different event categories used in the analysis. The separation between correct and wrong assigned pairs serves only illustrative purposes. Their ratio is determined by a fit to the signal MC, while its magnitude has no effect on the final result. The figures span a wide mass range mostly to highlight the expectations from the simulation for the signal. The fits to the data are performed within a smaller range to ensure their stability.



**Figure 6.9:** Parametrized signal spectra in  $e$ +jets (a–c) and  $\mu$ +jets (d–f) final states as a function of  $m_{t_{\text{had}}}$  in the  $1l4j0b$  (a, d),  $1l4j1b$  (b, e), and  $1l4j2b$  (c, f) event categories, using the PYTHIA MC simulation (Table 6.1). On the bottom panels, the estimated contributions from correct (wrong) assignments are shown in light (dark) blue. The parameterizations of correct and wrong assignments are fitted to the MC simulated events (points) and their difference, divided by the statistical uncertainty in the simulation, is represented on the top panels [AN2].

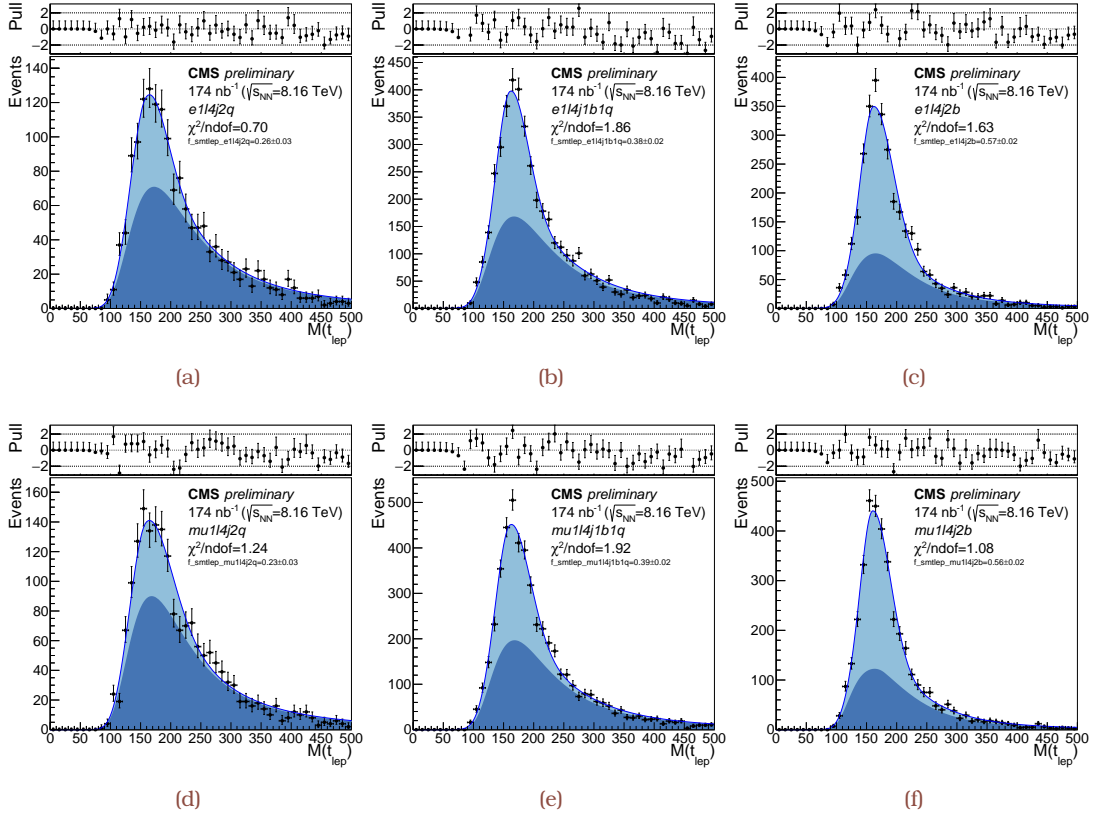
In Table 6.4 the main parameters for the spectrum of correct assignments are summarized. The effective width is computed from the cumulative distribution function (CDF); half of the 68% CL derived from the CDF is quoted as the effective width of the distribution. It can be seen that for the  $m_{j\bar{j}}$  and  $m_{t_{\text{had}}}$  variables the main parameters are quite similar across all event categories. The largest variations are related to the  $0b$  where the event count of the MC simulation may induce statistical fluctuations in the parameterizations. Given this category is signal depleted and is used only for controlling the background no significant impact in the analysis is expected.

## 6.4.2 Background parameterization

### W+jets background

The MC simulation is used to seed the parameters of the W+jets background modeling. Once a functional form is derived, its parameters and normalization are determined from the fit to the data. This approach is justified on the grounds of the nonaccurate simulation in the phase space of the analysis, i.e. simulations rely on the PYTHIA parton shower to





**Figure 6.10:** Parametrized signal spectra in  $e$ +jets (a–c) and  $\mu$ +jets (d–f) final states as a function of  $m_{t_{\text{lep}}}$  in the  $1l4j0b$  (a, d),  $1l4j1b$  (b, e), and  $1l4j2b$  (c, f) event categories, using the PYTHIA MC simulation (Table 6.1). In the bottom panels, the estimated contributions from correct (wrong) assignments are shown in light (dark) blue. The parameterizations of correct and wrong assignments are fitted to the MC simulated events (points) and their difference, divided by the statistical uncertainty in the simulation, is represented on the top panels [AN2].

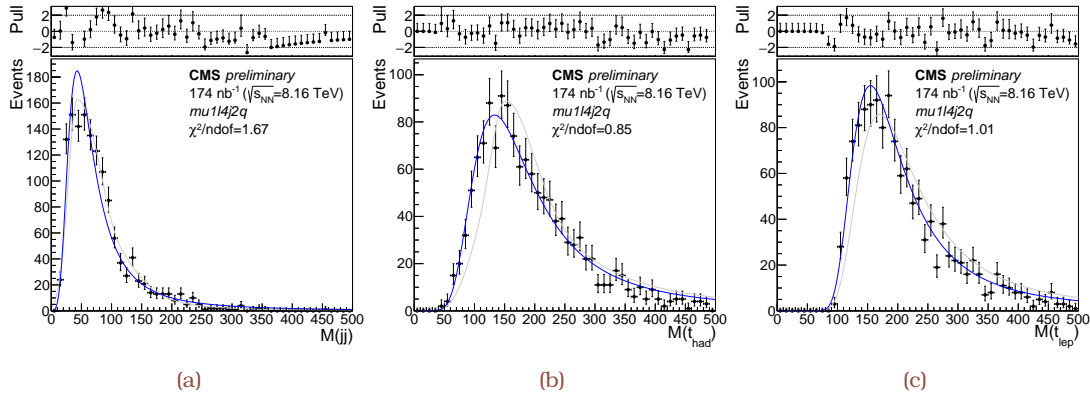
predict most of the jets in the  $t\bar{t}$  signal region. Therefore the simulation could bias both the rate and the shape of  $W$ +heavy flavor production.

From the MC simulation, a Landau pdf [284] is found to describe reasonably well the contribution from  $W$ +jets in all mass variables. Figure 6.11 compares the  $W$ +jets simulated events with the Landau pdf fit superimposed. Because of the low event count, the fit is performed using events with no  $b$ -tagged jets; a fair agreement is observed between the pdf and the simulation. It can also be noticed that the Landau parameters differ only slightly with respect to those obtained for the wrong assignment component of the signal, as described in the previous Section and shown in the same Figure. This is not surprising given the combinatorial nature of the  $W$ +jets background.

The higher event count  $W$ +jets LO sample using  $pp$  collisions at 8 TeV validates the Landau parameterization. Figure 6.12 shows the relative difference in the shape of the three mass distributions in the  $1l4j1b$  and  $1l4j2b$  with respect to the  $1l4j0b$  event category. The largest deviations are seen for  $m_{j'j''}$  and  $m_{t_{\text{lep}}}$  variables. The former indicates that the kinematics of the jets in a heavy-flavor enriched region are harder with respect to the light-flavor region, while the latter is expected as  $b$ -jets tend to contribute to real

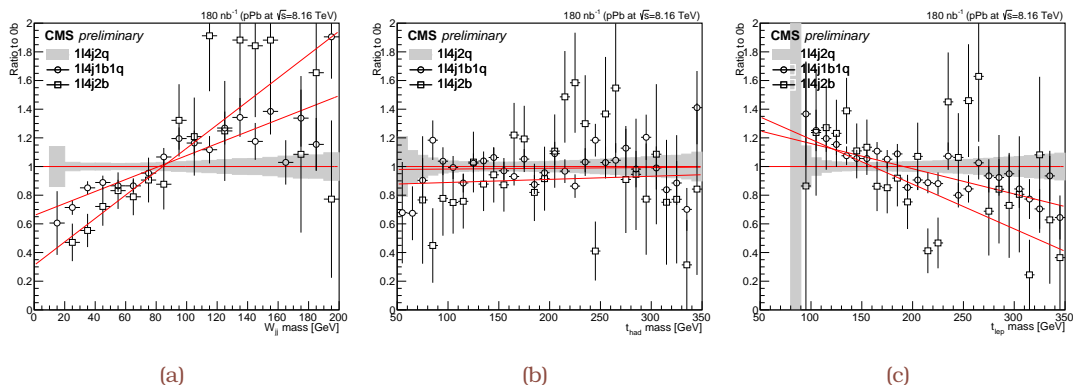
**Table 6.4:** Main parameters of the correct assignments for the signal spectra as extracted from the PYTHIA MC simulation (Table 6.1). Values are given in GeV [AN2].

Category	Final state ( $\ell$ +jets)	Mean	Median	Mode	Eff width
$m_{j\bar{j}}$					
1 $\ell$ 4j0b	$\ell = e$	75.2	73.7	75.0	28.9
	$\ell = \mu$	76.1	75.0	75.0	35.5
1 $\ell$ 4j1b	$\ell = e$	76.0	76.5	75.0	30.0
	$\ell = \mu$	75.9	76.6	75.0	30.4
1 $\ell$ 4j2b	$\ell = e$	76.2	77.3	75.0	26.0
	$\ell = \mu$	75.8	76.5	75.0	24.3
$m_{t_{\text{had}}}$					
1 $\ell$ 4j0b	$\ell = e$	156.8	164.9	159.7	44.0
	$\ell = \mu$	156.8	164.9	159.7	44.0
1 $\ell$ 4j1b	$\ell = e$	160.2	171.0	158.8	56.0
	$\ell = \mu$	162.3	172.0	161.0	53.9
1 $\ell$ 4j2b	$\ell = e$	161.0	169.3	159.9	40.8
	$\ell = \mu$	161.5	169.1	159.9	38.9
$m_{t_{\text{lep}}}$					
1 $\ell$ 4j0b	$\ell = e$	167.6	168.1	160.0	32.7
	$\ell = \mu$	172.8	172.5	160.0	35.0
1 $\ell$ 4j1b	$\ell = e$	168.9	170.8	162.6	30.0
	$\ell = \mu$	169.0	170.8	161.1	30.1
1 $\ell$ 4j2b	$\ell = e$	172.3	172.2	160.1	29.0
	$\ell = \mu$	187.7	170.2	179.4	27.3


**Figure 6.11:** Simulated and parametrized W+jets spectra (Table 6.1) in the  $\mu$ +jets final state as a function of  $m_{j\bar{j}}$  (a),  $m_{t_{\text{had}}}$  (b) and  $m_{t_{\text{lep}}}$  (c) in the 1 $\ell$ 4j0b event category; their difference, divided by the statistical uncertainty in the simulation, is represented on the top panels. The superimposed gray curve corresponds to the parametrized contribution from wrong assignments in the  $t\bar{t}$  signal process [AN2].

$p_{\text{T}}^{\text{miss}}$  through semileptonic b quark decays.

The relative differences in Fig. 6.12, as extracted using pp collisions, are used to



**Figure 6.12:** Relative difference of the W+jets spectra in the 1l4j1b and 1l4j2b with respect to the 1l4j0b event category from the W+jets LO sample, using pp collisions at 8 TeV. The ratios are shown for the  $m_{jj'}$  (a),  $m_{t_{\text{had}}}$  (b) and  $m_{t_{\text{lep}}}$  (c) variables. The gray band illustrates the statistical uncertainty in the MC sample using events in the 1l4j0b event category, while the red lines represent first order polynomial fits to the ratios [AN2].

extrapolate the W+jets distributions from the 1l4j0b to the 1l4j1b and 1l4j2b event categories; each simulated PYTHIA event falling in the 1l4j0b category is reweighted according to the value of the relative difference. The extrapolated distributions are fitted to Landau functions and are further compared to events from the PYTHIA MC simulation. The obtained results are shown in Fig. 6.13. Within the limited statistical accuracy of this procedure, the Landau parameterization is indeed expected to be a good approximation of the W+jets contribution in the different event categories. However, as summarized in Table 6.5, the parameters seem to differ in each category, and hence motivating separate width ( $\sigma$ ) and most probable value (MPV) Landau parameters per category. It is though interesting to note that the  $\sigma/\text{MPV}$  ratio is approximately the same possibly indicating that the intrinsic resolution is approximately the same. In the fit to the data Gaussian penalty terms are assigned to allow for a 5 (20)% variation in the  $\sigma$  (MPV) with respect to the predictions in Table 6.5.

### QCD multijet background

Typically, MC event simulation cannot properly reproduce the shape and normalization of the QCD multijet background, in particular, in the phase space of the analysis with an isolated and high- $p_T$  lepton on top of four jets. Significantly high event count samples would also be required, and without any guarantee that the result could be safely used as a description of this type of background, a data-based technique is applied.

A control region is selected in the  $\mu$ +jets final state by inverting the isolation requirement for muons, i.e.,  $I_{\text{rel}}^\mu > 0.2$ , leaving all other identification requirements unchanged. The inverted cut value of 0.2 reflects a tradeoff between the event count and the contamination expected in the control region from non-QCD processes, i.e., signal and W+jets events. In the e+jets final state the electron candidate is required to fail a looser set of identification criteria, the so-called “veto” identification (Table 3.2). A better estimate of the QCD multijet background is found in the ECAL endcaps [TH4] where the fraction of

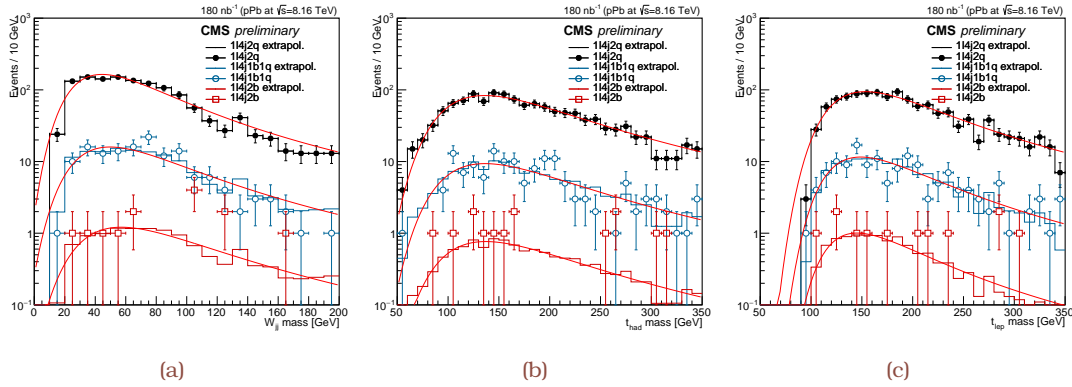


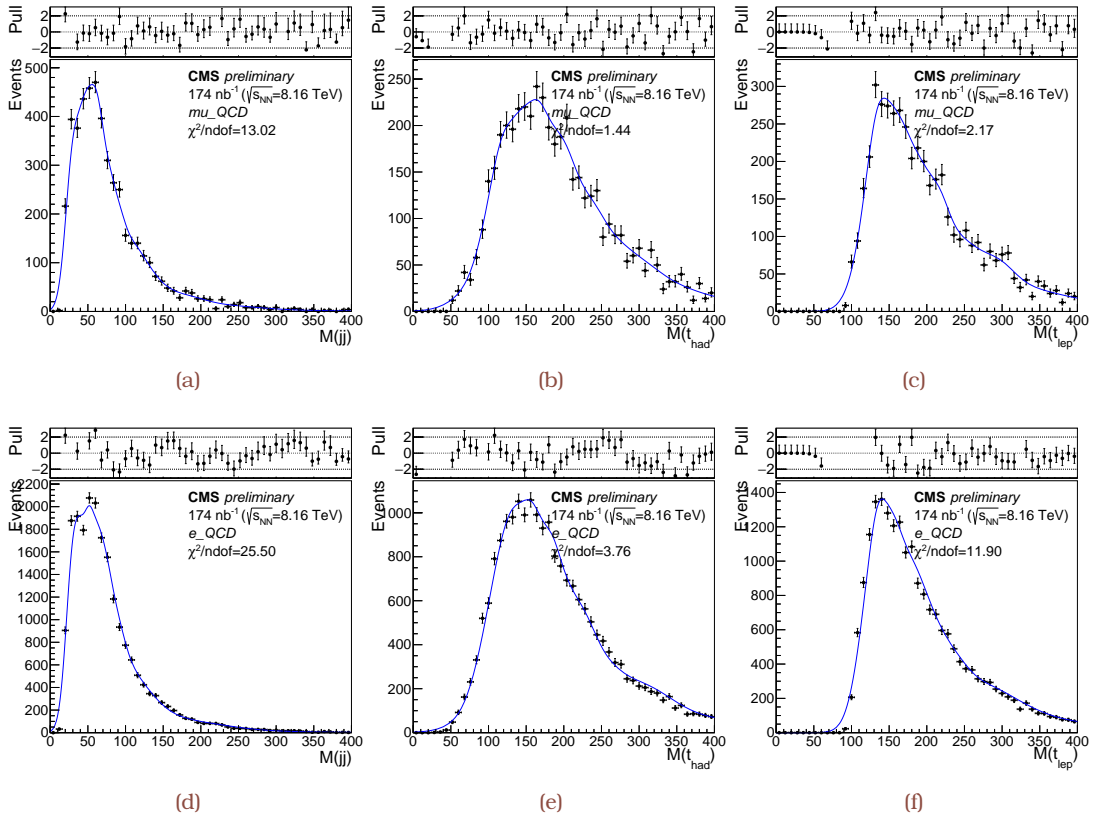
Figure 6.13: Reweighted W+jets spectra in the different event categories based on the ratio (Fig. 6.12) calculated from the W+jets LO sample, using pp collisions at 8 TeV separately for the  $m_{jj'}$  (a),  $m_{t_{\text{had}}}$  (b) and  $m_{t_{\text{lep}}}$  (c) variables. Landau functions are fitted to the reweighted spectra, while the latter are further compared to events from the PYTHIA MC simulation (Table 6.1), using pPb collisions [AN2].

Table 6.5: Parameters of the Landau functions as extracted from the reweighted W+jets PYTHIA MC simulation (Table 6.1). Values are given in GeV [AN2].

Category	MPV	$\sigma$	$\sigma/\text{MPV}$
$m_{jj'}$			
1l4j0b	48.2	17.0	0.35
1l4j1b	54.0	19.6	0.36
1l4j2b	62.0	22.3	0.36
$m_{t_{\text{had}}}$			
1l4j0b	143.7	34.2	0.24
1l4j1b	143.9	34.3	0.24
1l4j2b	144.7	34.6	0.24
$m_{t_{\text{lep}}}$			
1l4j0b	159.8	29.0	0.18
1l4j1b	154.7	26.5	0.17
1l4j2b	150.6	24.5	0.16

nongenuine electrons is higher than the barrel. In both  $\mu$ +jets and e+jets final states, the control regions contain no b-tagged jets to reduce the contamination from other non-QCD processes, and hence are expected to be enriched mostly in QCD multijet events.

Figure 6.14 shows the distributions of the three mass variables, reconstructed in the QCD multijet control regions. It is seen that the event count in the e+jets is significantly higher than in  $\mu$ +jets final state, owing to the dominant contribution from nongenuine electrons. Following the approach of [TH4] the shape of the QCD multijet background is assumed to be the same in the control and signal regions. The shape is modeled with a nonparametric kernel approach [285]. A fair agreement is observed for all mass variables and categories.



**Figure 6.14:** Parametrized QCD multijet distributions as a function of  $m_{jj}$  (a),  $m_{t_{\text{had}}}$  (b) and  $m_{t_{\text{lep}}}$  (c) using events with no b-tagged jets in (a–c)  $\mu$ +jets—failing  $I_{\text{Rel}}^{\mu} > 0.2$  requirement—and (d–f) e+jets—failing the veto identification—final states, respectively [AN2].

The initial normalization of the QCD multijet background is estimated from events with low  $p_{\text{T}}^{\text{miss}} < 20$  GeV. A fit is performed using the distributions of QCD multijet, and  $t\bar{t}$  and W+jets processes derived from the control regions and MC simulation, respectively. The contributions from  $t\bar{t}$  and W+jets are left to float within 30% of their initial estimates based on their theory cross sections. A 50–100% systematic uncertainty is assigned to the result of the fit that is displayed in Fig. 6.15 separately for the  $\mu$ +jets and e+jets final states. Although slight mismodeling is observed, the outcome of the fit serves the purpose of the data-based technique, i.e., to find an initial estimate for the number of QCD multijet events fulfilling the analysis selection. Table 6.6 summarizes the number of QCD multijet events estimated in each event category and the associated uncertainty which is used as a constraint when extracting the  $t\bar{t}$  cross section, as described in the following.

## 6.5 Systematic uncertainty

The measurement of the inclusive cross section is affected by systematic uncertainties that are assessed individually and originate from detector effects or theory assumptions. We describe in more detail the assessment of the systematic uncertainties below.

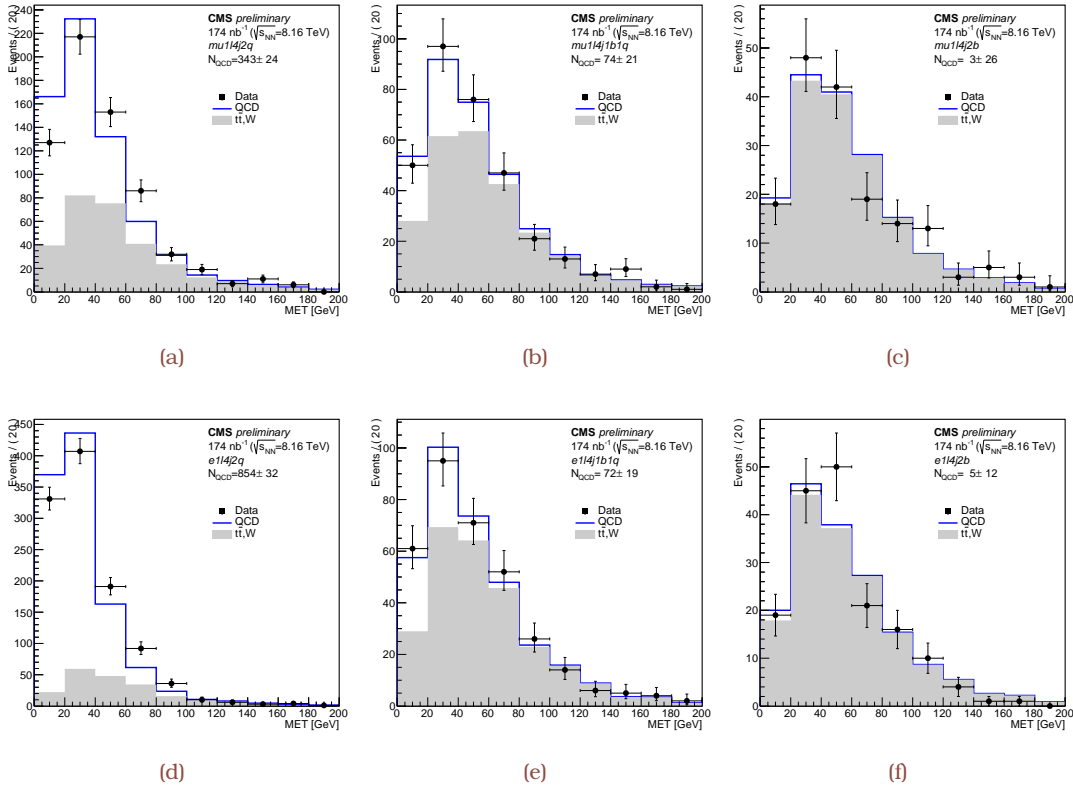


Figure 6.15: Fits to the  $p_T^{\text{miss}}$  distribution in the  $\mu$ +jets (a-c) and  $e$ +jets (d-f) final states in the 1l4j0b (a, d), 1l4j1b (b, e), and 1l4j2b (c, f) event categories. The QCD normalization estimated from the fit is shown in the captions, along with its statistical uncertainty [AN2].

Table 6.6: Initial estimate and associated uncertainty in the QCD multijet normalization in the 1l4j0b, 1l4j1b and 1l4j2b event categories [AN2].

Category	Final state ( $l$ +jets)	QCD normalization $\pm$ uncertainty
1l4j0b	$l = e$	854 $\pm$ 427
	$l = \mu$	343 $\pm$ 171
1l4j1b	$l = e$	74 $\pm$ 54
	$l = \mu$	77 $\pm$ 57
1l4j2b	$l = e$	5 $\pm$ 5
	$l = \mu$	3 $\pm$ 3

### 6.5.1 Experimental uncertainty

The uncertainties in the lepton trigger, reconstruction, and identification efficiencies are estimated with a tag-and-probe method (see Appendices B.0.2 and B.0.3). The total uncertainty is dominated by the dependency of charged lepton isolation on the underlying event.

The uncertainty due to the calibration of the jet energy scale has been estimated by a change of 5% in the jet energies, independent of the  $p_T$  and  $\eta$  of the jet. This takes into account a 3%-level difference between the reconstructed and generated jet energy

in MC events (Fig. 3.15a), and a 3% residual calibration uncertainty from data [198]. The uncertainty due to the limited accuracy of the jet energy resolution is determined by changing the JER correction from the nominal value within  $\pm 1$  standard deviation, depending on the  $p_T$  and  $\eta$  of the jet. We assume the 13 TeV values given that the JER degradation is mostly induced by detector effects. The parameterizations and associated uncertainties are described in Section 3.3.4 and Ref. [198], respectively. The effect on our analysis is estimated to be  $< 1\%$ .

Uncertainties in the background are included in the fit uncertainty, as the background normalization and distribution are determined *in-situ*. The b tagging efficiency is also considered as a separate source of experimental uncertainty in the measurement; is measured *in-situ* and bears the largest effect on the  $t\bar{t}$  cross section;

The uncertainty in the integrated luminosity has been initially estimated to be 5% at the time of the submission to the journal, while the central value remained essentially unchanged [TH2]. The 0.1% uncertainty in the LHC beam energy [259] has a numerically insignificant effect on this measurement.

## 6.5.2 Theory uncertainty

The modeling of the  $t\bar{t}$  signal events is an essential ingredient in the measurement as we rely on the expected acceptance and fraction of correct to wrong assignments. The impact of theoretical assumptions in the modeling is determined by repeating the analysis selection using the NLO POWHEG simulation. The differences found are small, and are accounted for as a systematic uncertainty.

The effect of the jet energy scale in the signal is included by varying the  $m_{j\bar{j}}$  distributions by a multiplicative factor as

$$\widetilde{m}_{j\bar{j}} = (1 + \delta_{JSF} \times \partial_{JSF}) \times m_{j\bar{j}} , \quad 6.5$$

where  $\delta_{JSF}$  is the estimated effect of the jet energy scale on the invariant mass of the two jets in the event, after propagating the jet energy scale uncertainty, and  $\partial_{JSF}$  is a nuisance parameter, Gaussian-distributed with mean and width of zero and unity, respectively.

The effect of the jet energy scale on the acceptance is estimated separately from its effect on the  $m_{j\bar{j}}$  distributions, by counting the events that pass the preselection when the jet energy scale is varied according to its uncertainties. This results in an additional 3% uncertainty in the acceptance that is not correlated in the fit with the JES impact on the  $m_{j\bar{j}}$  distribution since it pertains an extrapolation from the fiducial to the full phase space.

The uncertainty from the choice of free- and bound-proton PDF is determined by reweighting the sample of simulated  $t\bar{t}$  events according to the RMS of the CT14 and EPPS16 replicas. Two extra variations of  $\alpha_s(M_Z)$  are added in quadrature, to determine the total PDF uncertainty. Variation of the  $\mu_R/\mu_F$  scales in the  $t\bar{t}$  simulation is also taken into account.

## 6.6 Measurement of the inclusive $t\bar{t}$ cross section

### 6.6.1 Fit model

The parameterizations of the signal and main background processes, as described in Sections 6.4.1 and 6.4.2, respectively, are used to build a fit model that is applied to the data to extract the inclusive  $t\bar{t}$  cross section. The fit model predicts the total number of events in 1 $\ell$ 4j0b, 1 $\ell$ 4j1b, and 1 $\ell$ 4j2b event categories by summing up the contributions for  $t\bar{t}$ , and W+jets and QCD multijet background, i.e., in each category (“k”) we write

$$N_k(m_{j\bar{j}}) = N_k(t\bar{t}) \times \text{pdf}_k^{t\bar{t}}(m_{j\bar{j}}) + N_k(\text{bkg}) \times \left( \text{pdf}_k^{\text{W+jets}}(m_{j\bar{j}}) + f_k \times \text{pdf}_k^{\text{QCD}}(m_{j\bar{j}}) \right), \quad 6.6$$

where  $N_k(i)$  ( $\text{pdf}_k(i)$ ) is the total number of events (distribution) for the process  $i$  in category  $c$ , and  $f_k \in [0, 1]$  is a parameter that regulates the relative contribution in background events from the QCD multijet background;  $N_k(\text{QCD}) = f_k N_k(\text{bkg})$  is constrained by a Gaussian penalty term with mean and width as shown in Table 6.6. Given that the resolution and discrimination between the signal and background distributions of  $m_{j\bar{j}}$  are superior with respect to the other mass variables, the fit used to extract the final cross section is performed with the  $m_{j\bar{j}}$  variable. A variant of the nominal configuration included a two-dimensional fit to  $m_{j\bar{j}}$  and  $m_{\text{had}}$  variables; the  $\sigma_{t\bar{t}}$  was found to be consistent with the one obtained from the  $m_{j\bar{j}}$  fit, with the latter being described in the following.

The number of  $t\bar{t}$  events in each category can be inferred based on multinomial probabilities to describe the expected number of signal events. Denoting by  $\varepsilon_b$  the b-finding efficiency, i.e., the probability to select and tag a b quark from a top quark decay, one can write the following set of equations for the number of signal events expected in each event category

$$N_{1\ell 4j0b}(t\bar{t}) = \varepsilon_b^2 N(t\bar{t}) \quad N_{1\ell 4j1b}(t\bar{t}) = 2\varepsilon_b(1 - \varepsilon_b)N(t\bar{t}) \quad N_{1\ell 4j2b}(t\bar{t}) = (1 - \varepsilon_b)^2 N(t\bar{t}). \quad 6.7$$

The parameter  $\varepsilon_b$  is assigned 10% pre-fit uncertainty by means of a Gaussian constraint. From simulation it is expected that  $\varepsilon_b \approx 0.6$ , a value which is consistent between the two final states, and PYTHIA and POWHEG MC samples. Imposing the relation in Eq. (6.7) establishes the signal more accurately with respect to a fit where the signal yields would be left to float without any constraint in each category. Owing to the similarity of the distributions for wrongly assigned  $t\bar{t}$  events and the W+jets background, in particular, in events with no b-tagged jets, the fit could be biased without the more refined signal model of Eq. (6.7).

### 6.6.2 Extraction of the inclusive $t\bar{t}$ cross section

The final ingredient is the cross section which is used to parameterize  $N(t\bar{t})$

$$N(t\bar{t}) = \sigma_{t\bar{t}} A \varepsilon \mathcal{L}, \quad 6.8$$



where  $A$  is the signal acceptance,  $\varepsilon$  is the total lepton trigger and selection efficiency, and  $\mathcal{L}$  is the total integrated luminosity. In the fit, these parameters float constrained by Gaussian penalty terms with mean and width corresponding to their estimated values and uncertainties, respectively.

In total 25 nuisance parameters are used: 12 background normalization-related parameters, six related to the distribution of the W+jets background (MPV and  $\sigma$  of the Landau pdf), the b-finding efficiency, the JSF scaling factor, the integrated luminosity, and acceptance and efficiency values per final state. Out of the total nuisance parameters, 19 are associated with Gaussian constraints in the fit, i.e., only the total background yields are unconstrained. As these parameters are not directly relevant to our measurement and reflect unknown degrees in the modeling of signal and background, a profiling of the likelihood is performed in order to evaluate the uncertainty in the signal yield.

The likelihood is written as

$$\mathcal{L}(\sigma_{t\bar{t}}, \vec{\Theta}) = \prod_k \mathcal{P}(N_k | \hat{N}_k(\sigma_{t\bar{t}}, \vartheta_i)) \prod_i \rho(\vartheta_i^0, \vartheta_i, \sigma_{\vartheta_i}), \quad (6.9)$$

where  $\vec{\Theta}$  is the vector of nuisance parameters,  $\vartheta_i^0$  and  $\sigma_{\vartheta_i}$  are the mean and width of the Gaussian penalty term associated to the  $i$ th nuisance parameter (when applicable), and the rest of the notations follow Eq. (5.11). The constraint terms help to stabilize the fit, penalizing extreme values of the nuisance parameters that may be induced by the low event count in the data and the similarity for part of the signal with the background. The central value is thus obtained by maximizing Eq. (6.9), and the associated uncertainty is determined after profiling the likelihood, i.e., by maximizing the test statistic

$$\hat{n} = \frac{\mathcal{L}(\sigma_{t\bar{t}}, \hat{\vec{\Theta}})}{\mathcal{L}(\hat{\sigma}_{t\bar{t}}, \hat{\vec{\Theta}})}, \quad (6.10)$$

where the notations follow Eq. (5.12).

Figure 6.16 shows the  $m_{j\bar{j}}$  distribution for events with zero, one, or at least two b-tagged jets, compared with the combined fit results. An excellent agreement is observed in all event categories between the data and the fitted signal and background parameterizations. The total number of signal events obtained through the combined fit to the  $\mu$ +jets and e+jets final states is  $710 \pm 130$ . The total uncertainty in the fitted signal events is obtained from the covariance matrix of the fit, and using Eq. (6.8)

$$\sigma_{t\bar{t}} = 45 \pm 8 \text{ (total) nb.} \quad (6.11)$$

The signal acceptance is determined from the POWHEG NLO simulation (Table 6.1) and is found to be  $\mathcal{A} = 0.060 \pm 0.002$  and  $0.056 \pm 0.002$  in the  $\mu$ +jets and e+jets final states relative to all generated  $t\bar{t}$  events, including the branching fraction to leptons; the 4% uncertainty in the acceptance correction includes its dependence on the proton and Pb PDFs, and on the values of theoretical scales and the QCD coupling ( $a_s = 0.118 \pm 0.001$  at the Z boson pole mass). The lepton trigger and selection efficiencies are  $\varepsilon = 0.91 \pm 0.04$

and  $0.63 \pm 0.03$  in the  $\mu$ +jets and e+jets as estimated from data.

Fitting exclusively the  $\mu$ +jets or e+jets final state, the number of signal events is measured to be  $420 \pm 80$  or  $348 \pm 50$ , and using Eq. (6.8)

$$\sigma_{\bar{t}t}^{\mu+\text{jets}} = 44 \pm 3 \text{ (stat)} \pm 8 \text{ (syst)} \text{ nb}, \quad 6.12$$

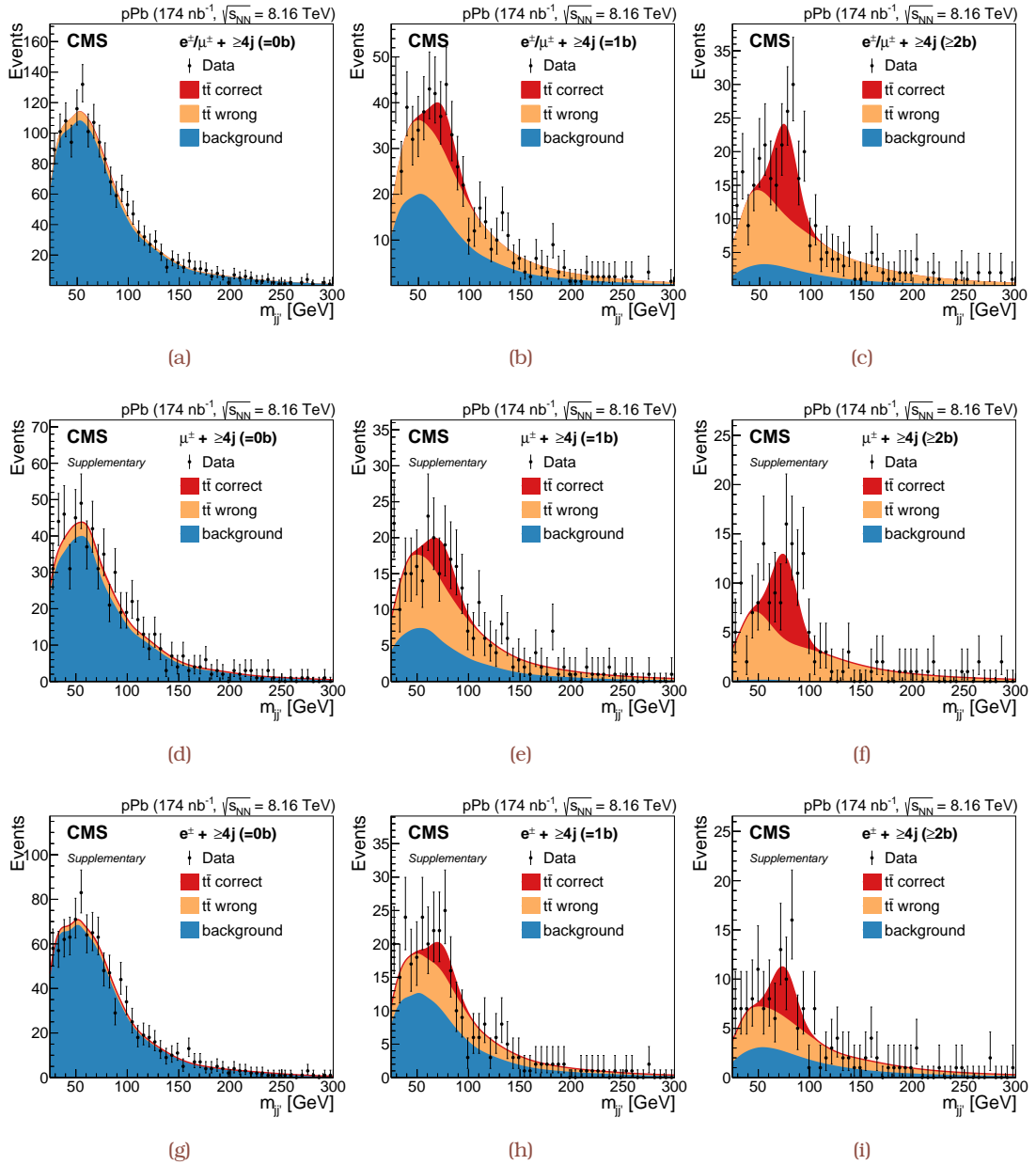
$$\sigma_{\bar{t}t}^{e+\text{jets}} = 56 \pm 4 \text{ (stat)} \pm 13 \text{ (syst)} \text{ nb}, \quad 6.13$$

where the uncertainty is further split into a statistical part, by leaving  $\sigma_{\bar{t}t}$  to float in the fit and fixing all other parameters to their post-fit values, and a systematic part, by subtracting the square of the statistical uncertainty from the square of the total uncertainty. Figure 6.16 shows the  $m_{jj'}$  distribution for events with zero, one, or at least two b-tagged jets, compared with the exclusive fit results. The cross section is thus measured with a relative uncertainty of about 18 (23)% in the  $\mu$ +jets (e+jets) final state, and 17% in their combination.

The measured cross section is found to be consistent with the theoretical prediction [279],  $\sigma(\text{pPb} \rightarrow \bar{t}t + X) = 59.0 \pm 5.3 \text{ (PDF)}_{-2.1}^{+1.6} \text{ (scale)} \text{ nb}$ , computed with MCFM (v.8) [256] using the CT14 proton PDF [33] and the EPPS16 nPDF for the lead ions [88], scaled to NNLO+NNLL accuracy with a  $K$  factor computed with Top++ (v.2.0) [257], and multiplied by  $A = 208$ . The PDF uncertainties are obtained from the corresponding 56+40 eigenvalues of the CT14+EPPS16 sets (corresponding to a 90% confidence level) added in quadrature, while the theoretical scale uncertainty is estimated by modifying the factorization and renormalization scales within a factor of two with respect to their default value set at  $\mu_F = \mu_R = m_t$ . The same calculation with the CT10 proton PDF [286] and EPS09 [86] nPDF yields  $\sigma(\text{pPb} \rightarrow \bar{t}t + X) = 57.5 \pm_{-3.3}^{+4.3} \text{ (PDF)}_{-2.0}^{+1.5} \text{ (scale)} \text{ nb}$ . The difference in the theoretical  $\bar{t}t$  cross section computed with the PDF for free protons and for bound nucleons is small. A net overall antishadowing effect increases the total top-quark pair cross section by only 4% for both the EPPS16 and EPS09 sets in pPb relative to pp collisions [279]. Such a difference is too small to be observed in the data with the current experimental uncertainties. Figure 6.17 shows the measured and theoretical cross sections for  $\bar{t}t$  production in pPb collisions at  $\sqrt{s_{\text{NN}}} = 8.16 \text{ TeV}$ , compared with the results from pp collisions at  $\sqrt{s} = 8 \text{ TeV}$  [246, 247] scaled by  $A$  and by the ratio of 8.16 over 8 TeV NNLO+NNLL cross sections.

### 6.6.3 Impact of the statistical and systematic uncertainties

The values obtained for the free parameters of the fit, i.e., the normalization of the signal, QCD multijet, and W+jets yields (as well as the parameters of their functional forms described above), the b-finding efficiency, and the jet energy scale factor, are reported in Table 6.7. It can be seen that the post-fit values for the b-finding and jet energy scale factor are close to their initial estimates, and are not significantly altered in the fit. Because of the low event count, the sensitivity of the fit to the QCD multijet and W+jets background distributions is reduced, hence the fraction  $f$  of QCD multijet background is prone to statistical fluctuations. In the signal enriched categories, there is even lower sensitivity



**Figure 6.16:** Invariant mass distributions of the W candidate,  $m_{j\bar{j}}$ , in the 0 (a, d, g), 1 (b, e, h), and 2 (c, f, i) b-tagged jet categories after all selections in the  $l+jets$  (a-c),  $\mu+jets$  (d-f) and  $e+jets$  (g-i) final states. The red and orange areas correspond to the signal simulation (correct and wrong assignments, respectively) while the blue one corresponds to the estimated non-top background contributions. The error bars indicate the statistical uncertainty [TH1].

to the background parameters that is further reflected on the normalization, functional form, and relative contribution parameters. There is a fair agreement between the values obtained in the exclusive fits to the in the  $\mu+jets$  (a-c) and  $e+jets$  (d-f) final states, and obtained in the combined fit to all six event categories.

Table 6.8 summarizes the statistical and systematic uncertainties in the measured  $t\bar{t}$  production cross section from different sources (see Section 6.5). The methodology follows closely the one used in [TH4]. For each source of uncertainty, the corresponding nuisance

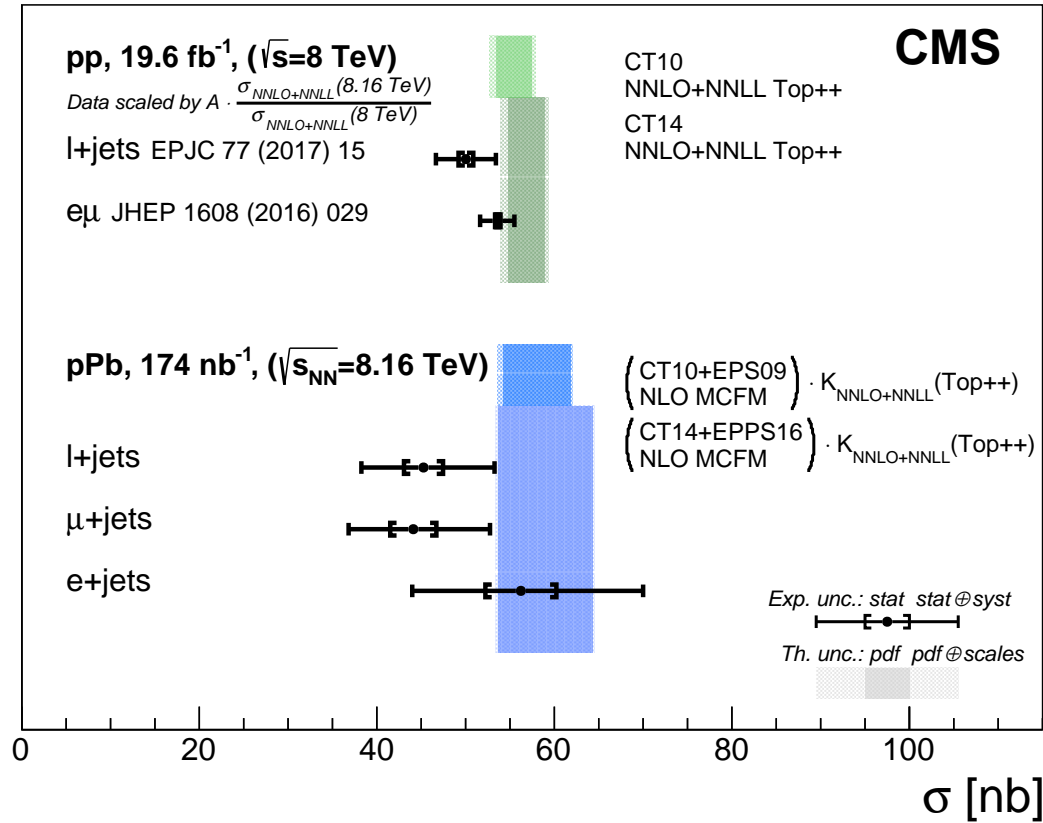


Figure 6.17: Inclusive  $t\bar{t}$  cross section measurements in the  $\mu$ +jets,  $e$ +jets, and combined  $\ell$ +jets final states using pPb collisions at  $\sqrt{s_{\text{NN}}} = 8.16$  TeV, compared to theoretical NNLO+NNLL predictions, and to scaled  $\sqrt{s} = 8$  TeV pp results [246, 247]. The total experimental error bars (theoretical error bands) include statistical and systematic (PDF and scale) uncertainties added in quadrature [TH1].

parameter in the fit is fixed to its post-fit value and varied within  $\pm 1$  standard deviation. The fit is repeated, and the difference found in the measured  $\sigma_{t\bar{t}}$  with respect to initial estimate is the reported impact. The individual sources of experimental uncertainty can only be interpreted as the observed post-fit values, and not as an absolute, orthogonalized breakdown of the uncertainty. The dominant source of systematic uncertainty originates from the  $b$ -finding efficiency, followed by the background normalization and modeling, and then the jet energy scale. The statistical uncertainty is at the level of 5%. It is expected that most of the systematic uncertainty in the analysis to improve with higher event count data samples. The uncertainties in the integrated luminosity, though refined as described in Section 4.5, and signal acceptance are currently the limiting factors.

#### 6.6.4 Statistical significance of the signal

Several studies have been performed to quantify the statistical significance of the observed signal compared to a background-only hypothesis. An asymptotic approach is used with a test statistic based on the profile likelihood ratio, and systematic uncertainties are included in the form of nuisance parameters given Gaussian priors [287]. Several tests have been performed, varying the estimation method and the background model-

**Table 6.7:** Post-fit results for the free parameters of the exclusive and combined fits used to extract the inclusive  $\sigma_{t\bar{t}}$  in the  $\mu$ +jets and e+jets, and combined  $l$ +jets final states, respectively, using pPb collisions at  $\sqrt{s_{NN}} = 8.16$  TeV [AN2].

Parameter	$\mu$ +jets	Fit e+jets	Combined	
$\varepsilon_b$	$0.62 \pm 0.05$	$0.59 \pm 0.05$	$0.62 \pm 0.04$	
JSF	$0.98 \pm 0.02$	$1.03 \pm 0.02$	$0.99 \pm 0.02$	
	Category	Background yields		
$N_k(\text{bkg})$	1e4j0b	$[1.04 \pm 0.04] \times 10^3$	$[1.03 \pm 0.04] \times 10^3$	
	1e4j1b	$[0.17 \pm 0.04] \times 10^3$	$200 \pm 30$	
	1e4j2b	$43 \pm 15$	$35 \pm 12$	
	1 $\mu$ 4j0b	$596 \pm 34$	$594 \pm 34$	
	1 $\mu$ 4j1b	$[0.13 \pm 0.04] \times 10^3$	$[0.12 \pm 0.04] \times 10^3$	
	1 $\mu$ 4j2b	$2.7 \pm 2.9$	$2.9 \pm 3.1$	
$f_k$	1e4j0b	$1.00 \pm 0.03$	$1.00 \pm 0.030$	
	1e4j1b	$0.72 \pm 0.20$	$0.66 \pm 0.17$	
	1e4j2b	$0.12 \pm 0.11$	$0.10 \pm 0.08$	
	1 $\mu$ 4j0b	$1.00 \pm 0.10$	$1.00 \pm 0.10$	
	1 $\mu$ 4j1b	$0.60 \pm 0.27$	$0.53 \pm 0.29$	
	1 $\mu$ 4j2b	$1.00 \pm 0.60$	$1.00 \pm 0.60$	
	W+jets Landau pdf			
MPV	1l4j0b	$48 \pm 10$	$48 \pm 9$	$48 \pm 10$
	1l4j1b	$57 \pm 7$	$53 \pm 6$	$54 \pm 5$
	1l4j2b	$62 \pm 12$	$62 \pm 12$	$60 \pm 6$
$\sigma$	1l4j0b	$17.0 \pm 0.8$	$17.0 \pm 0.8$	$17.0 \pm 0.9$
	1l4j1b	$19.5 \pm 1.0$	$19.3 \pm 1.0$	$19.1 \pm 1.0$
	1l4j2b	$22.3 \pm 1.1$	$22.2 \pm 1.1$	$22.0 \pm 1.1$

ing assumptions. Even with the most conservative assumptions, the background-only hypothesis is excluded with a significance above five standard deviations.

### Significance with the nominal shape analysis and variations

The fitted  $\sigma_{t\bar{t}}$  and the statistical significance are reported in Table 6.9. Results are given using the nominal  $m_{jj'}$  analysis and for different assumptions related to the W+jets background parameters and event categories. For all considered cases, a significance larger than five standard deviations is obtained.

The PLR variation as a function of the parameter of interest, and corresponding to the different hypotheses about the W+jets model parameters, are given in Figs. 6.18- 6.20 for all three event categories. The breakdown into statistical and systematic uncertainties is slightly different from the procedure followed to quote the result in Eq. (6.13); the background normalization is included in the statistical uncertainty, while they are treated as a source of systematic uncertainty in Eq. (6.13)x.

**Table 6.8:** Summary of the individual contributions to the statistic and systematic uncertainties in the inclusive  $\sigma_{\bar{t}t}$  measurement using pPb collisions at  $\sqrt{s_{\text{NN}}} = 8.16$  TeV. The contributions are estimated by the change induced in the cross section after varying once at a time the corresponding nuisance at  $\pm 1$  standard deviation away from its post-fit values. The impact of the individual sources of uncertainty is given as a relative uncertainty and quoted with one significant figure. The total uncertainty has been symmetrized [TH1].

Source	$\Delta\sigma_{\bar{t}t}/\sigma_{\bar{t}t}$ (%)		
	e+jets	$\mu$ +jets	Combined
Statistical	$\pm 7$	$\pm 6$	$\pm 5$
Systematic			
Electron efficiency	4	-	4
Muon efficiency	-	4	4
b-finding efficiency	8	15	13
Jet energy scale	4	3	4
Jet energy resolution	<1	<1	<1
Background	21	8	7
Signal acceptance	4	4	4
Integrated luminosity	5	5	5
Total uncertainty	23	18	17

**Table 6.9:** Statistical significance obtained from the  $m_{j\bar{j}}$  nominal analysis using events from all three categories, and exclusively in the  $1\ell 4j2b$  category. Different hypotheses about the W+jets background distribution are considered: nonrestricted background model parameters, or parameters fixed to the central values as summarized in Table 6.5. The quoted uncertainty in the cross section is the total uncertainty [AN2].

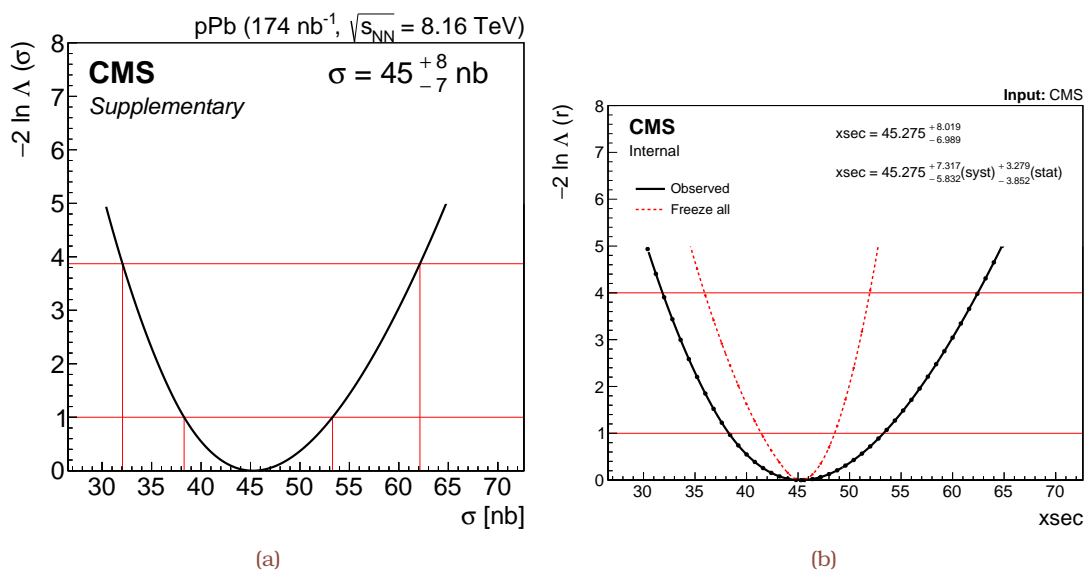
Hypothesis	All event categories		$1\ell 4j2b$	
	$\sigma_{\bar{t}t}$ [nb]	Significance	$\sigma_{\bar{t}t}$ [nb]	Significance
Nominal	$45 \pm 8$	6.64	$49 \pm 13$	5.95
Nonrestricted W+jets distribution	$48 \pm 6$	5.88	$44 \pm 12$	4.74
Fixed W+jets distribution	$49 \pm 7$	8.06	$50 \pm 13$	9.94

### Significance using pseudo-experiments

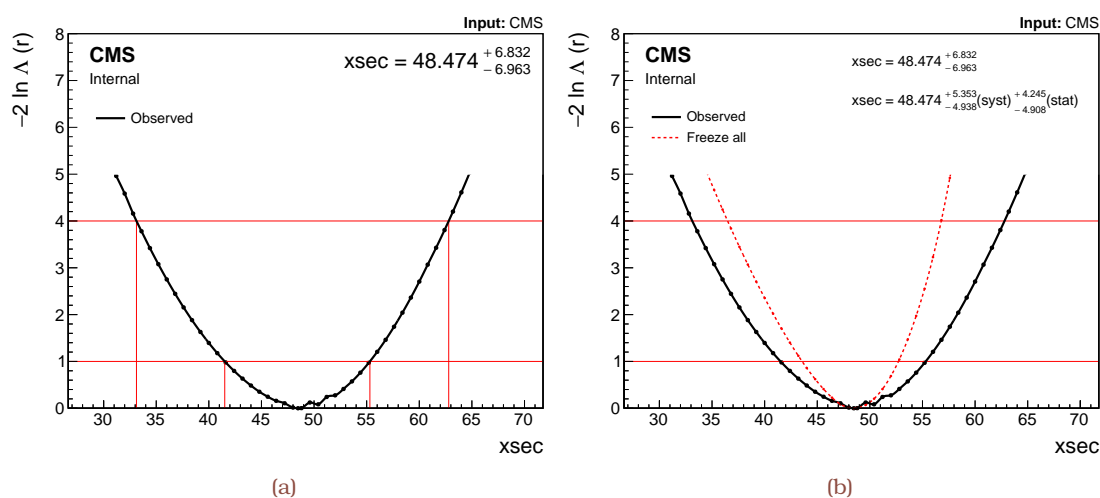
The significance is also estimated by generating pseudo-experiments from the background-only hypothesis, and monitoring the distribution of the test statistic. For this test, the nominal hypothesis is used, i.e., the W+jets background pdf parameters to be constrained. The resulted distribution is shown in Fig. 6.21. It can be seen that the statistical significance in data is above five standard deviations.

### Significance using the $1\ell 4j2b$ event category with a counting experiment

The significance has also been estimated using a simple counting experiment in the  $1\ell 4j2b$  event category, i.e., combining the  $\mu$ +jets and e+jets final states. This means that



**Figure 6.18:** Profile likelihood ratio variation as a function of the parameter of interest for the nominal  $m_{j\bar{j}}$  analysis corresponding to the (a) total [TH1] and (b) split into statistical and systematic uncertainties [AN2]. The two horizontal lines on (a) represent the values in the profile likelihood that are used to determine the 68 and 95% confidence level intervals for the  $\sigma_{t\bar{t}}$ .



**Figure 6.19:** Profile likelihood ratio variation as a function of the parameter of interest considering nonrestricted W+jets distribution, and corresponding to the (a) total and (b) split into statistical and systematic uncertainties [AN2]. The two horizontal lines on (a) represent the values in the profile likelihood that are used to determine the 68 and 95% confidence level intervals for the  $\sigma_{t\bar{t}}$ .

no information related to the distributions is considered. Different hypotheses on the systematic uncertainty are considered, as summarized in Table 6.10: the nominal hypothesis, including only the systematic uncertainty considered for the  $\sigma_{t\bar{t}}$  measurement, and extra assumptions with an additional arbitrary systematic uncertainty in the back-

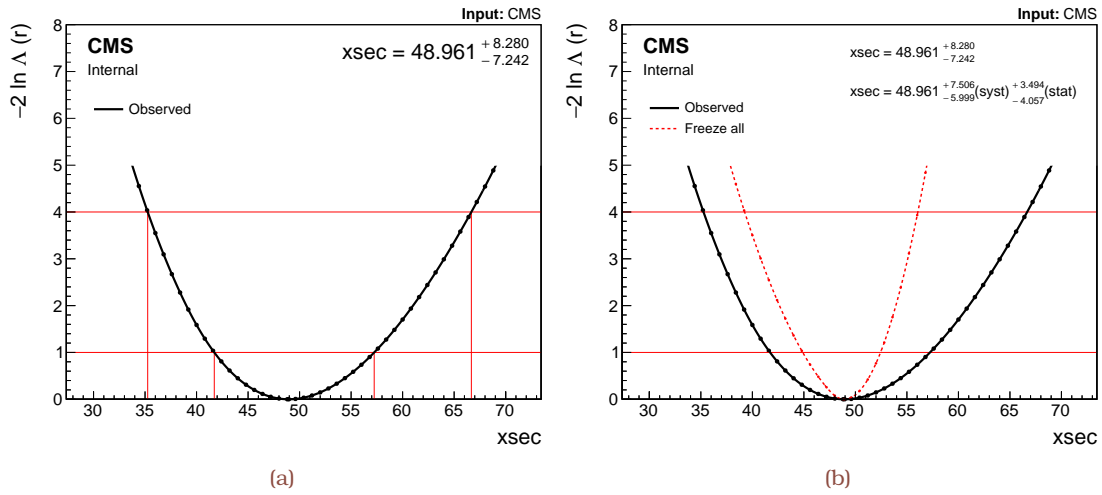


Figure 6.20: Profile likelihood ratio variation as a function of the parameter of interest considering nonrestricted W+jets distribution, and corresponding to the (a) total and (b) split into statistical and systematic uncertainties [AN2]. The two horizontal lines on (a) represent the values in the profile likelihood that are used to determine the 68 and 95% confidence level intervals for the  $\sigma_{\bar{t}t}$ .

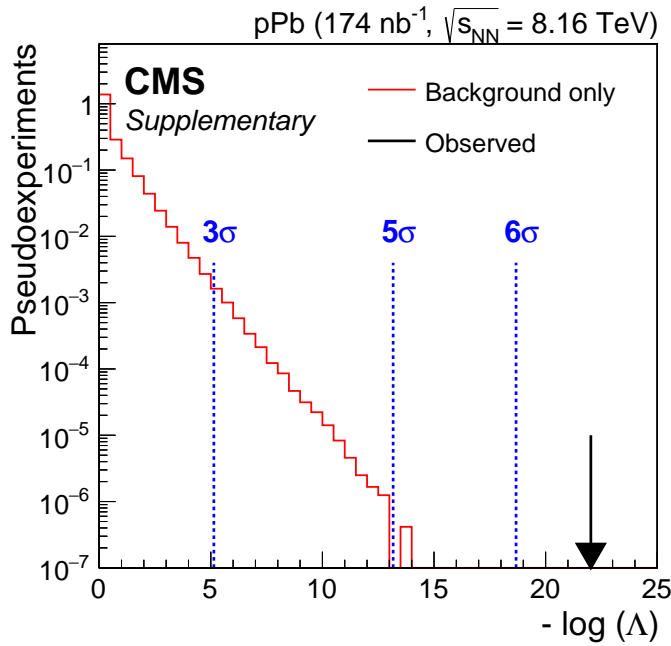


Figure 6.21: Profile likelihood ratio distribution from pseudo-data generated from the nominal background-only hypothesis. The value of the test statistic in data is also indicated by the black arrow. Dashed lines show the approximate location of the expected test statistic value, based on asymptotic formulae, for  $z$  values equal to three, five, and six standard deviations [TH1].

ground normalization. This is to account for the fact that the background normalization is left free in the actual fit to the  $m_{j\bar{j}}$  distribution, but it is constrained by the shape of the



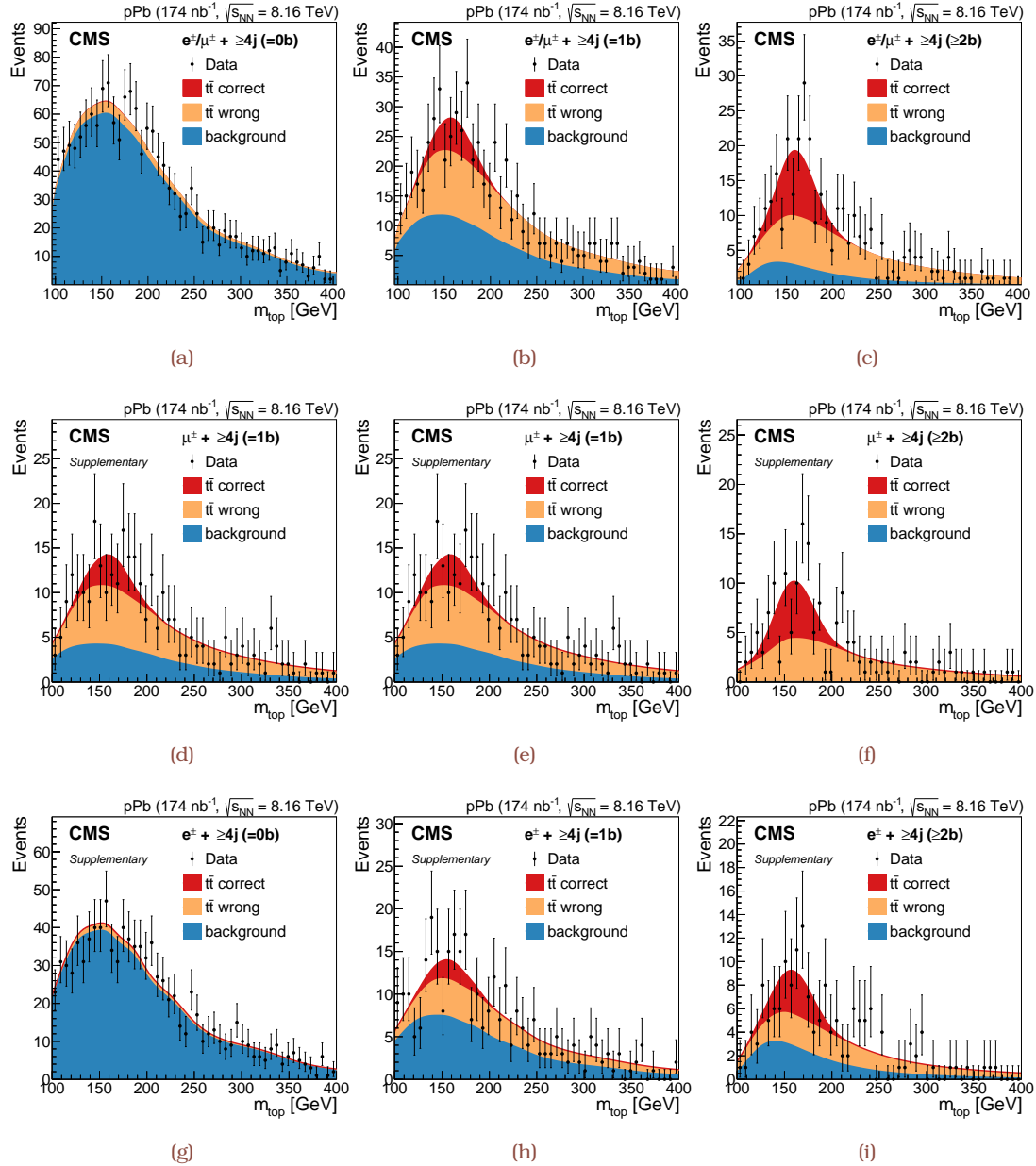
distribution in this fit. Even with the most conservative assumption, the background-only hypothesis is excluded with a significance above five standard deviations. The nominal hypothesis gives a best-fit cross section of  $51.2 \pm 9.0$  nb.

**Table 6.10:** Significance obtained with a simple counting experiment using only events from the  $1\ell 4j 2b$  category. Different assumptions about the systematic uncertainty are considered: the nominal hypothesis, including only the systematic uncertainty considered for the cross section measurement, and hypotheses with an additional arbitrary systematic uncertainty in the background normalization [AN2].

Hypothesis	Statistical significance
Nominal	16.8
Additional background 10%	14.8
Additional background 20%	12.2
Additional background 50%	8.5
Additional background 100%	5.7

#### **Additional post-fit control distributions**

To further support the hypothesis that the selected data are consistent with the production of top quarks the hadronic top quark mass is formed, following the description of Section 6.3. In the  $1\ell 4j 0b$  and  $1\ell 4j 1b$  event categories, the jet(s) with the highest value(s) of the b quark identification discriminator are considered for this purpose. Figure 6.22 shows the distribution of the hadronic top mass reconstructed using events in the three b-tagged jet categories, with all signal and background parameters kept fixed to those from the outcome of the  $m_{tj'}$  fit.



**Figure 6.22:** Invariant mass distributions of the  $t \rightarrow jj'b$  candidates,  $m_{\text{top}}$ , in the 0 (a, d, g), 1 (b, e, h), and 2 (c, f, i) b-tagged jet categories after all selections in the  $l$ +jets (a-c),  $\mu$ +jets (d-f) and  $e$ +jets (g-i) final states. All signal and background parameters are kept fixed to the outcome of the  $m_{jj'}$  fit. The red and orange areas correspond to the signal simulation (correct and wrong assignments, respectively) while the blue one corresponds to the estimated non-top background contributions. The error bars indicate the statistical uncertainty [TH1].



IV



Part D



## Chapter 7: Projected improvements in the precision of the inclusive $\sigma_{t\bar{t}}$ measurement

### Contents

---

<b>7.1 The largest reference proton-proton data set at 5.02 TeV</b> . . . . .	<b>210</b>
<b>7.2 Incremental improvements in luminosity determination</b> . . . . .	<b>210</b>
7.2.1 Refined impact of nonfactorizable beam shape . . . . .	210
7.2.2 Concept and formalism of variable separation scan . . . . .	212
<b>7.3 Cross section measurements</b> . . . . .	<b>214</b>
7.3.1 Statistical and systematic uncertainty treatment . . . . .	214
7.3.2 PDF constraints including the inclusive $\sigma_{t\bar{t}}$ . . . . .	214

---

The reference proton-proton run in 2017 consisted of collisions at  $\sqrt{s} = 5.02$  TeV, the same nucleon-nucleon center-of-mass energy to the heavy ion collisions in 2018, i.e., the last run before the second long shutdown. The precision of the inclusive  $t\bar{t}$  cross section will profit from the larger amount of data (about  $0.3\text{fb}^{-1}$ ) and the phase-I silicon pixel tracker. Owing to refined transverse beam shape models, the systematic uncertainty associated with the knowledge of the absolute luminosity scale is expected to be further decreased. This has been already demonstrated by a specially conducted run at the end of 2017, in which the beams were adjusted for low pileup configuration. In this Chapter, a projection of the  $\sigma_{t\bar{t}}$  measurement is presented based on the result previously attained in 2015, and the expected improvements are used to estimate the tighter constraints on parton distribution functions.

The material in the following Chapter, as documented in [Scientific output](#) and [Internal notes](#), relies almost exclusively on a original contribution. The latter further includes the preliminary effort of luminosity calibration for the pp run at  $\sqrt{s} = 5.02$  TeV in 2017.

## 7.1 The largest reference proton-proton data set at 5.02 TeV

The second proton-proton reference data set in Run 2 offers the unprecedented opportunity to establish a precise baseline evaluation relative to heavy ion data, and to parasitically improve and extend top quark cross section measurements at  $\sqrt{s} = 5.02$  TeV. The higher amount of data relative to 2015 (Figs. 7.1a and 7.1b) and the inclusion of the phase-I pixel silicon tracker allow even for fine-binned measurements in phase space regions—especially at the edge of the tracker acceptance—that are affected in the current  $\sigma_{\bar{t}t}$  measurement, e.g., the  $|\eta| > 2.1$  region (Figs. 7.1c and 7.1d) because of the QCD multijet background. The most significant reduction of uncertainty is thus expected because of the higher event count in data and MC simulation, and reduced uncertainty in the absolute luminosity scale determination. Improved jet energy calibration and b jet identification can also be foreseen. Since the measurable portion of the phase space will be increased, in turn, the theory-based uncertainty in the extrapolation to the full phase space will be reduced. It can be then demonstrated that the projected inclusive  $\bar{t}t$  cross sections consistently have a sizeable impact on the proton PDFs, with the strongest effect on the gluon distribution.

## 7.2 Incremental improvements in luminosity determination

### 7.2.1 Refined impact of nonfactorizable beam shape

The vdM scan method assumes that the bunch proton density function is factorizable into independent  $x$ - and  $y$ -dependent terms. However, this assumption is not strictly valid, and leads to a biased estimate of the beam overlap area. The two-dimensional vertex distributions accumulated during the beam imaging scans in November 2015 are used to measure the bunch proton densities, extending the methodology described in Section 4.3.3.

The simplest model for the bunch proton density that has a correlated spatial dependence is a Gaussian distribution (Eq. 4.36) with  $x$ - $y$  correlations parametrized by a correlation parameter  $r$ . For the refined analysis, various parametrizations are considered to determine the optimal fit function to the vertex distributions. In the end, the fit is performed using a model that describes the proton density function  $\rho(x, y)$  with a main Gaussian component  $g_M$  (following the form of Eq. 4.36) with a large weight  $w_M$ , a wide component  $g_W$  with a small weight to model wide tails, and a narrow component  $g_N$  with a small but negative coefficient  $-w_N$  to model a flattened central part, i.e.,

$$\rho(x, y) = -w_N g_N(x, y) + w_M g_M(x, y) + (1 + w_N - w_M) g_W(x, y).$$

7.1

The reconstruction of bunch proton densities using a fit model with the weighted sum of three Gaussian distributions is also considered.

To derive a correction for the measured cross section from the vdM scan the fitted bunch proton densities are used to simulate vdM scans. The product  $\Sigma_x \Sigma_y$  from the MC

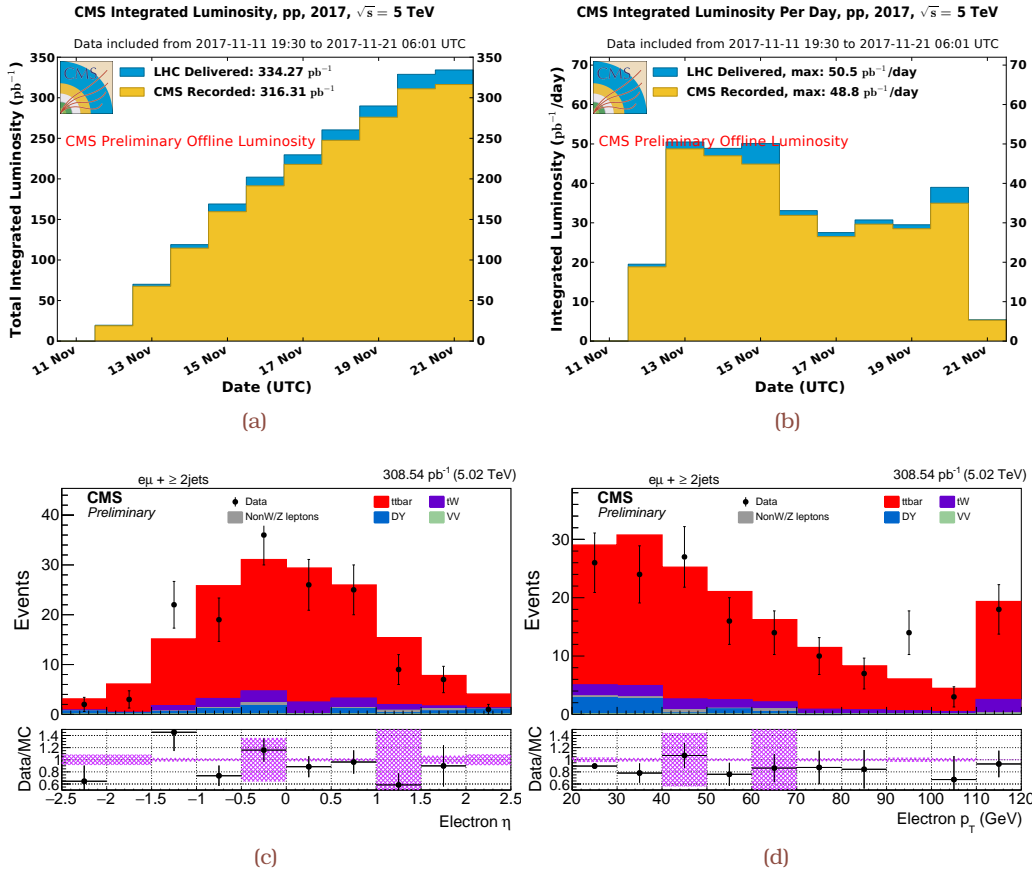
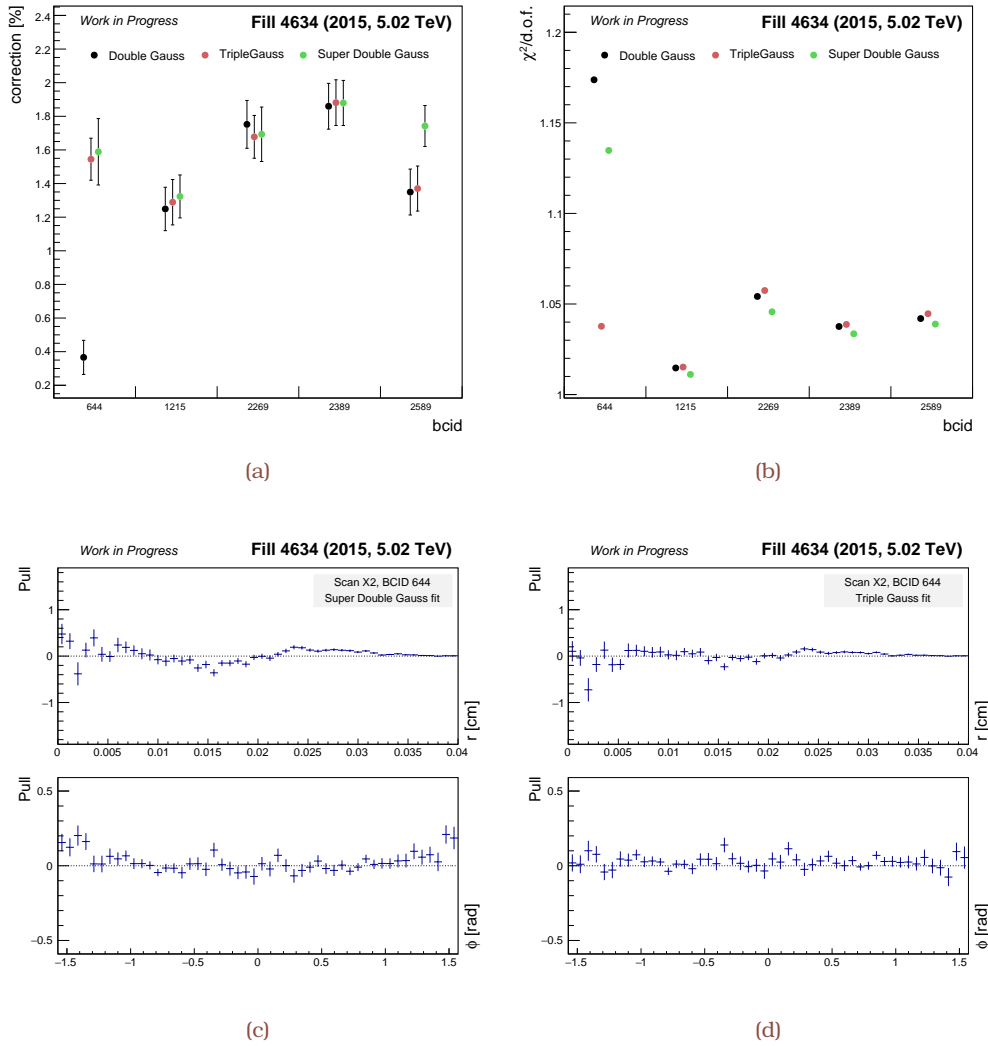


Figure 7.1: (a, b) Cumulative luminosity as a function of time delivered by LHC and recorded by CMS during the 5.02 TeVpp reference run in 2017 [288]. The plot makes use of the offline preliminary calibration that amounts to  $316.31 \pm 11.07 \text{ pb}^{-1}$  [AN10] under the condition of “stable beams” and with no requirement on the data quality in CMS. (c, d) Predicted and observed distributions of the electron  $\eta$  and  $p_T$  for events passing the dilepton criteria (Table 6.2), and after requiring at least two jets, in the  $e^\pm\mu^\mp$  final state (image courtesy of J. González, Oviedo).

simulation of the vdM scan method is then compared to the value from direct integration of the nonfactorized bunch proton densities, yielding an estimate of the inaccuracy introduced by using the beam overlap area  $\propto \Sigma_x \Sigma_y$ , which does not account for the  $x$ - $y$  correlations of the bunch proton densities. Multiple pseudo-experiments are performed to derive the central value of the correction and its statistical uncertainty.

Figure 7.2 shows the calculated corrections and the goodness-of-fit  $\chi^2$  divided by the number of degrees of freedom as a function of the bunch crossing identification number from different fits to the reconstructed vertex distributions. In general, the fit results are good, although there is some remaining mismodeling for BCID 644. A correction factor on the visible cross section of 1.5% is obtained taking the bunch-averaged correction and using the weighted sum of three Gaussian distributions. The uncertainty due to residual beam-shape effects in the bunch proton densities is estimated to be 0.2%, covering the range of correction factors obtained by the alternative model of Eq. 7.1 per BCID.



**Figure 7.2:** (a) Difference between the beam overlap area from the simulated vdM scans and the integrals incorporating genuine nonfactorizabilities as a function of the bunch crossing identification number in November 2015 scan program. (b) The goodness-of-fit  $\chi^2$  divided by the number of degrees of freedom in the modeling of the vertex distributions accumulated during beam imaging scans in November 2015. (c, d) Projections to radial and angular coordinates from the two-dimensional pull distributions of proton density models built based on (c) Eq. 7.1 and (d) a combination of three Gaussian pdf [AN11].

## 7.2.2 Concept and formalism of variable separation scan

To displace the beams at the IP the orbit is modified with a four-magnet closed orbit bump (Fig. 4.10) that allows establishing an orbit deformation with well-defined position and slope at any given point  $m$ , with  $m$  being located between the second and the third magnet. Considering the kick  $k_i$ , ( $i = 1 - 4$ ) instantaneous values in the field generated by a corrector  $i$ , the requirement for the bump to be “closed,” i.e., an unchanged orbit at

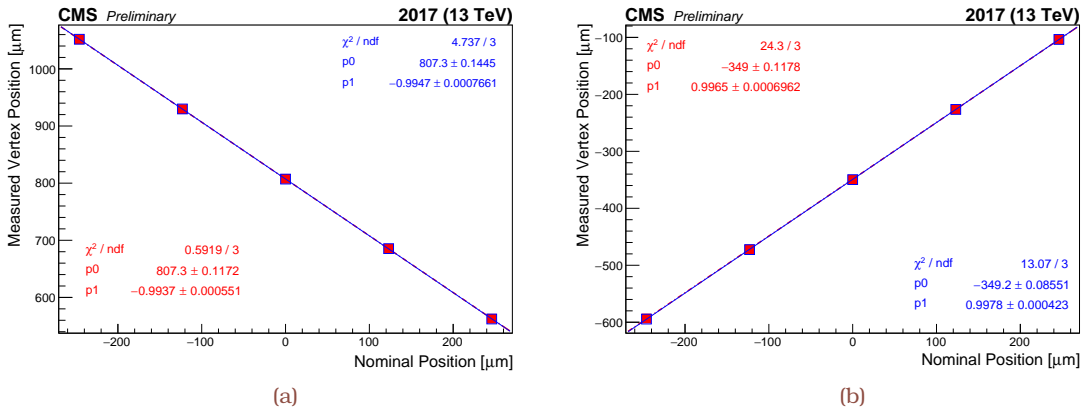


the first and last corrector can be expressed as a system of linear equations

$$\begin{pmatrix} x_m \\ \beta_m x'_m + a_m x_m \\ x_1 \\ x'_1 \end{pmatrix} = \begin{pmatrix} A \\ \beta_m A' + a_m A \\ 0 \\ 0 \end{pmatrix} = T \times C \begin{pmatrix} \sqrt{\beta_1} k_1 \\ \sqrt{\beta_2} k_2 \\ \sqrt{\beta_3} k_3 \\ \sqrt{\beta_4} k_4 \end{pmatrix}, \quad 7.2$$

where the transformation matrix  $T$  is given in and the constant  $C$  depends on the  $\beta_m$  function and the overall machine tune (ref);  $A$  and  $A'$  are the bump amplitude and slope, respectively. For the specific case of the variable separation scans each beam is displaced at the IP with both  $a$ —the slope of the  $\beta_m$  function—and  $x'_m$  equal to zero allowing only for a parallel separation.

A bump nonclosure would result in a scale factor error in the beam displacement which would directly modify the measured beam size. The origin of the nonclosure could be the combined effect of the “hysteresis,” i.e., distance (from the IP) dependent distortions in  $x_m$  and  $x'_m$ , and lattice imperfections. To minimize the hysteresis effect during the scan the translation is always performed in the same direction. In addition, for each scan an acquisition at zero separation is performed at the beginning, middle and the end of the scan; given the middle point lies on a different hysteresis branch as compared to the two other points, an indication of the hysteresis effect translates to a reduction of rate assuming the beam parameters do not significantly vary, e.g., beam emittance, intensity, etc.



**Figure 7.3:** Length scale calibration scan, using the variable separation procedure, for the  $x$  (a) and  $y$  (b) direction of beam 1 (red) and 2 (blue), respectively. Shown is the measured displacement of the luminous centroid as a function of the expected displacement based on the corrector bump amplitude. The line is a linear fit to the data. Errors are of statistical nature [TH8].

The variable separation scan for the length scale calibration was introduced in the July 2017 scan program using pp collisions at  $\sqrt{s} = 13$  TeV. Similar to the constant separation scan we make use of the CMS tracker to reconstruct the displacement of the luminous region. However, the variable separation scan is designed to measure the calibration

constant for each of the two beams in each of the two directions independently. The calibration data for both horizontal and vertical bumps of beam 1 and 2 are shown in Fig. 7.3. The scale factor which relates the nominal beam displacement to the measured displacement of the luminous centroid is given by the slope of the fitted straight line; the intercept is irrelevant. Because regular vdM scans are performed by displacing the two beams symmetrically in opposite directions, the relevant scale factor in the determination of the beam overlap is the average of the scale factors for beam 1 and 2 in each plane, i.e.,  $(0.9937 + 0.9947)/2 \approx 0.994$  and  $(0.9978 + 0.9965)/2 \approx 0.997$  in  $x$  and  $y$ , respectively. This represented an excellent agreement with the scale factors obtained based on the constant separation scan, though with much better precision [TH8].

## 7.3 Cross section measurements

### 7.3.1 Statistical and systematic uncertainty treatment

Based on the measurements of the inclusive  $\bar{t}t$  cross sections at  $\sqrt{s} = 5.02$  TeV [TH4], their performance can be conjectured using the data set collected in 2017 at the same center-of-mass energy. Although the higher instantaneous luminosity resulted in about three pp interactions per bunch crossing on average, pileup mitigation techniques are not crucial for a good performance of the  $\bar{t}t$  reconstruction. Because of the inclusion of the phase-I silicon tracker, the requirement of  $|\eta| < 2.1$  for electrons can be relaxed. At least two of the jets have to be identified as b jets, i.e., fulfilling a requirement of the CSVv2 algorithm, whereas alternative b tagging techniques can be used with a higher b-jet selection efficiency and rejection power for other jets in  $\bar{t}t$  events (Table 3.5).

Theoretical and modeling uncertainties make a significant contribution to the overall uncertainty. Since it is speculative to estimate a possible reduction of these uncertainties we consider the current ones as a conservative estimate. Improvements of the theoretical predictions can be expected, while further measurements at  $\sqrt{s} = 13$  TeV could reduce the modeling uncertainty.

The individual sources of uncertainty, as well as the assumed correlations, that are used as input to illustrate the impact of the conjectured  $\sigma_{\bar{t}t}$  measurements at  $\sqrt{s} = 5.02$  TeV on the knowledge of the proton PDFs are summarized in Table 7.1. The setup of the QCD analysis at NNLO is identical to the one used in Ref. [TH4], i.e., the total uncertainty in the measurements are propagated to the extracted QCD fit parameters using the MC method.

### 7.3.2 PDF constraints including the inclusive $\sigma_{\bar{t}t}$

For each replica, the PDF fit is performed, and the uncertainty is estimated as the RMS around the central value. The relative uncertainties in the gluon distributions, as obtained in the QCD analyses with and without the measured and conjectured values for  $\sigma_{\bar{t}t}$  at  $\mu_F^2 = 10^5$  GeV<sup>2</sup>, are shown in Figs. 7.4a and 7.4a, respectively. A further reduction of the uncertainty in the gluon distribution at  $x \gtrsim 0.1$  is observed, once the conjectured values for  $\sigma_{\bar{t}t}$  are included in the fit. Although they remained unaffected in Ref. [TH4],

**Table 7.1:** Conjectured inputs to the updated QCD analysis at NNLO using the data set collected in 2017 at  $s = 5.02$  TeV. Entries marked with one (two) “\*” delineate foreseen but difficult to quantify improvements over experimental (theoretical) point of view.

Final state	$e^{\pm}\mu^{\mp}$	$\mu^{\pm}\mu^{\mp}$	$l+t$ jets		
Central value (pb)	68.9	68.9	68.9		
Uncertainty (%)				Correlation	Note
Data sample event count	$25/\sqrt{10} \sim 8$	$48/\sqrt{10} \sim 16$	$9.5/\sqrt{10} \sim 3.0$	0	$\times 10$ expected yield
MC sample event count	$1.4/\sqrt{2} \sim 1.0$	$2.4/\sqrt{2} \sim 1.7$	$0.1/\sqrt{2} < 0.1$	0	$\times 2$ MC simulated yield
b tagging efficiency	-	-	3.4 (*)	-	New algorithms; performance to be validated
Electron efficiency	1.4 (*)	-	1.1 (*)	1	Some improvement upon endcap electrons
Muon efficiency	$3.0 \rightarrow 2.5$	$6.1 \rightarrow 5.0$	1.7	1	Less conservative approach
$p_T^{\text{miss}}$	-	0.7	-	-	
Jet energy scale	1.3	1.3	3.0	1	
Jet energy resolution	$< 0.1$	$< 0.1$	0.6	1	
QCD multijet background	-	-	2.4 (*)	-	Some improvement upon endcap electrons
W+jets background	$2.5/\sqrt{10} \sim 0.8$	$0.7/\sqrt{5} \sim 0.3$	$3.5/\sqrt{5} \sim 1.6$	1	$\times 10(5)$ due to data- (MC-) based estimation
tW background	1.4	1.6	1.3	1	
WV background	0.7	0.9	-	1	
Z/ $\gamma^*$ background	$2.7 \rightarrow \sim 2.0$	$15.4/\sqrt{10} \sim 5.0$	-	1	$\times 10$ due to data-based estimation; in $e^{\pm}\mu^{\mp}$ partly
$\mu_R, \mu_F$ scales of $t\bar{t}$ signal (PS)	1.2	1.7	4.4	1	
$\mu_R, \mu_F$ scales of $t\bar{t}$ signal (ME)	$< 0.1$	1.1	$< 0.1$	1	
Hadronization model of $t\bar{t}$ signal	1.2 (**)	5.2 (**)	2.8 (**)	1	
PDF	0.5	0.4	$< 0.1$	1	
Integrated luminosity	$2.3 \rightarrow 1.7$	$2.3 \rightarrow 1.7$	$2.3 \rightarrow 1.7$	1	see Section 7.2

a small improvement to the uncertainties in the valence quark distributions (Figs. 7.4c and 7.4d) is also observed, consistent with the outcome of Refs. [289] and [290].

It is thus demonstrated that the projected inclusive  $t\bar{t}$  cross sections have a stronger impact on the gluon distribution in the proton. Overall, this measurement is expected to profit from the higher amount of data collected in 2017, the improved phase-I silicon tracker detector, and the better knowledge of the absolute luminosity scale.

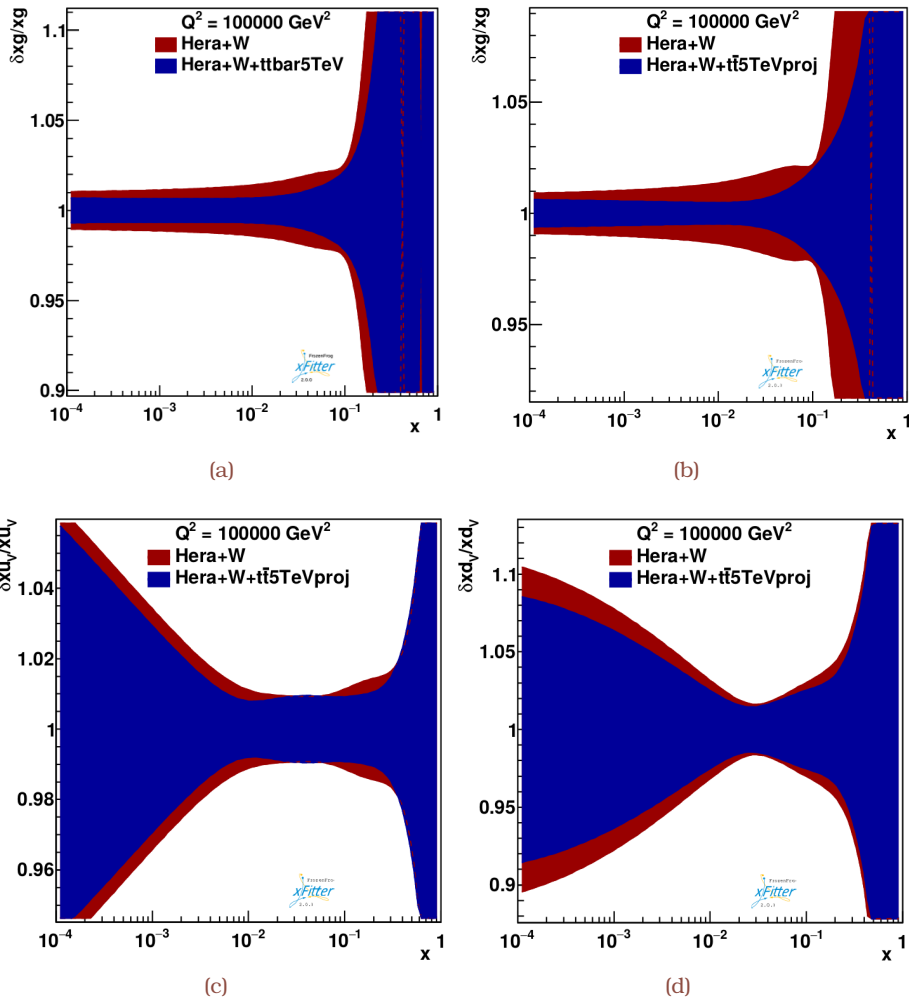


Figure 7.4: (a) The relative uncertainties in the gluon distribution functions of the proton as a function of  $x$  at  $\mu_r^2 = 10^5 \text{ GeV}^2$  from a QCD analysis using the HERA DIS and CMS muon charge asymmetry measurements, and also including the CMS  $\sigma_{\bar{t}t}$  results at  $\sqrt{s} = 5.02 \text{ TeV}$ . (b) Same as (a) but corresponding to the conjectured  $\sigma_{\bar{t}t}$  results. (c, d) Same as (b) but corresponding to the relative uncertainties in the valence quark distribution functions. The latter needs to be compared with Figs. 5.18d and 5.18d, respectively.

In November 2015, LHC delivered proton-proton (pp) collisions at a center-of-mass energy ( $\sqrt{s}$ ) of 5.02 TeV, corresponding to an integrated luminosity of  $27.4 \pm 0.6 \text{ pb}^{-1}$  in the CMS experiment. Measurements of the top quark pair production cross section ( $\sigma_{t\bar{t}}$ ) at various  $\sqrt{s}$  probe different values of the fraction  $x$  of the proton longitudinal momentum, and thus provide complementary information on the gluon content of the proton. Using  $t\bar{t}$  candidate events with  $l$ +jets, where leptons are either electrons ( $l = e$ ) or muons ( $l = \mu$ ), and dilepton ( $e^\pm\mu^\mp$  or  $\mu^\pm\mu^\mp$ ) final states, the result is  $\sigma_{t\bar{t}} = 69.5 \pm 6.1 \text{ (stat)} \pm 5.6 \text{ (syst)} \pm 1.6 \text{ (lumi)} \text{ pb}$  with a total relative uncertainty of 12%, that represents a remarkable achievement and a significant improvement relative to the first observation based on the  $e^\pm\mu^\mp$  final state alone.

The correlation in phase space of the jets from the W boson hadronic decay (“light jets”) carries a distinctive hallmark with respect to the main backgrounds that are controlled by counting the number of jets coming from the hadronization of the b quark (“b jets”) in the selected  $l$ +jets events. The signal extraction is then performed maximizing a profile likelihood fit to the distribution of a kinematic variable, sensitive to the resonant behavior of the light jets, for different categories of lepton flavor and jet multiplicity. Similarly to the most recent LHC studies of the inclusive  $\sigma_{t\bar{t}}$  the measurement is first performed in a fiducial phase space—a restricted region that closely resembles the detector acceptance in  $p_T$  and  $\eta$  of leptons and jets—and it is then extrapolated to the full phase space based on MC simulation. The individual and combined results are compared to the predictions from the ABMP16, CT14, MMHT14, and NNPDF3.0 parton distribution functions (PDFs). Theoretical predictions from different PDFs have comparable values and uncertainties, once consistent values of  $\alpha_s$  and  $m_{\text{top}}$  are associated with the respective PDF set. The limited-precision measurement can be complemented with the significantly larger pp data set recorded in 2017, equivalent to almost  $0.3 \text{ fb}^{-1}$ .

Asymmetric collisions of lead ( $^{208}\text{Pb}^{82+}$ ) nuclei with protons had not been included in the initial LHC design. However, unexpected discoveries in small collision systems, reminiscent of flow-like collective phenomena, engaged further investigations. After the short, yet remarkable, pilot physics run in 2012, and the first full one-month run in early 2013, the second full proton-nucleus run took place in late 2016 delivering collisions primarily at a nucleon-nucleon center-of-mass energy ( $\sqrt{s_{\text{NN}}}$ ) of 8.16 TeV, for each direction of the beams. Apart from complex bunch filling schemes due to the generation of the beams from two separate injection paths, the distinct feature of operation with asymmetric collisions at LHC is the difference in revolution frequencies. Given that the colliding bunches have significantly different size and charge, both beams are displaced transversely, onto opposite-sign off-momentum orbits, and longitudinally, to restore collisions at the proper interaction points. The achieved performance surpassed though almost eight times the designed instantaneous luminosity. The long-term integrated luminosity goal of  $100 \text{ nb}^{-1}$  has been surpassed, rendering the 2016 pPb run the baseline for several years.

Until recently, top quark measurements therefore remained out of reach in nuclear collisions due to the reduced amount of integrated luminosity produced during the first

period at the LHC, and the relatively low  $\sqrt{s_{\text{NN}}}$  values available at the BNL RHIC. Novel studies of top quark cross sections have finally become feasible with the 2016 pPb run at  $\sqrt{s_{\text{NN}}} = 8.16$  TeV. The top pair production cross section has been measured for the very first time in nuclear collisions, using a data set of  $174 \pm 6 \text{ nb}^{-1}$  in the CMS experiment. The measurement is performed analyzing events with exactly one isolated lepton and at least four jets, and minimally relies on assumptions derived by simulating signal and background processes. The resonant nature of the invariant mass of the two light jets,  $m_{jj}$ , provides a distinctive feature of the signal with respect to the main backgrounds, i.e., from QCD multijet and W+jets processes. The significance of the  $t\bar{t}$  signal against the background-only hypothesis is above five standard deviations. The measured cross section is  $45 \pm 8$  (total) nb, consistent with perturbative quantum chromodynamics calculations (using the CT14 proton PDF and the EPPS16 nuclear PDF for the lead ions) as well as the expectations from scaled pp data. To further support the hypothesis that the selected data are consistent with the production of top quarks, a “proxy” of the top quark mass is constructed as the invariant mass of candidates formed by pairing the W candidate with a b jet. This first study clearly paves the way for further detailed investigations of the top quark production in nuclear interactions [291], providing in particular a new tool for studies of the hot and dense matter created in nucleus-nucleus collisions [292–295].

Measurements of production cross sections provide fundamental tests of theoretical predictions. Increasingly higher precision both in the experimental measurements and the theoretical predictions is required to determine fundamental parameters of the standard model. At LHC, cross section measurements are often limited by the uncertainty in the integrated luminosity that is currently known with a precision of  $\mathcal{O}(2\text{--}4\%)$ , depending on the collision system. The luminosity calibration is based on the van der Meer scan technique, a purely experimental method. In dedicated sessions, the beam axes are moved in the transverse plane across each other such that the “beam overlap integral” can be determined. From the extracted integral, and the measured beam currents, the scale of the instantaneous luminosity is determined. To this end, several observables are used, each one corresponding to a cross section ( $\sigma_{\text{vis}}$ ) in the visible phase space region. The integrated luminosity for an arbitrary period of data taking is obtained from the accumulated counts of calibrated  $\sigma_{\text{vis}}$ . Relative nonlinearity and long-term stability in the response of the detector-based algorithms account for residual dependencies of the  $\sigma_{\text{vis}}$  ratios on conditions typical for normal physics operations.

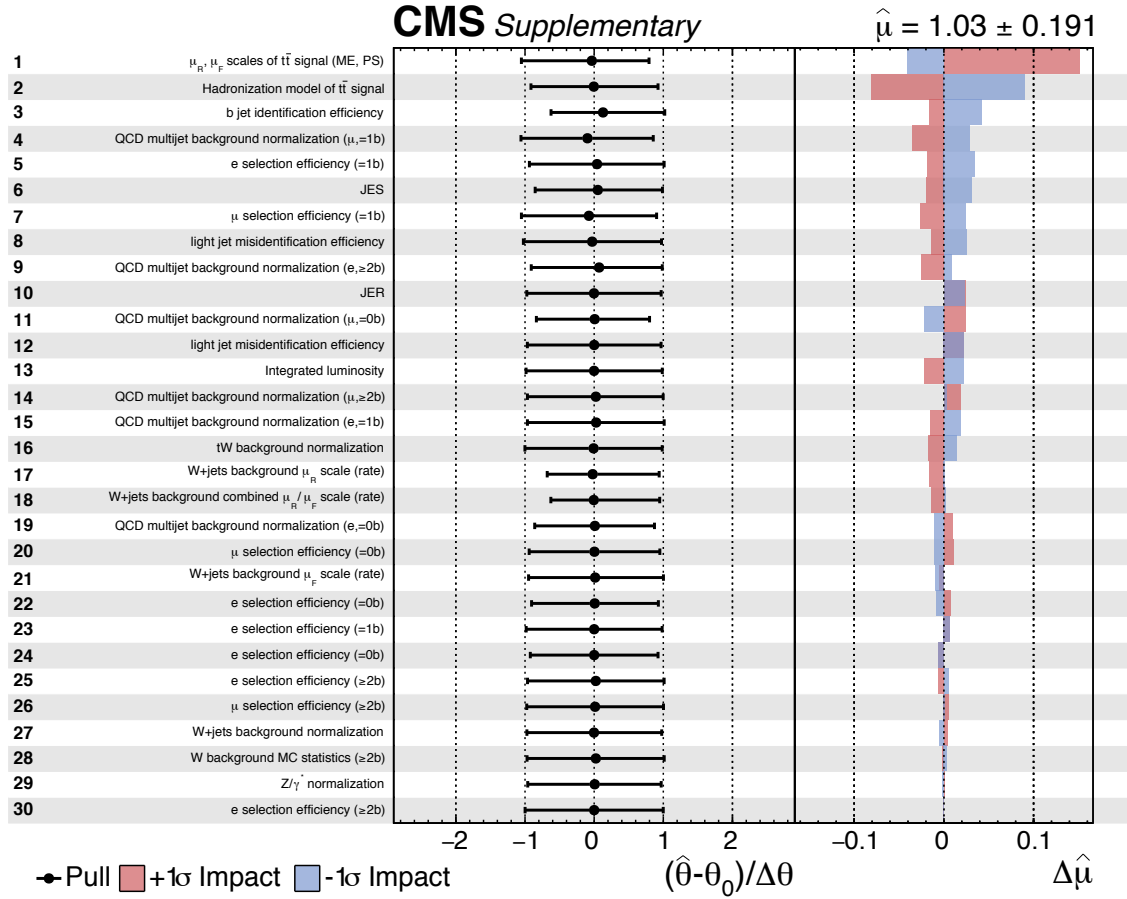
Significant improvements in the luminosity measurement are being planned, and a target uncertainty of 1% has been set for the High-Luminosity LHC [292, 293, 296]. Such improvement is expected to be achieved by combination of improved luminosity detector instrumentation—currently in the design phase—and refined analysis techniques, rapidly developing during the analysis of Run 2 data.

## Appendices





Appendix **A**: Summary of the impacts and pulls of the most significant nuisance parameters in the  $\ell$ +jets final state at 5.02 TeV



**Figure A.1:** Summary of the impacts and pulls of the most significant nuisance parameters used in the count analysis, when the fit is performed to the Asimov data set. In each plot the left panel shows the post-fit pull (value and uncertainty) of each nuisance, while the right panel displays the estimated impact on the fit for the signal strength. Only the first 30 nuisances are displayed, being their name shown at each row of the plots [TH4].

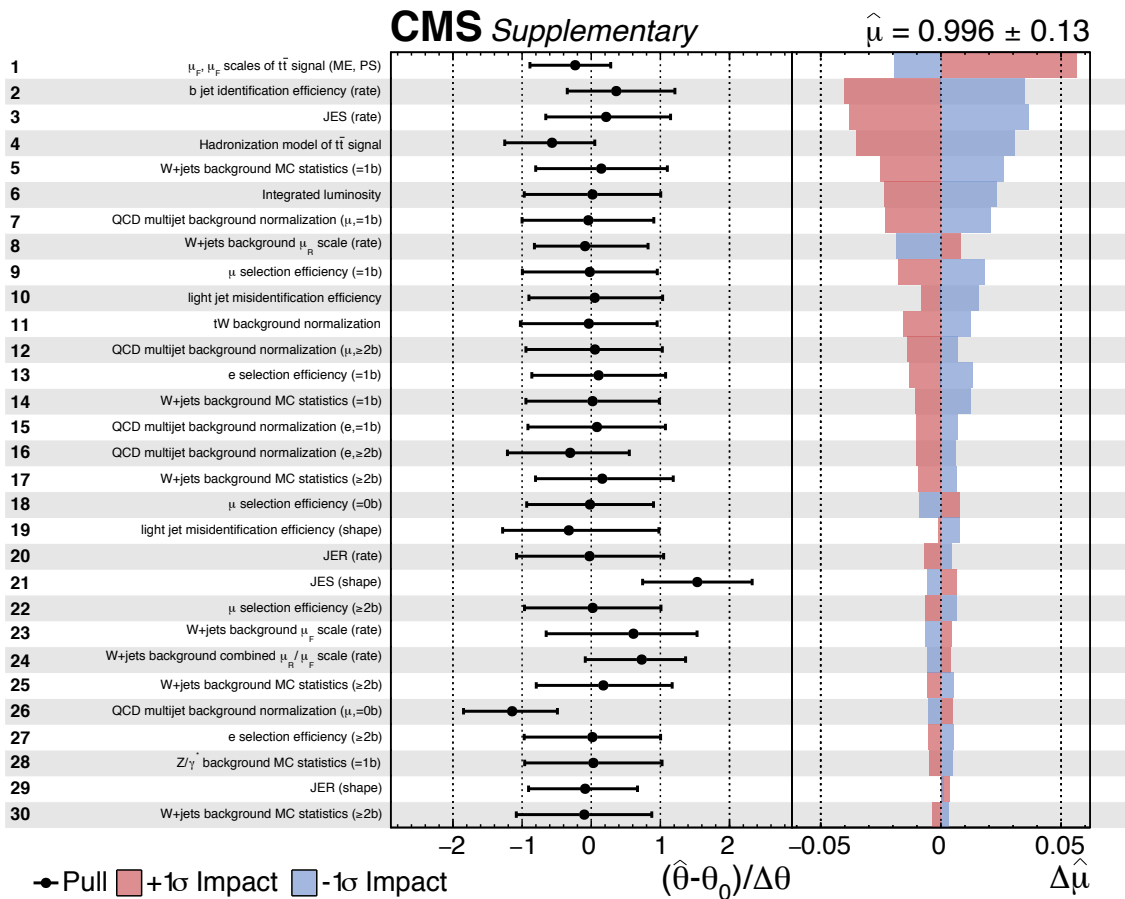


Figure A.2: Summary of the impacts and pulls of the most significant nuisance parameters used in the analysis of distributions, when the fit is performed to the Asimov data set. In each plot the left panel shows the post-fit pull (value and uncertainty) of each nuisance, while the right panel displays the estimated impact on the fit for the signal strength. Only the first 30 nuisances are displayed, being their name shown at each row of the plots [TH4].

The purpose of this Appendix is to describe the techniques and results of the electron and muon efficiency measurements, which have been included in a series of analyses using pp and pPb collisions at  $\sqrt{s} = 5.02$  and  $\sqrt{s_{NN}} = 8.16$  TeV, respectively. Simulated samples are used to derive the efficiency of the lepton trigger, isolation, reconstruction, and selection criteria, as a function of kinematic variables. These lepton-muon efficiencies are also directly estimated from data in Z boson enriched samples (Table B.1) using the “tag-and-probe” technique. To account for differences between data and simulation correction factors are computed from the ratio of the efficiencies measured in data to those calculated from the simulation, in bins of lepton  $p_T$  and  $\eta$ . For all considered cases, Drell-Yan events are simulated either at leading or at next-to-leading-order with the use of PYTHIA (v6.424) [49] or POWHEG (v2) [39,40,297] and MADGRAPH5\_aMC@NLO (v2.2.2) [38] generators interfaced to PYTHIA (v8.205) [51] for parton showering and hadronization (Tables 5.1 and 6.1), respectively.

Run Period	Data set indicative description	Integrated luminosity
pp, $\sqrt{s} = 5.02$ TeV	SingleMuon_Run2015E	27.4 pb <sup>-1</sup> [TH3]
pp, $\sqrt{s} = 5.02$ TeV	SingleElectron_Run2015E	27.4 pb <sup>-1</sup> [TH3]
pp, $\sqrt{s} = 5.02$ TeV	DoubleElectron_Run2015E	27.4 pb <sup>-1</sup> [TH3]
pPb, $\sqrt{s_{NN}} = 8.16$ TeV	SingleMuon_PARun2016C	174 nb <sup>-1</sup> [TH2]
pPb, $\sqrt{s_{NN}} = 8.16$ TeV	SingleElectron_PARun2016C	174 nb <sup>-1</sup> [TH2]

**Table B.1:** Data samples used for lepton efficiency measurements collected with the CMS detector at the end of 2015 and 2016 period at  $\sqrt{s} = 5.02$  and  $\sqrt{s_{NN}} = 8.16$  TeV, respectively. The quoted values for the integrated luminosity correspond to the offline calibration, as described in Section 4.

### B.0.1 The tag-and-probe technique

The tag-and-probe (“TnP”) technique takes advantage of a known mass resonance to select (“tag”) particles of the desired type, and to study (“probe”) the efficiency of a particular selection criterion on those particles. More specifically, the TnP method utilizes  $Z \rightarrow ee$  ( $Z \rightarrow \mu\mu$ ) decays as a high-purity source of unbiased electrons (muons) for computing electron (muon) efficiencies, following the method of Ref. [298]. On the one hand, the tag lepton is the control lepton to which stringent selection criteria are applied; thus the fake rate for passing-tag selection criteria should be small. The probe lepton, on the other hand, is the test lepton, whose selection criteria are varied and depend on the specifics of the selection being examined. By pairing these type of objects and imposing an invariant mass requirement on the TnP pair—in our case close to the Z boson pole mass [16]—one ensures to collect a high purity sample. Combinatoric backgrounds can be eliminated through any typical background subtraction methods such as fitting or sideband subtraction.

By definition, to fulfill the tag (“T”) selection implies that the lepton also satisfies the criteria to be a passing probe. Passing-probe leptons (“P”) are not typically of tag quality, but still satisfy the selection criteria for which the efficiency is being measured. Failing-probe leptons (“F”) account for the remainder of the electrons, which fail the selection criteria. Therefore the above classification entails  $3^2 - 3 = 6$  pair categories: “TT,” “TP,” “TF,” “PP,” “PF,” and “FF.” Defining  $\rho_T$ ,  $\rho_P$  and  $\rho_F$  as the probability of a lepton to be the tag, the passing probe or the failing probe, respectively, one trivially receives that

$$\rho_T + \rho_P + \rho_F = 1, \quad \text{B.1}$$

with the efficiency to be simply given by

$$\varepsilon_{\text{tot}} = \rho_T + \rho_P \leq 1. \quad \text{B.2}$$

With a sample of  $N(Z \rightarrow ee, \mu\mu)$  events, and with  $N_T$  the number of tag electrons out of  $N$ ,  $N_P$  the number of probe electrons out of  $N$  etc., the true sample efficiency  $\varepsilon_{\text{true}}$  would be equal to

$$\varepsilon_{\text{true}} = \frac{N_T + N_P}{N_T + N_P + N_F} = \frac{N_T + N_P}{2N}. \quad \text{B.3}$$

However, with the TnP method only a subset of the total  $N$  events are being considered, namely the studied events are the ones correctly identified in data by virtue of having a very clean tag electron. For such a subset one instead is left with

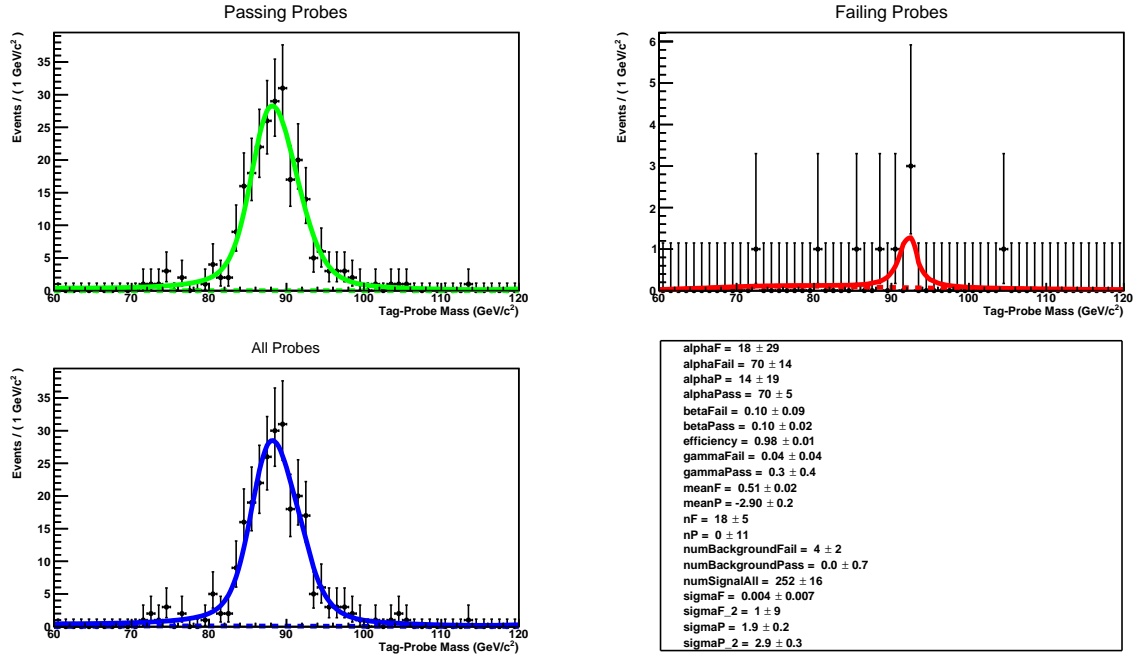
$$\varepsilon = \frac{2N_{\text{TT}} + N_{\text{TP}}}{2N_{\text{TT}} + N_{\text{TP}} + N_{\text{TF}}}, \quad \text{B.4}$$

where  $N_{\text{TT}}$  is the number of tag–tag electrons out of  $N$  etc. The factor of 2 in front of the TT events is due to the fact that electrons are not distinguishable in that case. In the large sample limit of  $N \rightarrow \infty$ , the equation above reduces to Eq. (B.2), i.e., the true sample efficiency. For the sake of brevity, we denote the nominator in Eq. (B.4) as  $N_P$  and the denominator as  $N_{\text{tot}} = N_P + N_F$ . In the present analysis, the estimation of  $Z \rightarrow ee$  ( $Z \rightarrow \mu\mu$ ) signal events in the passing and failing samples is carried out via a simultaneous maximum likelihood fit to the dielectron (dimuon) invariant mass  $m_{ee}$  ( $m_{\mu\mu}$ ) distributions in these two samples.

### Arbitration in case of multiple candidates

After applying the selection on  $m_{ee}$  ( $m_{\mu\mu}$ ), some events remain with more than one  $Z \rightarrow ee$  ( $Z \rightarrow \mu\mu$ ) candidate. In case there are more than one  $Z \rightarrow ee$  ( $Z \rightarrow \mu\mu$ ) candidates to select from, the following arbitration is made:

- If there are two probe candidates in the event and if only one of them also passes the tag criteria then choose the one which passes the tag criteria.
- If there are two probe candidates in the event and if both pass the tag criteria then choose the probe candidate with higher  $p_T$ .
- If there are two probe candidates in the event and if both fail the tag criteria then



**Figure B.1:** Example of TnP fit using  $Z \rightarrow ee$  events at  $\sqrt{s} = 5.02$  TeV; the distributions—from top left to bottom left—are given as a function of tag–probe pair mass and correspond to passing, failing and all probes. These examples correspond to the SC to RECO efficiency step (Table B.4) for a probe of  $40 < E_T < 50$  GeV within the  $-2.00 < \eta < -1.56$  region. Results of the simultaneous fit over the tp and tf mass shapes are shown in the bottom right plot, using different parameterizations for signal and background.

choose one of them randomly.

- If there are more than two probe candidates in the event then the event is rejected from further consideration.

It is thus evident that events, in which both electrons fulfill the tag criteria, contribute twice in the numerator and the denominator of Eq. (B.4), for the tag criteria should be stricter relative to the passing probe criteria.

## B.0.2 Electron efficiency

The efficiency calculation for electrons is decomposed into three steps:

- Super Cluster (SC) to Gaussian Sum Filter track-matched electron (“RECO”): the offline electron reconstruction efficiency, i.e., the probability that, given a SC is found, an electron is reconstructed and passes the offline selection.
- RECO to the respective identification working point (“ID”): the efficiency to pass the selection criteria specific to the measurement, including identification, isolation, and conversion rejection, given that the electron candidate has already passed the previous stage of the offline selection.
- ID to the online trigger requirement (“HLT”): the efficiency of the reconstructed and selected offline electron to be matched to the HLT object(s) under study.

In that case then, the total efficiency measurement can be factorized into three se-

quential measurements, hence given by the product

$$\varepsilon = \varepsilon_{\text{SC}} \times \varepsilon_{\text{ID}} \times \varepsilon_{\text{HLT}}$$

B.5

### Signal acceptance and candidate selection

It is required both electrons from  $Z \rightarrow ee$  decays to be within the ECAL fiducial area, i.e., within  $|\eta| < 2.5$ , but excluding the EB-EE transition region  $1.4442 < |\eta| < 1.560$ . Both electrons should have  $E_T > 10 \text{ GeV}$ , where  $E_T = \sum_i E_i \sin \vartheta_i$ ;  $E_i$  is the uncorrected energy seen by the calorimeters for the  $i^{\text{th}}$  particle,  $\vartheta_i$  is the polar angle of particle  $i$ , and the sum contains all particles emitted into a fixed solid angle in the event. Additionally,  $m_{ee}$  should be “close” to the nominal Z boson mass within a range of  $60 < m_{ee} < 120 \text{ GeV}$ . Typically, the choice of this mass window is driven by two considerations:

1. it should be consistent with the Z boson pole mass [16], while
2. it should be sufficiently broad that failing probes do not frequently fall outside the mass window.

Ideally, unprescaled single-electron triggers are used with the lowest- $E_T$  threshold available. However, such threshold had been set at  $E_T > 40 \text{ GeV}$  under the operating conditions at  $\sqrt{s} = 5.02 \text{ TeV}$ , thus double-electron triggers were applied at HLT event selection level. The  $Z \rightarrow ee$  sample for electron efficiency extraction therefore makes use of events that satisfy a diphoton trigger with symmetric transverse energy  $E_T$  thresholds of  $E_T = 15 \text{ GeV}$  covering the full tracker acceptance. Pairs of photon candidates above the  $E_T$  threshold are accepted only if their invariant mass is above  $50 \text{ GeV}$ . The trigger selection requires a loose identification using cluster shower shapes and a selection based on the ratio of the hadronic to the electromagnetic energy of the photon candidates. In Table B.2 the complete list of HLT filters, along with their associated L1 seeds, is shown, while in Table B.3 a detailed description of double-electron triggers is given.

Run period	HLT filter	L1 filter seed
pp, $\sqrt{s} = 5.02 \text{ TeV}$	DoublePhoton15_Eta2p5_Mass50_1000_R9SigmaHE	SingleEG20BptxAND
pp, $\sqrt{s} = 5.02 \text{ TeV}$	SinglePhoton40_Eta3p1	SingleEG20BptxAND
pPb, $\sqrt{s} = 8.16 \text{ TeV}$	DoublePhoton15_Eta3p1_Mass50_1000	SingleEG14BptxAND
pPb, $\sqrt{s} = 8.16 \text{ TeV}$	SinglePhoton20_Eta3p1	SingleEG10BptxAND

**Table B.2:** List of unprescaled single- and double-electron HLT filters, along with their L1 seeds, used for measuring the electron efficiencies. Electromagnetic (“EG”) energy deposits are reconstructed through the sum of the  $E_T$  deposited in two neighboring groups of  $5 \times 5$  ECAL crystals. Events with at least one such reconstructed deposit above a certain  $E_T$  threshold (10, 14, 20 GeV) are selected by the hardware level trigger. In the HLT level, stricter quality criteria are imposed to reduce the online filtering to acceptable rates. Coincidence with the bunch crossing signal sent by the BPTX (see Section 3.1.5) detector is required for all cases.

Following are the definitions of tag, probe, and passing probe in each efficiency step. They are also summarized in Table B.4.

Category	$R_9$ (5×5)	$\sigma_{\eta\eta}$ (5×5)	$H/E$
Barrel	$> 0.4$	$< 0.020$	$< 0.3$
Endcap	$> 0.5$	$< 0.045$	$< 0.2$
Other trigger requirements			
HLT seeded candidate “1” $E_T > 15 \text{ GeV}$   HLT seeded candidate “2” $E_T > 15 \text{ GeV}$   $m_{1,2} > 50 \text{ GeV}$			

**Table B.3:** Trigger selection logic for the double-electron HLT filters as listed in Table B.2. The HLT object is seeded by at least two L1 candidates having  $E_T > 15 \text{ GeV}$  and combinatorially paired together. Assuming each candidate mass to be that of the electron, any pair fulfilling a minimum invariant mass threshold of  $m_{1,2} > 50 \text{ GeV}$  is retained for further trigger selection. An upper bound of  $1000 \text{ GeV}$  was arbitrarily imposed. The rate increase, in case one or both of the seeded legs lay outside the ECAL barrel region, was modulated in pp collisions by sequentially imposing the two candidates to have  $R_9 > 0.4$  (0.5),  $\sigma_{\eta\eta} < 0.02$  (0.045) and  $H/E < 0.3$  (0.2) in the barrel (endcap).

### SC to RECO efficiency

- Tag: PF electron with  $E_T > 25 \text{ GeV}$  fulfilling the tight identification criteria
- Probe: SC
- Passing probe: SC matched to a reconstructed GsfElectron

### RECO to ID efficiency

- Tag: PF electron with  $E_T > 25 \text{ GeV}$  fulfilling the tight identification criteria
- Probe: PF electron
- Passing probe: PF electron fulfilling the identification, isolation and conversion rejection criteria

An overview of the electron kinematics and electron identification variables, the latter listed in Table 3.2), is shown in Figs. 3.18 and 3.19 comparing data with simulation within a mass range of  $60 < m_{ee} < 120 \text{ GeV}$  using pp and pPb collisions, respectively. Good agreement is observed in the region defined by the kinematic and identification requirements except for the isolation deposit variables because of the increased underlying event in the simulated pPb collisions (see Section 3.3.1).

### ID to the HLT efficiency

- Tag: PF electron with  $E_T > 25 \text{ GeV}$  fulfilling the tight identification criteria
- Probe: PF electron fulfilling the the identification, isolation and conversion rejection criteria
- Passing probe: PF electron having passed identification, isolation and conversion rejection criteria, and geometrically matched inside a cone of  $\Delta R < 0.3$  to the online trigger object.

To extract the signal yield of the tp and tf samples an unbinned maximum likelihood fit is performed to the  $m_{ee}$  variable with different modeling for the signal and background. The efficiency  $\varepsilon$  enters as an explicit fit parameter, such that correlations are taken into

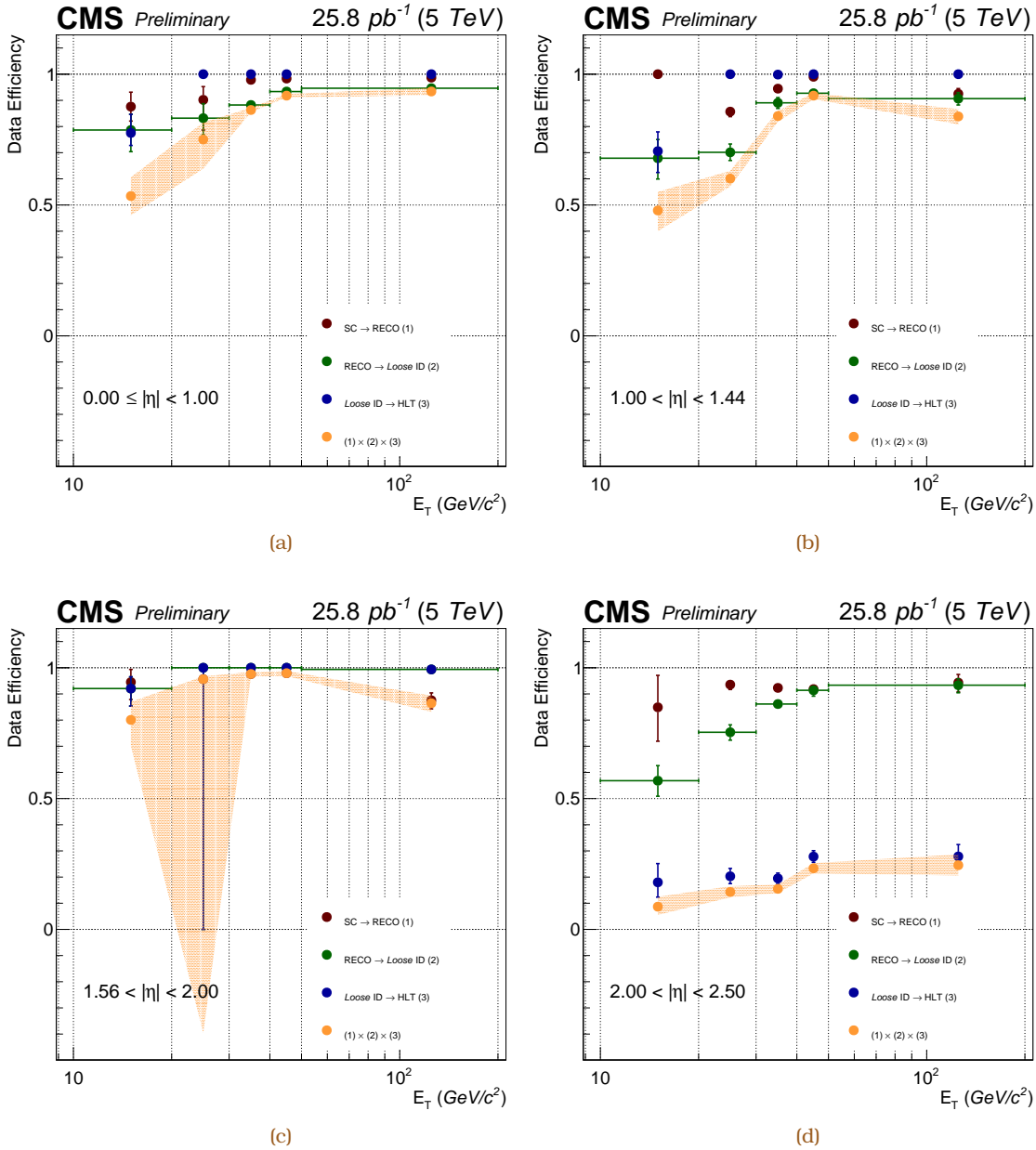


Figure B.2: Factorized (Eq. (B.5)) electron efficiencies measured in data as a function of probe  $E_T$  in  $0.00 \leq |\eta| < 1.00$  (a),  $1.00 \leq |\eta| < 1.44$  (b),  $1.56 \leq |\eta| < 2.00$  (c) and  $2.00 \leq |\eta| < 2.50$  (d) regions using the TnP technique (Table B.4 at  $\sqrt{s} = 5.02$  TeV).

account in the efficiency uncertainty extracted from the fit. The floating parameters of the fit are

1.  $N_{\text{sig}}$ : the number of signal tag and probe pairs,
2.  $N_{\text{bkg}}^{\text{P}}$ : the number of background tp pairs,
3.  $N_{\text{bkg}}^{\text{F}}$ : the number of background tf pairs,
4.  $\varepsilon$ : the signal efficiency,
5.  $2 \times 5$  parameters in the signal parameterization,
6.  $2 \times 3$  parameters in the background parameterization (Eq. (B.8)),

The signal and background are extracted in the tp and tf samples through the following



RECO				ID				HLT			
Type	ID	Requirements		Type	ID	Requirements		Type	ID	Requirements	
Object		$E_T$	HLT	Object		$E_T$	HLT	Object		$E_T$	HLT
PF electron	$t$	25	Table B.3	SE	$t$	25	Table B.3	PF electron	$t$	25	Table B.3
SC	$\times$	10	$\times$	PF electron	$\times$	10	$\times$	PF electron	$l/m$	10	$\times$
GSF electron	$\times$	10	$\times$	PF electron	$l/m$	10	$\times$	PF electron	$l/m$ ( $\Delta R < 0.3$ )	10	Table B.2

**Table B.4:** Requirements for the selection of tag and probe particles, along with them for tp, for each step in the electron efficiency measurement. The abbreviations l, m, and t stand for the loose, medium, and tight—the most stringent—conventions, respectively, in the electron identification (Table 3.2).

relations

$$N^P(m_{ee}) = N_{\text{sig}} \times \varepsilon \times P_{\text{sig}}^P(m_{ee}) + N_{\text{bkg}}^P \times \varepsilon \times P_{\text{bkg}}^P(m_{ee}), \quad \text{B.6}$$

$$N^F(m_{ee}) = N_{\text{sig}} \times (1 - \varepsilon) \times P_{\text{sig}}^F(m_{ee}) + N_{\text{bkg}}^F \times \varepsilon \times P_{\text{bkg}}^F(m_{ee}), \quad \text{B.7}$$

where  $P_{\text{sig}}^P(m_{ee})$  and  $P_{\text{sig}}^F(m_{ee})$  are the signal probability density functions (pdfs) as a function of the  $m_{ee}$  variable in the tp and tf samples respectively.  $P_{\text{bkg}}^P(m_{ee})$  and  $P_{\text{bkg}}^F(m_{ee})$  are the background pdfs as a function of the  $m_{ee}$  variable, in the tp and tf samples respectively, and are given by

$$P_{\text{bkg}}^P(m_{ee}) = \text{erfc}[\beta \times (a - m_{ee})] \times \exp[(m_Z - m_{ee}) \times \gamma], \quad \text{B.8}$$

similarly for  $P_{\text{bkg}}^F(m_{ee})$ .

In this approach, the tp and tf mass shapes for the signal is built based on a Crystal Ball (CB) pdf, whose parameters are allowed to float in the fit freely. More specifically, defining  $t = (m_{ee} - \text{mean})/\sigma_1$  and  $t_0 = (m_{ee} - \text{mean})/\sigma_2$ , for our case the CB function equals to

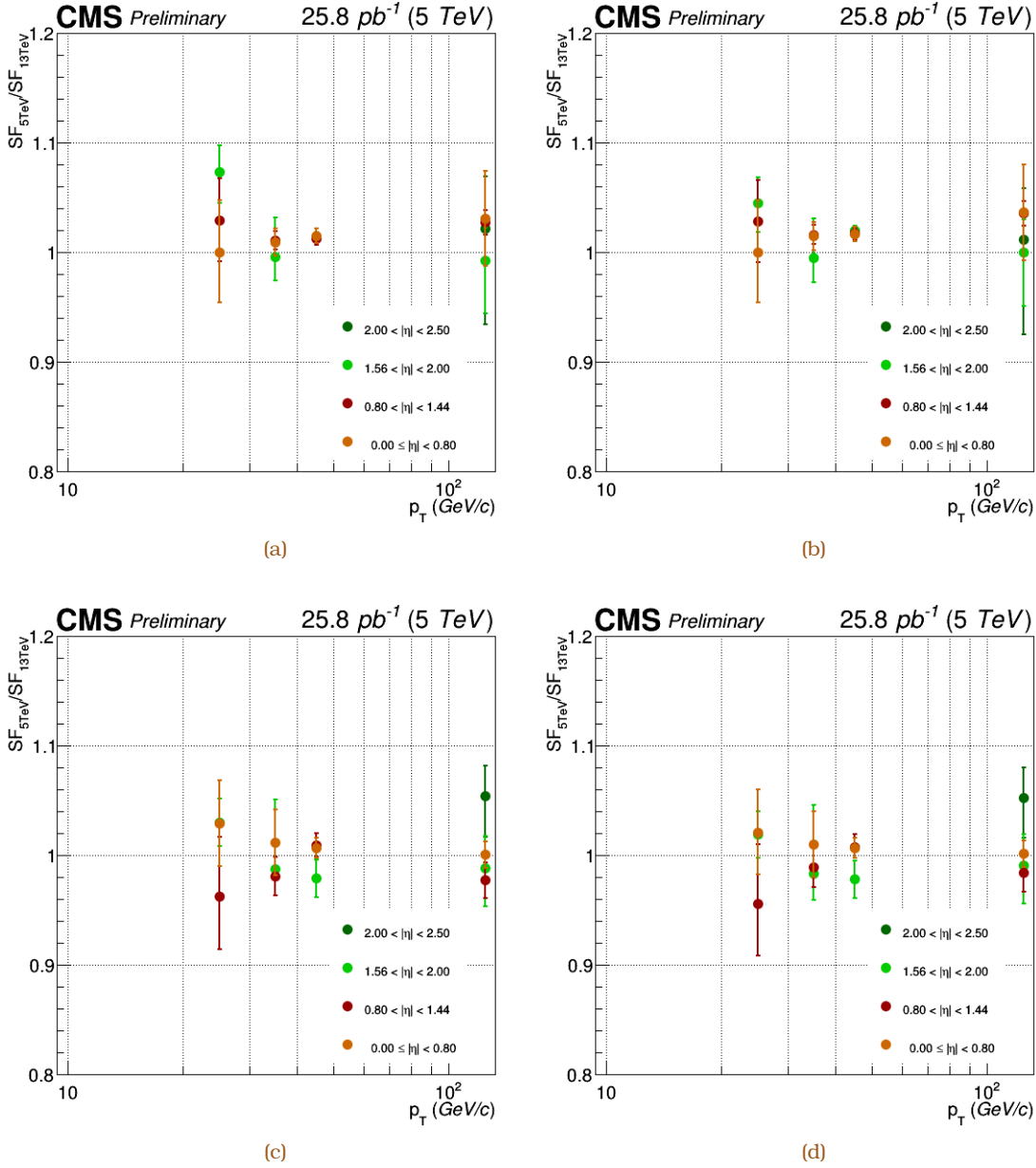
$$f \propto \begin{cases} \exp\left(-\frac{1}{2}t_0^2\right) & t > 0, \\ \exp\left(-\frac{1}{2}t^2\right) & t > -|a|, \\ (b - t)^{-n}, \quad b = n/|a| - |a| & \text{otherwise.} \end{cases}$$

Such a CB pdf modified to include a Gaussian is empirically said to properly describe the detector simulation and the data quite well, for instance Fig. B.1 exhibits. An empirical approach is considered as to what parameters of the line shape model are allowed to float. It is found that if we allow as many parameters as previously listed (also shown in Fig. B.1) to float, the fit converges and uncertainties in each of the floating parameters are properly calculated <sup>1</sup>.

The electron efficiencies are estimated in bins of probe electron  $E_T$  and  $|\eta|$ . The following bin edges have been coherently used over all efficiency steps:

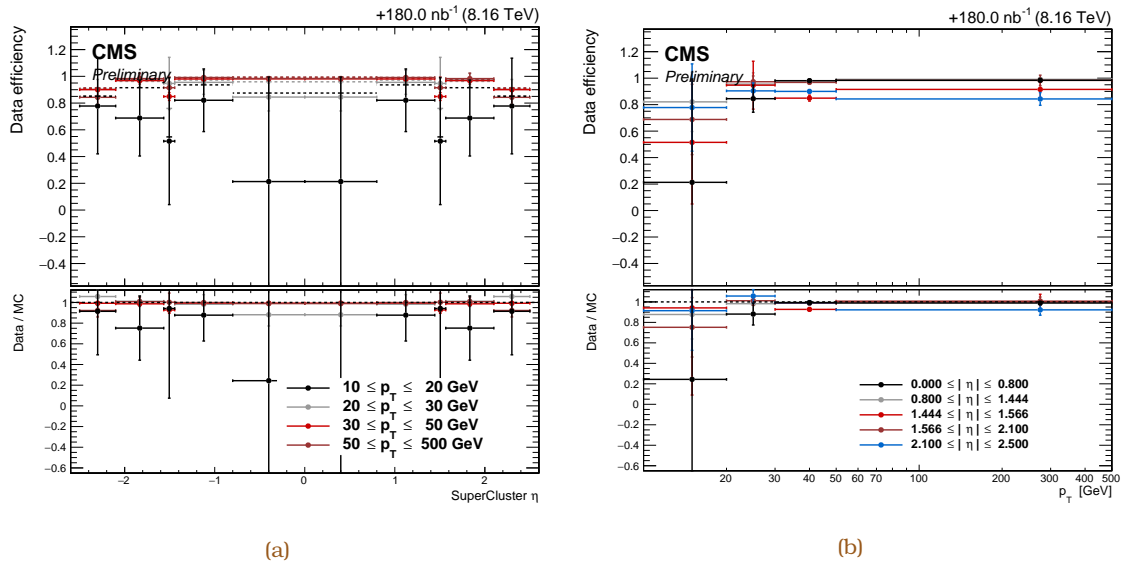
- probe electron  $E_T$  (GeV): 10, 20, 30, 40, 50, and 200
- probe electron  $|\eta|$ : 0, 1.0, 1.4442, 1.56, 2.0, and 2.5

<sup>1</sup>Exceptions were found for some low- $E_T < 20$  GeV bins.



**Figure B.3:** Data to simulation ratios in the 5 and 13 TeV running conditions are determined using the TnP technique as a function of the probe  $p_T$  in  $|\eta| < 0.80$ ,  $0.80 \leq |\eta| < 1.44$ ,  $1.56 \leq |\eta| < 2.00$  and  $2.00 \leq |\eta| < 2.50$  regions for the RECO (a, b) and ID (c, d) steps. The comparison is separately performed using the signed  $\eta$  values at 13 TeV, i.e.,  $\eta > 0$  (a, c) and  $\eta < 0$  (b, d).

The calculated one-dimensional event efficiencies and scale factors are shown in Figs. B.2–B.3 and Figs. B.4–B.3 using pp and pPb collisions, respectively. Overall, no dependency on the sign of the probe electron  $\eta$  for all efficiency steps is found, meaning the results are given in unsigned probe electron  $|\eta|$  bins. Concerning the plateau of the used double-electron HLT, it is found already at  $E_T = 25$  GeV, assuming that there is no asymmetry dependence between the two trigger legs in the efficiency factorization formula. The decreased double-electron HLT efficiency is because of the  $|\eta|$  requirement on the L1



**Figure B.4:** Electron efficiency, as measured in data, and data to simulation ratios using the TnP technique, as a function of the probe  $\eta$  (a) or  $p_T$  (b), for the RECO step at  $\sqrt{s_{NN}} = 8.16$  TeV.

trigger seed. Pertaining to the identification efficiencies at  $E_T > 40$  and  $> 20$  GeV the loose and medium identification requirements retain about 90 and 70% of the electrons, respectively. For the  $E_T < 20$  GeV bin, the low event count induces some instability to the fit, and hence bin-by-bin variations are observed for that region. In addition, such variations are prominent for the barrel-endcap transition regions, albeit electrons falling inside such a window are excluded from any further consideration. Finally, the combination of all efficiency steps using pp and pPb collisions are presented in Fig. B.2 and Table B.7, respectively.

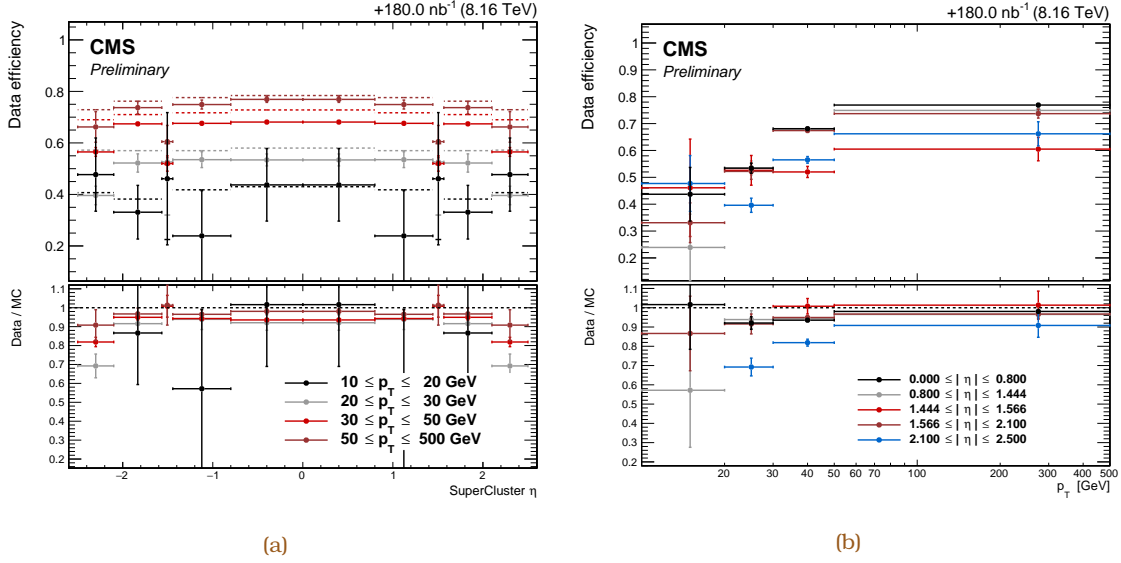
In conclusion, data-based efficiencies suggest that excluding the barrel-endcap transition window and restricting the electron selection in a phase space of  $E_T > 40$  ( $> 20$ ) GeV and within the  $|\eta| < 2.1$  region, about 85 (63)% of electrons are reconstructed, identified, and fulfilling the trigger requirements in pp (pPb) collisions.

### B.0.3 Muon efficiency

Run period	HLT filter	L1 filter seed
pp, $\sqrt{s} = 5.02$	SingleMuon15_HIL3	L1_SingleMu7_BptxAND
pPb, $\sqrt{s} = 8.16$ TeV	SingleMuon12_PAL3	L1_SingleMu7_BptxAND

**Table B.5:** List of unrescaled single-muon HLT filters, along with their L1 seeds, used for measuring the muon efficiencies

For the muon case, the efficiency is calculated requiring the tag muon to pass the HLT filters, as given in Table B.5. For the muon component of the L1 trigger, CSC and DT chambers provide the trigger primitives constructed from hit patterns consistent with muons that originate from the collision region, and RPC chambers provide hit information.



**Figure B.5:** Electron efficiency, as measured in data, and data to simulation ratios using the TnP technique, as a function of the probe  $\eta$  (a) or  $p_T$  (b), for the ID step at  $\sqrt{s_{NN}} = 8.16$  TeV.

At L1, muons are requested to be either standalone (CSC, DT, or RPC) or matched (DT to RPC or CSC to RPC) candidates. The L1 trigger was updated during 2016 and is further divided into three subsystems, i.e., the barrel, endcap, and overlap muon track algorithms. The coincidence with a bunch crossing identified with the BPTX detector is required to ensure the presence of real collisions. Once an event is selected by the L1 filter, it proceeds to sequential HLT steps, that take as input the information from the preceding step and make use of more elaborate algorithms.

The probe must be then a reconstructed muon forming a dimuon system with the tag, the latter of tight identification and isolation quality, and a pair mass in the range of 60–120 GeV. The total single muon efficiency is factorized in four steps, for which the probe and the passing probes are described in Table B.6. The fitting models observed to appropriately describe the data are summarized also in the same Table; events are separately counted in all categories according to the pass or fail status of the probe muon.

Efficiency step	STA RECO	Inner tracking	ID	HLT
$ \eta $ bins	0-1.2-1.8-2.1-2.4	0-2.4	0-1.2-1.8-2.1-2.4	0-1.2-1.8-2.1-2.4
pp nominal fitting				
Signal pdf	Double Gaussian	Double Gaussian	CB+Gaussian	CB+Gaussian
Background pdf	1st order Chebyshev	2nd order Chebyshev	1st order Chebyshev	1st order Chebyshev
Fit mass range (GeV)	(60, 120)	(60, 120)	(60, 120)	(60, 120)

**Table B.6:** Summary of the fit configuration used to obtain the muon efficiencies for each step using the TnP technique.

### Efficiency of standalone muon reconstruction

- **Probe:** tracks within the fiducial volume fulfilling the same selection as the inner

tracks associated to muons with at least five layers of the tracker including at least one pixel hit.

- **Passing probe:** probe matched with a standalone (STA) muon with at least one valid muon hit.

#### Efficiency of inner track reconstruction

In principle, this constitutes a twofold process combining the inner tracking (geometrical matching between STA and inner track) with the global muon fit efficiencies. The latter is highly efficient; hence the results reflect mostly the inner track efficiency, i.e.,

- **Probe:** STA muons with at least one valid muon hit.
- **Passing probe:** probe that is a global muon.

#### Efficiency of muon identification and isolation selection

- **Probe:** global muon.
- **Passing probe:** probe that satisfies all muon identification and isolation (“ID+ISO”) selection

#### Efficiency of the event filter

- **Probe:** global muon passing the ID+ISO muon selection
- **Passing probe:** probe geometrically matched to the single muon HLT filter

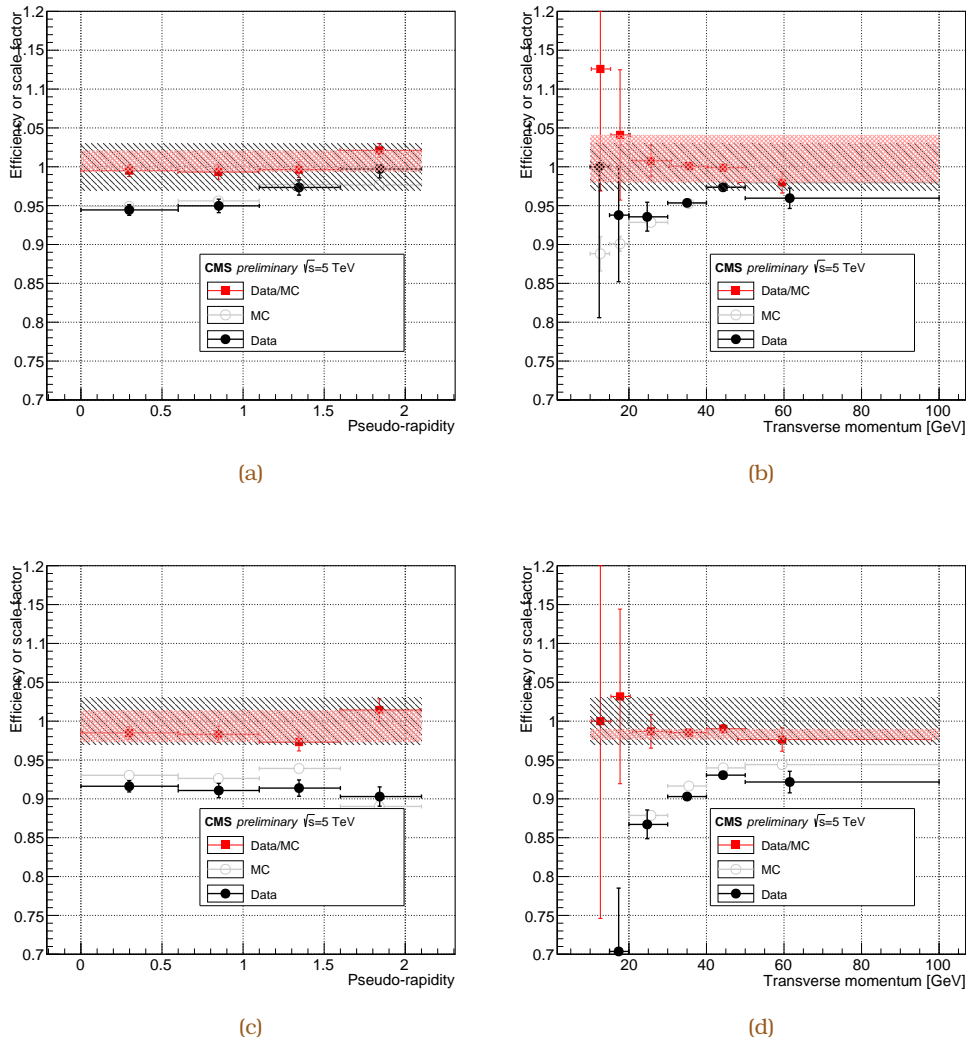
Table B.7 summarizes the inclusive trigger and selection efficiencies measured for electrons and muons using pPb collisions at  $\sqrt{s_{\text{NN}}} = 8.16$  TeV. These values are computed by convoluting the differentially measured efficiencies with the expected lepton  $p_{\text{T}}$  and  $\eta$  distributions in the PYTHIA  $\bar{t}t$  MC sample (Table 6.1). The values are also compatible between the Ppb and pPb running periods, and the residual difference is assigned as an additional systematic uncertainty.

Parameter	Final state	
	$\mu$ +jets	e+jets
$\varepsilon_{\text{RECO}}$	0.998	$0.95 \pm 0.01$
$\varepsilon_{\text{ID+ISO}}$	$0.96 \pm 0.04$	$0.69 \pm 0.03$
$\varepsilon_{\text{HLT}}$	$0.95 \pm 0.01$	$0.95 \pm 0.01$

**Table B.7:** Average lepton efficiencies as calculated from data separately in the  $\mu$ +jets and e+jets final states at  $\sqrt{s_{\text{NN}}} = 8.16$  TeV. For the combined identification and isolation efficiency an additional 4% systematic uncertainty due to the presence of the underlying event is taken into account (see Section 3.3.1).

Figure B.6 summarizes the results of the data- and MC-based efficiencies, as well as their ratio, for the different muon requirements using pp collisions at  $\sqrt{s} = 5.02$  TeV. The scale factors are overall consistent with unity in the  $p_{\text{T}} > 18$  GeV and  $|\eta| < 2.1$  range. The observed point-to-point deviations are well within an overall 3% variation that is considered as the uncertainty due to the muon trigger and selection efficiency.

Figures B.7 and B.8 summarize the results of the data- and MC-based efficiencies, as well as their ratio, as a function of  $\eta$  and  $p_{\text{T}}$  when the probe is required to pass the tracking and ID+ISO requirements, respectively, using pPb collisions at  $\sqrt{s_{\text{NN}}} = 8.16$  TeV. It can be noticed that the STA efficiency is found to be almost 100% with negligible



**Figure B.6:** Efficiency measurements for ID+ISO (a, b) and the combined ID+ISO with HLT (c, d) requirements. The results are separately summarized as a function of the probe muon  $|\eta|$  (a, c) or  $p_T$  (b, d). For each plot, the absolute efficiencies measured in data and expected in simulation, as well as the ratio of the two, are given. The red dashed band corresponds to the envelope derived using the points for which the scale factor is determined with an uncertainty  $< 15\%$ , while the gray dashed band corresponds to the final envelope considered in the pp analysis at the level of  $3\%$ , taking into account also the HLT efficiency measurements (Fig. 3.6).

uncertainty. Any uncertainty stemming from this assumption is expected to be negligible for the analysis. As for the electrons, the ID+ISO requirement introduces a dependency on the simulated underlying event, i.e., a strong  $p_T$  dependency of the scale factor.

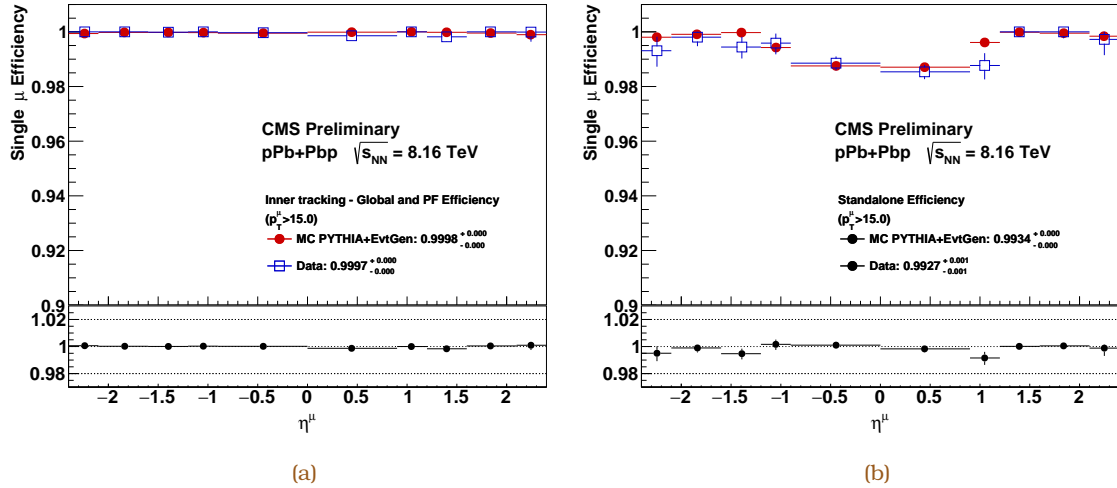


Figure B.7: Data- and MC-based efficiencies, as well their ratio, for the muon reconstruction as a function of the probe  $|\eta|$  integrated at  $p_T > 15$  GeV. The dips in efficiency close to  $|0.3|$  are due to the regions with less instrumentation between the central muon wheel and the two adjacent wheels.

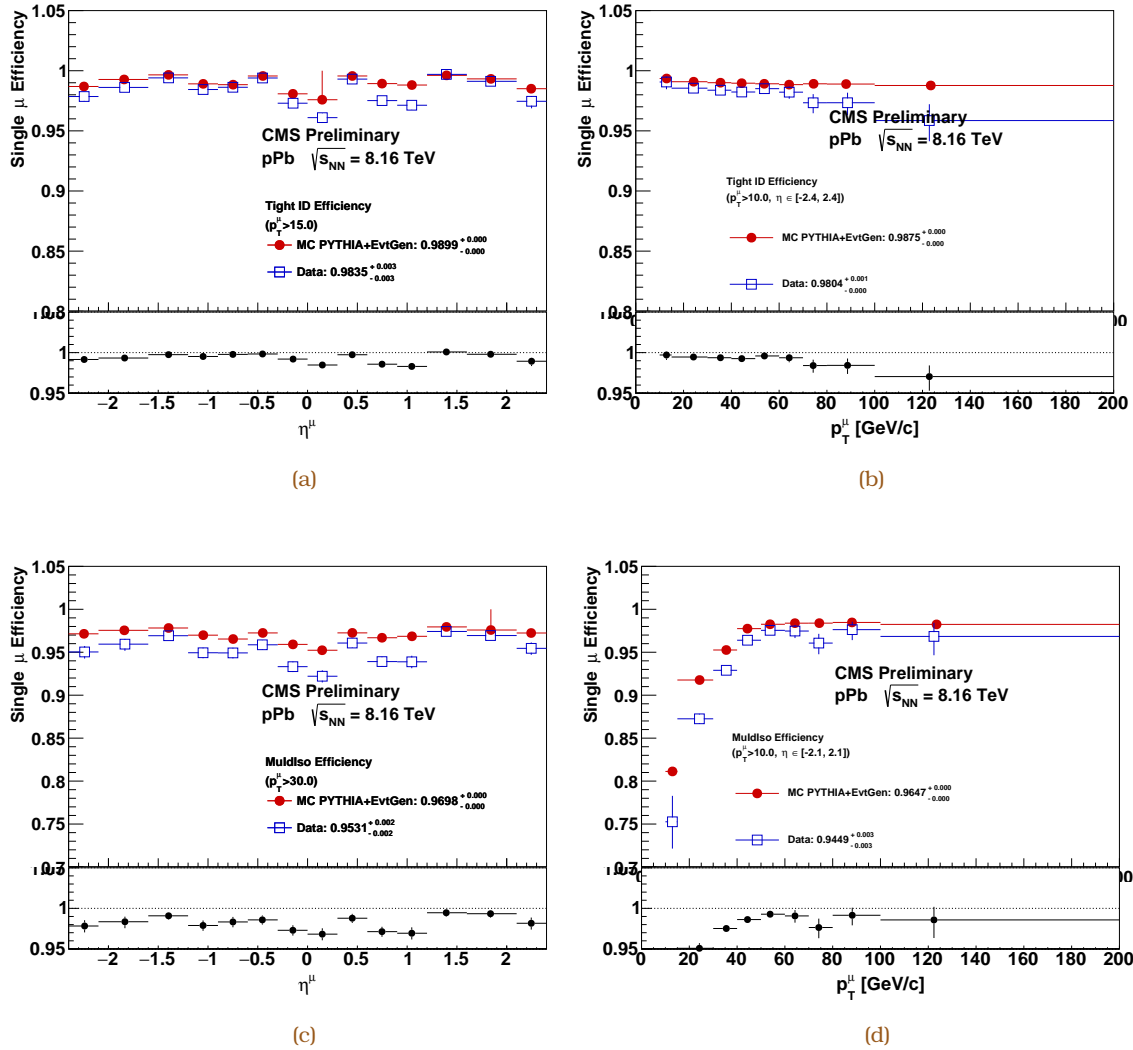


Figure B.8: Comparison between the measured and expected identification (a, b) and the combined identification with isolation (c, d) muon efficiencies as a function of the probe  $\eta$  (a, c) or  $p_T$  (b, d). The simulation is systematically higher than the data as a result of small imperfections in the model, which are revealed by the stringent requirements for a muon to satisfy tight identification criteria.

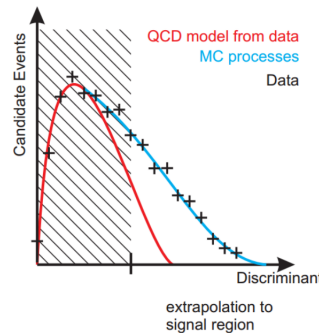


QCD multijet production is an important background to measurements of single top quark cross sections, e.g., in the  $t$  channel. These events become signal-like when an additional lepton is presented in the final state, where a lepton could simply be the decay product of a hadron formed within jets. Typically, such “nonprompt” leptons, as opposed to leptons having been produced from the hard scattering, are surrounded by many activities within the jet. On the other hand, leptons arisen out of the  $W$  boson leptonic decay are expected to be isolated, meaning that they should be found in a restricted spatial region where they contribute with a significant fraction in the deposited energy. Therefore a QCD enriched data sample, yet orthogonal with the signal region, is achieved using leptons that fail the isolation requirement (Eq. (3.2)) since the region with reduced isolation starts to become overwhelming in QCD [TH5]. The question left to be answered concerned with the isolation region(s) that allow the optimal reduction of QCD, however, at the same time retaining as much QCD as required to have a proper data-based modeling of this background process, plus of its verification.

Given the associated theoretical uncertainty in the modeling of the QCD multijet process, for instance stemming from a not precise enough knowledge of the cross section or the approximate description of the underlying hadronic activity, it is necessary to predict the size and properties of this process by data. A reliable model for determining QCD contamination is then derived by fitting templates of all physics processes to the  $x$  ( $\equiv m_T^W$ ) discriminant (Fig. C.1), i.e.,

$$F(x) = N_{\text{QCD}} \times Q(x) + N_{\text{non-QCD}} \times W(x) . \quad \text{C.1}$$

The  $W(x)$  templates, representing the signal and other non-QCD background processes, are taken from MC simulation (Table C.1). After having subtracted non-QCD contamination, the  $Q(x)$  template is constructed from a dedicated control sample that will be described in the following.



**Figure C.1:** A fictitious example that establishes the basic principle behind data-based QCD multijet estimation. The QCD shape distribution is extracted by data over a phase space in the proximity of the signal region. The total amount of QCD events is then determined by a template fit having utilized a variable that discriminates QCD against non-QCD processes.

**Table C.1:** Event yields for the main processes in the 2j1t category, satisfying the selection as described in Ref. [TH13] and using  $2.3 \text{ fb}^{-1}$  of pp collisions at  $\sqrt{s} = 13 \text{ TeV}$ . The quoted uncertainties are of statistical nature. All yields are taken from simulation, except for QCD multijet events where the yield and the associated uncertainty are determined from data.

Process	$\mu^+$	$\mu^-$
Top quark ( $t\bar{t}$ and $tW$ )	$6837 \pm 13$	$6844 \pm 13$
V (V=W, Z)+jets	$2752 \pm 82$	$2487 \pm 76$
QCD multijet	$308 \pm 154$	$266 \pm 133$
Single top quark $t$ -channel	$1493 \pm 13$	$948 \pm 10$
Total expected	$11390 \pm 175$	$10545 \pm 154$
Observed	11877	11017

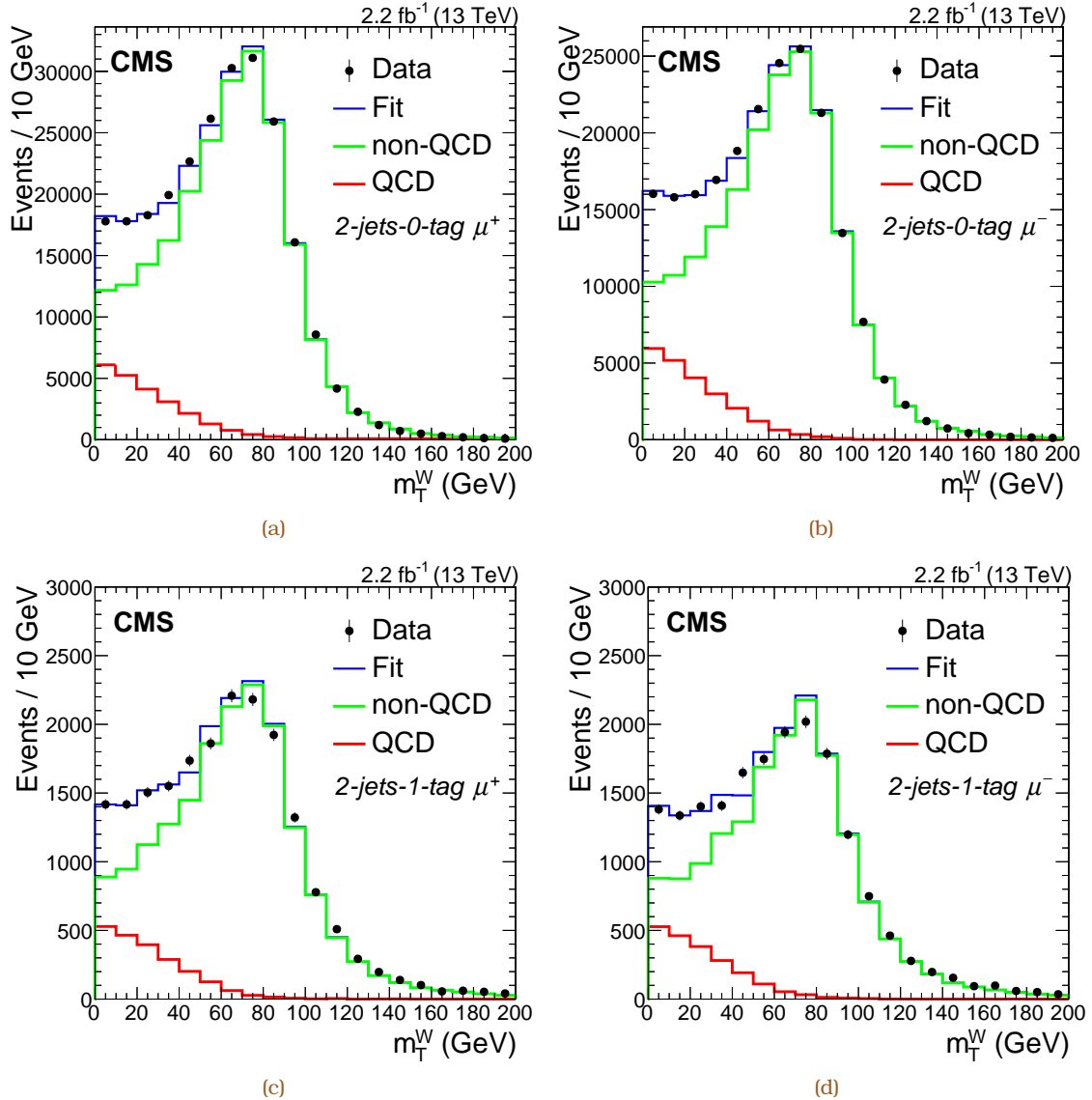
As important as rejecting QCD away is, constraining the amount of its residual contribution to the analyses sensitive window is equally crucial. Indeed, Figs. C.2a - C.2d are representative examples of the reconstructed  $m_T$  in the 2j0t and 2j1t regions within a  $m_{\ell\nu b}$  window around the top quark mass (SR) containing leptons with  $I_{\text{rel}}^\mu < 0.06$ . More specifically, reasonable agreement between data and simulation was found, after both QCD and non-QCD components have been scaled to the estimated contribution, the latter determined by the fit process in Eq. (C.1). Thus a separation of  $I_{\text{rel}}^\mu > 0.12$ – $0.15$  effectively results in a high statistics model for the QCD background, while non-QCD contamination becomes less dominant. The region in between  $0.06 < I_{\text{rel}}^\mu < 0.12$ – $0.15$  can be then used for the verification of such *in-situ* model for the QCD multijet, in a way similar to what was performed in Ref. [TH12].

By reperforming the template QCD against non-QCD fit over the  $m_T$  variable in a range of  $0.01 < I_{\text{rel}}^\mu < 0.15$ , it was found that the separation in the three distinct  $I_{\text{rel}}^\mu$  regions seems to be optimal. Just to note that the choice of  $I_{\text{rel}}^\mu < 0.15$  was simply made in view of the centrally provided working point for the muon isolation. In Fig. C.3,  $S/B$  is plotted as a function of  $I_{\text{rel}}^\mu$ , with  $S$  and  $B$  the number of best-fit non-QCD and QCD multijet events, i.e., as determined from Eq. (C.1). An additional cross-check has been done, by means of using the template QCD against non-QCD fit over the  $p_T^{\text{miss}}$  variable (Fig. C.3c) in a range of  $0.03 < I_{\text{rel}}^\mu < 0.1$  and with a finer step size of 0.05.

### C.0.1 Pileup mitigation techniques for the isolation of charged leptons

#### Implementation of the PUPPI algorithm in CMS

In CMS, various techniques for pileup mitigation have been developed, with the most prominent example the CHS algorithm (see Section 3.3.1). Several new techniques have recently been proposed including the pileup per particle identification (PUPPI) technique [190]. In contrast to CHS that only removes charged particles from pileup, PUPPI also calculates probabilities for each neutral particle of pileup or not origin and weights them based on that probability. Therefore, physics objects clustered from hadrons, such



**Figure C.2:** It is customary events to be divided in categories according to the number of jets and b-tagged jets using the naming convention of “n-jet-m-tag,” referring to events with n jets, m of which are b tagged. The reconstructed  $m_T$  distributions are shown for the 2j0t (a, b) and 2j1t (c, d) categories for events satisfying the selection as described in Ref. [TH13], separately for positively (a, c) and negatively (b, d) charged muons, using 2.3 fb<sup>-1</sup> of pp collisions at  $\sqrt{s} = 13$  TeV. The fit template for the QCD multijet component is derived from a sideband region in data. Only the statistical uncertainty is taken into account in the fit.

as jets, or reconstructed from particles assessed with PUPPI, such as  $\vec{p}_T^{\text{miss}}$  and the isolation of charged leptons, are expected to be less affected by harsher pileup conditions. Below, basic information on such techniques, applied to the lepton isolation, is given along with performance comparisons against the nominal isolation method (Eq. (3.2)). The impact of pileup mitigation techniques on object reconstruction performance, including the commissioning of PUPPI in the complete data sample in 2016, will be detailed in Ref. [299].

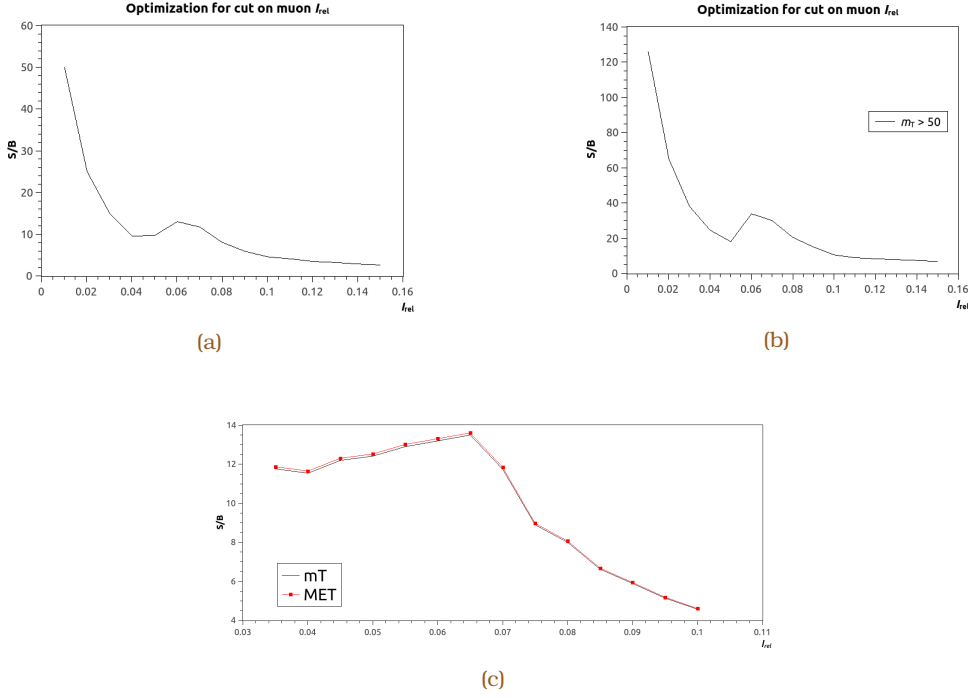


Figure C.3:  $S/B$  as a function of  $I_{\text{rel}}^{\mu}$  over the entire  $m_T > 0$  GeV variable region (a), in the extrapolated signal region of  $m_T > 50$  GeV (b), and over the entire  $p_T^{\text{miss}} > 0$  GeV region overlaid (c) [AN12].

For each particle, a local shape variable,  $a$ , is defined that is sensitive to differences between the hadronization of quarks and gluons and the soft diffuse radiation coming from pileup in the neighborhood of the particle, namely, particles from a parton shower are expected closer to each other, whereas pileup particles are distributed more homogeneously. The  $a$  variable can be then used as a weight to rescale the four-momenta of each particle at the particle level, i.e., before any clustering process is deployed. These weights thus serve as a metric of the probability that the particle originates from pileup interactions; a weight of unity (zero) is assigned to particles considered to originate from the leading (pileup) vertex. Furthermore, the algorithm flexibly allows combination with experimental information associated with vertex and timing performance, as described in the following.

In particular, the PUPPI algorithm is implemented using PF candidates in CMS. The variable  $a$  is separately defined in the central (within tracker acceptance) and forward (outside tracker acceptance) regions, and for any given PF candidate  $i$  is defined as

$$a_i = \log \sum_{j \neq i, \Delta R_{ij} < 0.4} \left( \frac{p_{Tj}}{\Delta R_{ij}} \right)^2 \begin{cases} \text{for } |\eta_i| < 2.5, & j \text{ are charged particles from leading vertex} \\ \text{for } |\eta_i| > 2.5, & j \text{ are all kinds of reconstructed particles} \end{cases} \quad \text{C.2}$$

where  $p_{Tj}$  is the transverse momentum of any neighboring particle  $j$ ,  $\Delta R_{ij}$  is the distance between particle  $i$  and  $j$  on the  $(\eta-\phi)$  plane, and the summation extends over all particles  $j$  within a cone of radius 0.4 around the PF candidate  $i$ . In the central region of the detector, i.e., up to  $|\eta| < 2.5$  where the tracking information is available, particles  $j$  are charged

particles from LV, while no such distinction can be made for particles in the forward region, i.e. in the  $|\eta| > 2.5$  region. In addition, charged PF candidates not associated with any PV are used in the calculation, if they satisfy a loose requirement—a multiple of the vertex reconstruction resolution—on  $d_z$ , where  $d_z$  is the distance between the track and LV in the  $z$  direction.

A  $\chi^2$  approximation

$$\chi_i^2 = \frac{(a_i - \bar{a}_{\text{PU}})^2}{\text{RMS}_{\text{PU}}^2}, \quad \text{C.3}$$

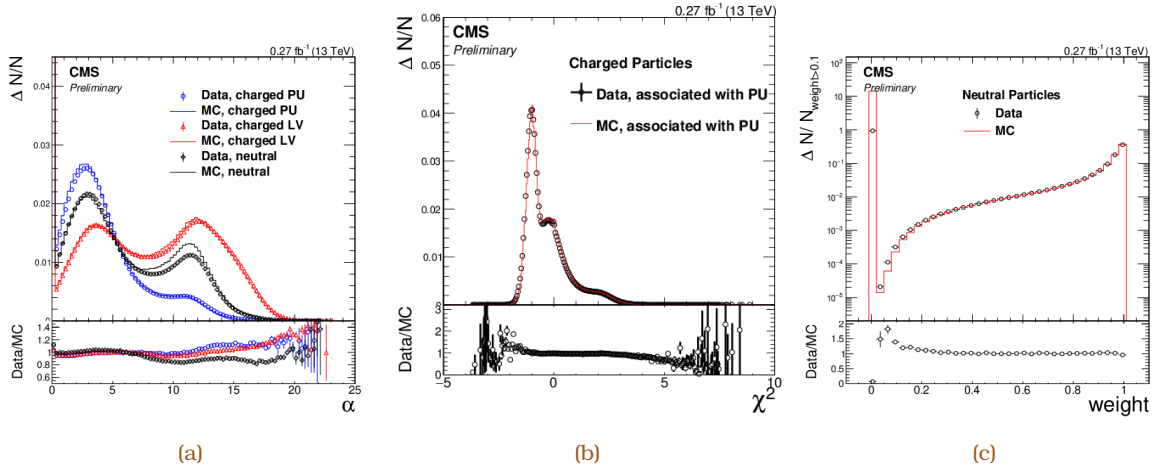
where  $\bar{a}_{\text{PU}}$  is the median value of the  $a_i$  distribution for all charged PF candidates associated with the pileup vertices in the event, and  $\text{RMS}_{\text{PU}}$  is the corresponding root-mean-square (RMS) of the  $a_i$  distribution, is used to determine the likelihood that a PF candidate originates from pileup. The numerator is defined to be sensitive to the direction of the deviation of  $a_i$  from  $\bar{a}_{\text{PU}}$ , given that  $a$  receives different values for particles from the leading vertex relative to particles of pileup origin. Within the tracker acceptance,  $\bar{a}_{\text{PU}}$  and  $\text{RMS}_{\text{PU}}$  are calculated using all charged pileup PF candidates, while in the forward region they are calculated either using all the particles in the event.

The  $\chi^2$  variable in Eq. (C.3) is finally transformed to a weight using the relation

$$w_i = F_{\chi^2, \text{dof}=1}(\chi_i^2), \quad \text{C.4}$$

where  $F_{\chi^2, \text{dof}=1}$  is the cumulative distribution function that approximates the  $\chi^2$  distribution with one degree of freedom of all PF candidates in the event. The weights range from zero to one, with zero indicating PF candidates originating from a pileup vertex, whereas PF candidates originating from a PV have values close to unity. Only charged PF candidates associated with the LV receive the value of unity. The  $w_i$  distribution is of interest above a minimum threshold, i.e., it is trimmed by rejecting particles with a probability of more than 99% to originate from pileup, and the minimum scaled  $p_T$  of neutral PF candidates is further required to be  $w_i \times p_{T,i} > (A + B \times N_{\text{PV}})$ , where  $N_{\text{PV}}$  is the reconstructed PV multiplicity. In this equation,  $A$  and  $B$  are adjustable parameters that depend on  $\eta$  and whose purpose is to ensure uniform jet energy response. An optimization of the tunable parameters is performed separately in the regions  $|\eta| < 2.5$ ,  $2.5 < |\eta| < 3$ , and  $|\eta| > 3$ , to achieve optimal performance related to jet kinematics and  $p_T^{\text{miss}}$  resolution.

Figure C.4 displays the metric, the weighting function, and the particle weights of PUPPI algorithm form implemented in CMS. Events are selected online based on jet triggers [163] and are required offline to be associated with a scalar jet  $p_T$  sum of at least 1 TeV. The separation power (Fig. C.4a) of the variable  $a$  between particles from the leading and pileup vertices reveals the majority of the charged particles from the latter having a value below about eight, with only a small fraction of (mostly high momentum) particles ending up to higher values. In addition, comparing the neutral to charged particle distributions, it becomes clear that the  $a$  distribution is qualitatively similar for these two types of PF candidates, confirming that  $\bar{a}_{\text{PU}}$  and  $\text{RMS}_{\text{PU}}$  computed exclusively based on charged particles can be used to compute weights for neutral particles. The  $\chi^2$  (Eq. (C.3)) is well modeled (Fig. C.4b) and is less affected by the mismodeling in simulation. The shape of



**Figure C.4:** The performance of the PUPPI algorithm has been tested using a pp data sample recorded in late 2016 (period “Run2016H”) at  $\sqrt{s} = 13$  TeV, corresponding to about  $0.27 \text{ fb}^{-1}$ , and it is compared to QCD multijet MC simulation [299].

the resulting PUPPI weight distribution (Fig. C.4c) is also reasonably well reproduced by simulation for particles with relatively high weights, i.e., those likely to be originated from LV, yet relevant for the subsequent clustering. The impact of the mismodeling at low  $w_i$  values, dominated by low  $p_T$  particles from pileup interactions, on final observables is small, since their weighted contribution on object reconstruction is negligible.

### Definition and performance of PUPPI muon isolation

The lepton isolation, i.e., the scalar  $p_T$  sum of particles surrounding a charged lepton, is affected by pileup collisions. One of the standard techniques used in CMS to mitigate pileup contamination is the  $\delta\beta$  corrected isolation (Eq. (3.2)) In this technique the  $\delta\beta$  correction subtracts the contamination on average, whereas PUPPI is able to correct on a particle-by-particle basis. A straightforward application of the PUPPI algorithm to the muon isolation can be defined as

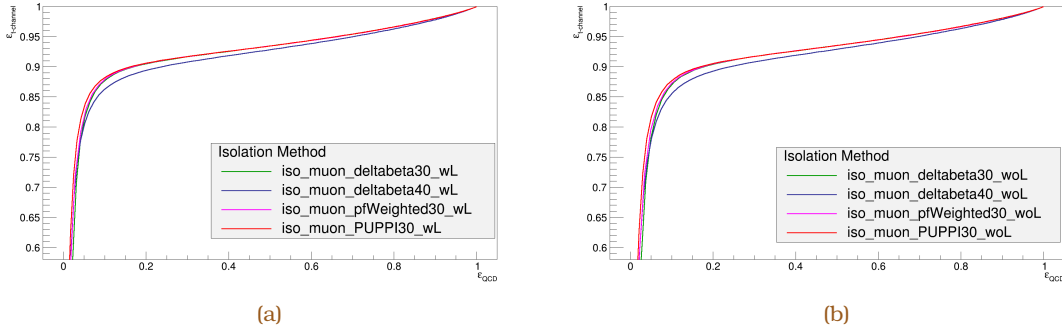
$$I_{\text{PUPPI}}^{\mu} = \sum_{\Delta R(i,j) < 0.3, 0.4} w^j p_T^j, \quad \text{C.5}$$

where  $p_T^j$  and  $w^j$  are the transverse momentum and PUPPI weight of the particle  $j$ , and the sum is considered for all particles in the cone with radius of 0.3 or 0.4 around muon in question. Although this treatment is favourable for muons from, e.g., semileptonic hadron decays, in the case of prompt muons the metric  $a$  calculates higher weights to all neutral particles surrounding muons, and hence introducing some pileup dependence on the isolation efficiency. An alternative type of isolation can be defined by excluding charged leptons from being of type  $j$  in Eq. (C.2). In following, the isolation defined with the former (default) PUPPI weight is referred to as “with lepton” (WL) and the latter as “without lepton” (WOL).

The performance of muon isolation is tested using MC simulation of single top  $t$ -

channel and QCD multijet events using a pileup distribution with a mean of about 20. The samples used for the study are described in Ref. [TH12]. The efficiency is calculated with respect to reconstructed muons with  $p_T$  greater than 20 GeV in  $|\eta| < 2.4$ . Muons are selected, if the relative isolation is smaller than a certain threshold, and the fraction of them passing the criteria is referred to as isolation efficiency for prompt muons and as fake rate for nonprompt muons.

Figure C.5 shows receiver operating characteristic (ROC) curves, i.e., the efficiency as a function of the fake rate, scanning the threshold on the isolation. Compared to  $\delta\beta$  correction, the PUPPI-WL isolation provides a higher rejection power, while the PUPPI-WOL isolation provides higher signal efficiency, as originally intended. In addition, and according to these findings, for the  $\delta\beta$  corrected isolation a cone size of  $\Delta R = 0.3$  seemed to be more promising, instead of the usual 0.4. However, a rough calculation based on the expected significance as a figure of merit, but modified [300] to include a systematic uncertainty of 50% in the QCD multijet background, revealed little-to-no gain in the measurement of the  $t$ -channel production cross section.



**Figure C.5:** Comparison of the pileup mitigation techniques in the case of muon inclusion (a) or exclusion (b) from the calculation of the charged isolation deposits in Eq. (C.5). The  $\delta\beta$  corrected isolation (Eq. (3.2)) is presented for two different cone sizes, i.e.,  $\Delta R = 0.3$  and 0.4. The combined isolation, defined as the mean of with- and without-lepton (WL and WOL) definitions of the PUPPI isolation, has been found to provide a slightly better performance along the usual operation region. A second algorithm has been developed, the so-called “PF reweighting,” that also rescales the four-momentum of neutral hadrons and photons with weights calculated based on their kinematics with respect to the PU- or non-PU-charged tracks in the event [AN12].





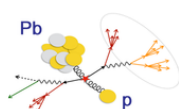
PHYSICAL REVIEW LETTERS

*moving physics forward*



Dear Sir or Madam,

We are pleased to inform you that the Letter



Observation of top quark production in proton-nucleus collisions

A.M. Sirunyan *et al.* (CMS Collaboration)  
Phys. Rev. Lett. **119**, 242001 (2017)

Published 14 December 2017

has been highlighted by the editors as an Editors' Suggestion. Publication of a Letter is already a considerable achievement, as *Physical Review Letters* accepts fewer than 1/4 of submissions, and is ranked first among physics and mathematics journals by the Google Scholar five-year h-index. A highlighted Letter has additional significance, because only about one Letter in six is highlighted as a Suggestion due to its particular importance, innovation, and broad appeal. Suggestions are downloaded twice as often as the average Letter, and are covered in the press substantially more often. If Suggestions were a separate publication, they would have an Impact Factor of 14. More information about our journal and its history can be found on our webpage [prl.aps.org](http://prl.aps.org).

Yours sincerely,

Hugues Chaté  
Editor  
*Physical Review Letters*

Michael Thoennessen  
Editor in Chief  
*Physical Review*

Physical Review Letters, 1 Research Road, Ridge, NY 11961-2701



- [TH1] A. M. Sirunyan *et al.* (CMS Collaboration), “Observation of top quark production in proton-nucleus collisions,” *Phys. Rev. Lett.*, vol. 119, p. 242001, 2017 doi:10.1103/PhysRevLett.119.242001. [Available Online]: <https://inspirehep.net/record/1624694>
- [TH2] CMS Collaboration, “CMS luminosity measurement using 2016 proton-nucleus collisions at nucleon-nucleon center-of-mass energy of 8.16 TeV,” CMS Physics Analysis Summary CMS-PAS-LUM-17-002, 2018. [Available Online]: <https://inspirehep.net/record/1681215>
- [TH3] CMS Collaboration, “CMS luminosity calibration for the pp reference run at  $\sqrt{s} = 5.02$  TeV,” CMS Physics Analysis Summary CMS-PAS-LUM-16-001, 2016. [Available Online]: <https://inspirehep.net/record/1501075>
- [TH4] A. M. Sirunyan *et al.* (CMS Collaboration), “Measurement of the inclusive  $t\bar{t}$  cross section in pp collisions at  $\sqrt{s} = 5.02$  TeV using final states with at least one charged lepton,” *JHEP*, vol. 1803, p. 115, 2018 doi:10.1007/JHEP03(2018)115. [Available Online]: <https://inspirehep.net/record/1635271>
- [TH5] G.K. Krintiras, “Constraining QCD multijet background in the  $t$ -channel single-top quark production at  $\sqrt{s} = 13$  TeV,” in *Proc. of the 8th International Workshop on Top Quark Physics (TOP2015)*, no. CMS-CR-2015-367, p. 059, Ischia, Italy, Sep 2015. [Available Online]: <https://inspirehep.net/record/1458379>
- [TH6] G.K. Krintiras, “Associated production of top quarks with the Higgs boson at  $\sqrt{s} = 13$  TeV,” in *Proc. of the 25th International Workshop on Deep-Inelastic Scattering and Related Topics (DIS 2017)*, no. CMS-CR-2017-176, Birmingham, UK, Apr 2017. [Available Online]: <https://inspirehep.net/record/1643827>
- [TH7] G.K. Krintiras, “Results on TOP physics from CMS,” in *Proc. of the International Workshop on Future Linear Collider (LCWS2017)*, no. CMS-CR-2018-013, Strasbourg, France, Oct 2017. [Available Online]: <https://inspirehep.net/record/1650460>
- [TH8] CMS Collaboration, “CMS luminosity measurement for the 2017 data-taking period at  $\sqrt{s} = 13$  TeV,” CMS Physics Analysis Summary CMS-PAS-LUM-17-004, 2018. [Available Online]: <https://inspirehep.net/record/1677076>
- [TH9] G.K. Krintiras, “Study of the top quark production in complementary phase space regions and impact on PDFs in CMS,” in *Proc. of the 29th Rencontres de Blois on Particle Physics and Cosmology*, no. CMS-CR-2017-345, Blois, France, Jun 2017. [Available Online]: <https://inspirehep.net/record/1643826>
- [TH10] G.K. Krintiras, “Observation of top quark production in proton-nucleus collision,” in *Proc. of the 10th International Workshop on Top Quark Physics*

- (TOP2017), no. CMS-CR-2017-434, Braga, Portugal, Sep 2017. [Available Online]: <https://inspirehep.net/record/1643825>
- [TH11] G.K. Krintiras, “Observation of top quark production in proton-nucleus collision,” in *Proc. of the 27th International Conference on Ultrarelativistic Nucleus-Nucleus Collisions (QM2018)*, Venice, Italy, May 2018. [Available Online]: [https://indico.cern.ch/event/656452/contributions/2871100/attachments/1648771/2636098/QM\\_Poster\\_GKKrintiras.pdf](https://indico.cern.ch/event/656452/contributions/2871100/attachments/1648771/2636098/QM_Poster_GKKrintiras.pdf)
- [TH12] CMS Collaboration, “Measurement of the  $t$ -channel single top quark cross section at 13TeV,” CMS Physics Analysis Summary CMS-PAS-TOP-15-004, 2015. [Available Online]: <https://inspirehep.net/record/1393357>
- [TH13] A. M. Sirunyan *et al.* (CMS Collaboration), “Cross section measurement of  $t$ -channel single top quark production in pp collisions at  $\sqrt{s} = 13\text{TeV}$ ,” *Phys. Lett. B*, vol. 72, p. 752, 2017 doi:10.1016/j.physletb.2017.07.047. [Available Online]: <https://inspirehep.net/record/1489193>
- [TH14] CMS Collaboration, “CMS luminosity measurements for the 2016 data taking period,” CMS Physics Analysis Summary CMS-PAS-LUM-17-001, 2017. [Available Online]: <https://inspirehep.net/record/1519242>
- [TH15] CMS Collaboration, “First measurement of the top quark pair production cross section in proton-proton collisions at  $\sqrt{s} = 5.02\text{TeV}$ ,” CMS Physics Analysis Summary CMS-PAS-TOP-16-015, 2015, superseded in the paper *JHEP 03 (2018) 115*. [Available Online]: <https://inspirehep.net/record/1471008>

- [AN1] CMS Collaboration, “Measurement of the top quark pair production cross section in proton-proton collisions at  $\sqrt{s} = 5.02$  TeV using final states with exactly one isolated lepton,” CMS Internal Note, 2016 [CMS AN-2016/320](#)
- [AN2] CMS Collaboration, “First measurement of the top quark pair cross section using  $\ell$ +jets in pPb collisions at  $\sqrt{s_{NN}} = 8.16$  TeV,” CMS Internal Note, 2017 [CMS AN-2017/043](#)
- [AN3] CMS Collaboration, “Inclusive  $t\bar{t}$  cross section in the dilepton final state at  $\sqrt{s} = 5.02$  TeV,” CMS Internal Note, 2016 [CMS AN-2016/098](#)
- [AN4] CMS Collaboration, “Measurement of the top quark pair production cross section in the dilepton channel in proton-proton collisions at  $\sqrt{s} = 5.02$  TeV,” CMS Internal Note, 2016 [CMS AN-2016/369](#)
- [AN5] CMS Collaboration, “Muon performance studies in 2016 pPb data,” CMS Internal Note, 2017 [CMS AN-2017/137](#)
- [AN6] CMS Collaboration, “Beamspot determination in 2015 data,” CMS Internal Note, 2015 [CMS AN-2015/195](#)
- [AN7] CMS Collaboration, “Electron reconstruction and identification for PbPb and reference pp collisions at 5.02 TeV,” CMS Internal Note, 2016 [CMS AN-2016/110](#)
- [AN8] CMS Collaboration, “Dijet pseudorapidity and mass measurement in pp and pPb collisions at  $\sqrt{s_{NN}} = 5.02$  TeV with the CMS detector,” CMS Internal Note, 2016 [CMS AN-2016/087](#)
- [AN9] CMS Collaboration, “Combination of the measurements of the top quark pair production cross section in pp collisions at  $\sqrt{s} = 5.02$  TeV and impact on PDF constraints,” CMS Internal Note, 2016 [CMS AN-2016/358](#)
- [AN10] CMS Collaboration, “CMS luminosity calibration for the pp reference run at 5.02 TeV in 2017,” CMS Internal Note, 2018 [CMS AN-2018/186](#)
- [AN11] CMS Collaboration, “Luminosity normalization precision and stability of CMS data taken in 2015 and 2016,” CMS Internal Note, 2017 [CMS AN-2017/240](#)
- [AN12] CMS Collaboration, “Muons for single top  $t$ -channel cross section measurements at  $\sqrt{s} = 13$  TeV, selection optimization and scale factors,” CMS Internal Note, 2016 [CMS AN-2016/112](#)



- [1] S. Herb *et al.*, “Observation of a dimuon resonance at 9.5 GeV in 400-GeV proton-nucleus collisions,” *Phys. Rev. Lett.*, vol. 39, p. 252, 1977 doi:10.1103/PhysRevLett.39.252. [Available Online]: <https://inspirehep.net/record/120368>
- [2] F. W. Abe *et al.* (CDF Collaboration), “Evidence for top quark production in  $p\bar{p}$  collisions at  $\sqrt{s} = 1.8\text{TeV}$ ,” *Phys. Rev. Lett.*, vol. 73, p. 225, 1994 doi:10.1103/PhysRevLett.73.225. [Available Online]: <https://inspirehep.net/record/373362>
- [3] F. W. Abe *et al.* (CDF Collaboration), “Evidence for top quark production in  $p\bar{p}$  collisions at  $\sqrt{s} = 1.8\text{TeV}$ ,” *Phys. Rev. D*, vol. 50, p. 2966, 1994 doi:10.1103/PhysRevD.50.2966. [Available Online]: <https://inspirehep.net/record/372952>
- [4] F. W. Abe *et al.* (CDF Collaboration), “Observation of top quark production in  $p\bar{p}$  collisions with the Collider Detector at Fermilab,” *Phys. Rev. Lett.*, vol. 74, p. 2626, 1995 doi:10.1103/PhysRevLett.74.2626. [Available Online]: <https://inspirehep.net/record/393084>
- [5] S. Abachi *et al.* (DØ Collaboration), “Observation of the top quark,” *Phys. Rev. Lett.*, vol. 74, p. 2632, 1995 doi:10.1103/PhysRevD.50.2966. [Available Online]: <https://inspirehep.net/record/393099>
- [6] Bruce, R. and others, “New physics searches with heavy ion collisions at the LHC,” 2018, submitted as input to the update of the European Particle Physics Strategy (EPPS). [Available Online]: <https://inspirehep.net/record/1709994>
- [7] G. Arnison *et al.* (UA1 Collaboration), “Associated production of an isolated, large-transverse-momentum lepton (electron or muon), and two jets at the CERN  $p\bar{p}$  collider,” *Phys. Lett. B*, vol. 147, p. 493, 1984 doi:10.1016/0370-2693(84)91410-2. [Available Online]: <https://inspirehep.net/record/205116>
- [8] ALEPH and DELPHI and L3 and OPAL and SLD Collaborations and LEP Electroweak Working Group, and SLD Electroweak and Heavy Flavour Groups, “Precision electroweak measurements on the Z,” *Phys. Rept.*, vol. 427, p. 257, 2006 doi:10.1016/j.physrep.2005.12.006. [Available Online]: <https://inspirehep.net/record/691576>
- [9] J. Haller *et al.* (Gfitter Group), “Update of the global electroweak fit and constraints on two-Higgs-doublet models,” *Eur. Phys. J. C*, vol. 78, p. 675, 2018 doi:10.1140/epjc/s10052-018-6131-3. [Available Online]: <https://inspirehep.net/record/1658767>

- [10] F. Englert and R. Brout, “Broken symmetry and the mass of gauge vector mesons,” *Phys. Rev. Lett.*, vol. 13, p. 321, 1964 doi:10.1103/PhysRevLett.13.321. [Available Online]: <https://dx.doi.org/10.1103/PhysRevLett.13.321>
- [11] P. W. Higgs, “Broken symmetries, massless particles and gauge fields,” *Phys. Lett.*, vol. 12, p. 132, 1964 doi:10.1016/0031-9163(64)91136-9. [Available Online]: [https://dx.doi.org/10.1016/0031-9163\(64\)91136-9](https://dx.doi.org/10.1016/0031-9163(64)91136-9)
- [12] P. W. Higgs, “Broken symmetries and the masses of gauge bosons,” *Phys. Rev. Lett.*, vol. 13, p. 508, 1964 doi:10.1103/PhysRevLett.13.508. [Available Online]: <https://dx.doi.org/10.1103/PhysRevLett.13.508>
- [13] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, “Global conservation laws and massless particles,” *Phys. Rev. Lett.*, vol. 13, p. 585, 1964 doi:10.1103/PhysRevLett.13.585. [Available Online]: <https://dx.doi.org/10.1103/PhysRevLett.13.585>
- [14] P. W. Higgs, “Spontaneous symmetry breakdown without massless bosons,” *Phys. Rev.*, vol. 145, p. 1156, 1966 doi:10.1103/PhysRev.145.1156. [Available Online]: <https://dx.doi.org/10.1103/PhysRev.145.1156>
- [15] T. W. B. Kibble, “Symmetry breaking in non-Abelian gauge theories,” *Phys. Rev.*, vol. 155, p. 1554, 1967 doi:10.1103/PhysRev.155.1554. [Available Online]: <https://dx.doi.org/10.1103/PhysRev.155.1554>
- [16] M. Tanabashi *et al.* (Particle Data Group), “The review of particle physics (2018),” *Phys. Rev. D*, vol. 98, p. 030001, 2018 doi:10.1103/PhysRevD.98.030001. [Available Online]: <https://pdg.lbl.gov>
- [17] G. Aad *et al.* (ATLAS Collaboration), “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC,” *Phys. Lett. B*, vol. 716, p. 1, 2012 doi:10.1016/j.physletb.2012.08.020. [Available Online]: <https://inspirehep.net/record/1124337>
- [18] S. Chatrchyan *et al.* (CMS Collaboration), “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC,” *Phys. Lett. B*, vol. 716, p. 30, 2012 doi:10.1016/j.physletb.2012.08.021. [Available Online]: <https://inspirehep.net/record/1124338>
- [19] S. Chatrchyan *et al.* (CMS Collaboration), “Observation of a new boson with mass near 125 GeV in pp collisions at  $\sqrt{s} = 7$  and 8 TeV,” *JHEP*, vol. 06, p. 081, 2013 doi:10.1007/JHEP06(2013)081. [Available Online]: <https://inspirehep.net/record/1224273>
- [20] CMS Collaboration, “Combined measurements of Higgs boson couplings in proton-proton collisions at  $\sqrt{s} = 13$  TeV,” 2018, submitted to *Eur. Phys. J. C*. [Available Online]: <https://inspirehep.net/record/1696607>



- [21] G. Aad *et al.* (ATLAS Collaboration), “Measurement of the Higgs boson mass from the  $H \rightarrow \gamma\gamma$  and  $H \rightarrow ZZ^* \rightarrow 4\ell$  channels with the ATLAS detector using  $25\text{fb}^{-1}$  of pp collision data,” *Phys. Rev. D*, vol. 90, p. 052004, 2014 doi:10.1103/PhysRevD.90.052004. [Available Online]: <https://inspirehep.net/record/1300650>
- [22] V. Khachatryan *et al.* (CMS Collaboration), “Precise determination of the mass of the Higgs boson and tests of compatibility of its couplings with the standard model predictions using proton collisions at 7 and 8TeV,” *Eur. Phys. J. C*, vol. 75, p. 212, 2015 doi:10.1140/epjc/s10052-015-3351-7. [Available Online]: <https://inspirehep.net/record/1335811>
- [23] J. Fumagalli, S. Mooij, and M. Postma, “Unitarity and predictiveness in new Higgs inflation,” *JHEP*, vol. 1803, p. 038, 2018 doi:10.1007/JHEP03(2018)038. [Available Online]: <https://inspirehep.net/record/1638325>
- [24] ATLAS and CMS Collaborations, “Combined measurement of the Higgs boson mass in pp collisions at  $\sqrt{s} = 7$  and 8TeV with the ATLAS and CMS experiments,” *Phys. Rev. Lett.*, vol. 114, p. 191803, 2015 doi:10.1103/PhysRevLett.114.191803. [Available Online]: <https://inspirehep.net/record/1356276>
- [25] ATLAS and CDF and CMS and DØ Collaborations, “First combination of Tevatron and LHC measurements of the top quark mass,” 2014, ATLAS-CONF-2014-008, CDF-NOTE-11071, CMS-PAS-TOP-13-014, DØ-NOTE-6416, FERMILAB-TM-2582-E. [Available Online]: <https://inspirehep.net/record/1286320>
- [26] D. Buttazzo *et al.*, “Investigating the near-criticality of the Higgs boson,” *JHEP*, vol. 1312, p. 089, 2013 doi:10.1007/JHEP12(2013)089. [Available Online]: <https://inspirehep.net/record/1242456>
- [27] A. Salvio, A. Strumia, N. Tetradis, and A. Urbano, “On gravitational and thermal corrections to vacuum decay,” *JHEP*, vol. 1609, p. 054, 2016 doi:10.1007/JHEP09(2016)054. [Available Online]: <https://inspirehep.net/record/1479980>
- [28] U. Husemann, “Top quark physics: Status and prospects,” *Prog. Part. Nucl. Phys.*, vol. 95, p. 48, 2018 doi:10.1016/j.pnpnp.2017.03.002. [Available Online]: <https://inspirehep.net/record/1589840>
- [29] J. Gao and A. Papanastasiou, “Top quark pair production and decay at high precision,” *Phys. Rev. D*, vol. 96, p. 051501, 2017 doi:10.1103/PhysRevD.96.051501. [Available Online]: <https://inspirehep.net/record/1601291>
- [30] R. Poncelet *et al.*, “NNLO QCD top quark pair production and decay,” in *Proc. of the 11th International Workshop on Top Quark Physics (TOP2018)*, Bad Neuenahr, Germany, Sep. 2018.

- [31] D. Pagani, I. Tsirikos, and M. Zaro, “The impact of the photon PDF and electroweak corrections on  $t\bar{t}$  distributions,” *Eur. Phys. J. C*, vol. 76, p. 479, 2016 doi:10.1140/epjc/s10052-016-4318-z. [Available Online]: <https://inspirehep.net/record/1468062>
- [32] S. Alekhin, J. Blümlein, S. Moch, and R. Placakyte, “Parton distribution functions,  $\alpha_s$ , and heavy quark masses for LHC Run 2,” *Phys. Rev. D*, vol. 96, p. 014011, 2017 doi:10.1103/PhysRevD.96.014011. [Available Online]: <https://inspirehep.net/record/1510074>
- [33] S. Dulat *et al.*, “New parton distribution functions from a global analysis of quantum chromodynamics,” *Phys. Rev. D*, vol. 93, p. 033006, 2016 doi:10.1103/PhysRevD.93.033006. [Available Online]: <https://inspirehep.net/record/1377752>
- [34] L. A. Harland-Lang, A. D. Martin, P. Motylinski, and R. S. Thorne, “Parton distributions in the LHC era: MMHT 2014 PDFs,” *Eur. Phys. J. C*, vol. 75, p. 204, 2015 doi:10.1140/epjc/s10052-015-3397-6. [Available Online]: <https://inspirehep.net/record/1334137>
- [35] R. D. Ball *et al.* (NNPDF Collaboration), “Parton distributions for the LHC Run 2,” *JHEP*, vol. 04, p. 040, 2015 doi:10.1007/JHEP04(2015)040. [Available Online]: <https://inspirehep.net/record/1325552>
- [36] ZEUS and H1 Collaborations, “Combination of measurements of inclusive deep inelastic  $e^\pm p$  scattering cross sections and QCD analysis of HERA data,” *Eur. Phys. J. C*, vol. 75, p. 580, 2015 doi:10.1140/epjc/s10052-015-3710-4. [Available Online]: <https://inspirehep.net/record/1377206>
- [37] P. Jimenez-Delgado and E. Reya, “Delineating parton distributions and the strong coupling,” *Phys. Rev. D*, vol. 89, p. 074049, 2014 doi:10.1103/PhysRevD.89.074049. [Available Online]: <https://inspirehep.net/record/1284991>
- [38] 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,” *JHEP*, vol. 07, p. 079, 2014 doi:10.1007/JHEP07(2014)079. [Available Online]: <https://inspirehep.net/record/1293923>
- [39] S. Frixione, P. Nason, and C. Oleari, “Matching NLO QCD computations with parton shower simulations: The POWHEG method,” *JHEP*, vol. 11, p. 070, 2007 doi:10.1088/1126-6708/2007/11/070. [Available Online]: <https://inspirehep.net/record/760769>
- [40] S. Alioli, P. Nason, C. Oleari, and E. Re, “A general framework for implementing NLO calculations in shower Monte Carlo programs: The POWHEG BOX,” *JHEP*, vol. 06, p. 043, 2010 doi:10.1007/JHEP06(2010)043. [Available Online]: <https://inspirehep.net/record/845712>

- [41] T. S. Gleisberg *et al.*, “Event generation with SHERPA 1.1,” *JHEP*, vol. 0902, p. 007, 2009 doi:10.1088/1126-6708/2009/02/007. [Available Online]: <https://inspirehep.net/record/803708>
- [42] J. Alwall *et al.*, “Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions,” *Eur. Phys. J. C*, vol. 53, p. 473, 2008 doi:10.1140/epjc/s10052-007-0490-5. [Available Online]: <https://inspirehep.net/record/753397>
- [43] R. Frederix and S. Frixione, “Merging meets matching in MC@NLO,” *JHEP*, vol. 12, p. 061, 2012 doi:10.1007/JHEP12(2012)061. [Available Online]: <https://inspirehep.net/record/1188307>
- [44] K. Cormier *et al.*, “Parton shower and matching uncertainties in top quark pair production with Herwig 7,” 2018, CERN-TH-2018-219. [Available Online]: <https://inspirehep.net/record/1698475>
- [45] T. Ježo *et al.*, “An NLO+PS generator for  $t\bar{t}$  and  $Wt$  production and decay including nonresonant and interference effects,” *Eur. Phys. J. C*, vol. 76, p. 691, 2016 doi:10.1140/epjc/s10052-016-4538-2. [Available Online]: <https://inspirehep.net/record/1476366>
- [46] CMS Collaboration, “Investigations of the impact of the parton shower tuning in PYTHIA8 in the modeling of  $t\bar{t}$  at  $\sqrt{s} = 8$  and 13 TeV,” CMS Physics Analysis Summary CMS-PAS-TOP-16-021, 2016. [Available Online]: <https://inspirehep.net/record/1500109>
- [47] CMS Collaboration, “Study of the underlying event in top quark pair production in pp collisions at 13 TeV,” 2018, submitted to *Eur. Phys. J. C*. [Available Online]: <https://inspirehep.net/record/1663395>
- [48] CMS Collaboration, “Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements,” CMS Physics Analysis Summary CMS-PAS-GEN-17-001, 2016. [Available Online]: <https://inspirehep.net/record/1691618>
- [49] T. Sjöstrand, S. Mrenna, and P. Skands, “PYTHIA 6.4 physics and manual,” *JHEP*, vol. 0605, p. 026, 2006 doi:10.1088/1126-6708/2006/05/026. [Available Online]: <https://inspirehep.net/record/712925>
- [50] T. Sjöstrand, S. Mrenna, and P. Skands, “A Brief Introduction to PYTHIA 8.1,” *Comput. Phys. Commun.*, vol. 178, p. 852, 2008 doi:10.1016/j.cpc.2008.01.036. [Available Online]: <https://inspirehep.net/record/764903>
- [51] T. Sjöstrand *et al.*, “An introduction to PYTHIA 8.2,” *Comput. Phys. Commun.*, vol. 191, p. 159, 2015 doi:10.1016/j.cpc.2015.01.024. [Available Online]: <https://inspirehep.net/record/1321709>

- [52] M. Bähr *et al.*, “HERWIG++ physics and manual,” *Eur. Phys. J. C*, vol. 58, p. 639, 2008 doi:10.1140/epjc/s10052-008-0798-9. [Available Online]: <https://inspirehep.net/record/780833>
- [53] J. Adam *et al.* (ALICE Collaboration), “Enhanced production of multistrange hadrons in high-multiplicity proton-proton collisions,” *Nature Phys.*, vol. 13, p. 535, 2017 doi:10.1038/nphys4111. [Available Online]: <https://inspirehep.net/record/1471838>
- [54] V. Khachatryan *et al.* (CMS Collaboration), “Observation of long-range near-side angular correlations in proton-proton collisions at the LHC,” *JHEP*, vol. 09, p. 091, 2010 doi:10.1007/JHEP09(2010)091. [Available Online]: <https://inspirehep.net/record/870473>
- [55] S. Chatrchyan *et al.* (CMS Collaboration), “Observation of long-range near-side angular correlations in proton-lead collisions at the LHC,” *Phys. Lett. B*, vol. 718, p. 795, 2013 doi:10.1016/j.physletb.2012.11.025. [Available Online]: <https://inspirehep.net/record/1191899>
- [56] V. Khachatryan *et al.* (CMS Collaboration), “Measurement of long-range near-side two-particle angular correlations in pp collisions at  $\sqrt{s} = 13\text{ TeV}$ ,” *Phys. Rev. Lett.*, vol. 116, p. 172302, 2016 doi:10.1103/PhysRevLett.116.172302. [Available Online]: <https://inspirehep.net/record/1397173>
- [57] S. Frixione, P. Nason, and G. Ridolfi, “A positive-weight next-to-leading-order Monte Carlo for heavy flavor hadroproduction,” *JHEP*, vol. 0709, p. 126, 2007 doi:10.1088/1126-6708/2007/09/126. [Available Online]: <https://inspirehep.net/record/756360>
- [58] V. Khachatryan *et al.* (CMS Collaboration), “Event generator tunes obtained from underlying-event and multiparton-scattering measurements,” *Eur. Phys. J. C*, vol. 76, p. 155, 2016 doi:10.1140/epjc/s10052-016-3988-x. [Available Online]: <https://inspirehep.net/record/1407839>
- [59] P. Skands, S. Carrazza, and J. Rojo, “Tuning PYTHIA 8.1: The Monash 2013 tune,” *Eur. Phys. J. C*, vol. 74, p. 3024, 2014 doi:10.1140/epjc/s10052-014-3024-y. [Available Online]: <https://inspirehep.net/record/1291955>
- [60] R. D. Ball *et al.* (NNPDF Collaboration), “Parton distributions with QED corrections,” *Nucl. Phys. B*, vol. 0877, p. 290, 2013 doi:10.1016/j.nuclphysb.2013.10.010. [Available Online]: <https://inspirehep.net/record/1246369>
- [61] R. D. Ball *et al.* (NNPDF Collaboration), “Unbiased global determination of parton distributions and their uncertainties at NNLO and at LO,” *Nucl. Phys. B*, vol. 855, p. 153, 2012 doi:10.1016/j.nuclphysb.2011.09.024. [Available Online]: <https://inspirehep.net/record/918528>

- [62] M. Seymour and A. Siodmok, “Constraining MPI models using  $\sigma_{\text{eff}}$  and recent Tevatron and LHC underlying-event data,” *JHEP*, vol. 1310, p. 113, 2013 doi:10.1007/JHEP10(2013)113. [Available Online]: <https://inspirehep.net/record/1243440>
- [63] J. Pumplin *et al.*, “New generation of parton distributions with uncertainties from global QCD analysis,” *JHEP*, vol. 0207, p. 012, 2002 doi:10.1088/1126-6708/2002/07/012. [Available Online]: <https://inspirehep.net/record/581996>
- [64] D. Bandurin *et al.*, “Review of physics results from the Tevatron,” *Int. J. Mod. Phys. A*, vol. 30, p. 01541001, 2015 doi:10.1142/S0217751X15410018. [Available Online]: <https://inspirehep.net/record/1317643>
- [65] CMS Collaboration, “CMS luminosity based on the pixel-cluster counting: Summer 2013 update,” CMS Physics Analysis Summary CMS-PAS-LUM-13-001, 2013. [Available Online]: <https://inspirehep.net/record/1260876>
- [66] J. Boyd and C. Schwick, “Luminosity measurements and systematics,” in *Proc. of the 8th Evian Workshop*, Evian Les Bains, France, Dec. 2017. [Available Online]: <https://indico.cern.ch/event/663598>
- [67] M. Pojer *et al.*, LHC Coordination information page. [Available Online]: <https://twiki.cern.ch/twiki/bin/viewauth/LhcMachine/LhcCoordinationMain>
- [68] CMS Collaboration, “Events recorded by the CMS detector during the proton-lead collision run of 2016,” CMS Collection CMS-PHO-EVENTS-2016-009-19, 2016. [Available Online]: <https://cds.cern.ch/record/2235235>
- [69] T. Sakuma and T. McCauley, “Detector and event visualization with SketchUp at the CMS experiment,” *J. Phys. Conf. Ser.*, vol. 513, p. 022032, 2014 doi:10.1088/1742-6596/513/2/022032. [Available Online]: <https://inspirehep.net/record/1265221>
- [70] T. McCauley, “A browser-based event display for the CMS experiment at the LHC using WebGL,” *J. Phys. Conf. Ser.*, vol. 898, p. 072030, 2017 doi:10.1088/1742-6596/898/7/072030. [Available Online]: <https://inspirehep.net/record/1638558>
- [71] T. Carli, K. Rabbertz, and S. Schumann, “Studies of quantum chromodynamics at the LHC,” in *The Large Hadron Collider*, 1st ed., T. Schörner-Sadenius, Ed. Springer International Publishing, 2015, p. 139. ISBN 978-3-319-15001-7 doi:10.1007/978-3-319-15001-7\_5
- [72] S. Chatrchyan *et al.* (CMS Collaboration), “The CMS experiment at the CERN LHC,” *JINST*, vol. 3, p. S08004, 2008 doi:10.1088/1748-0221/3/08/S08004. [Available Online]: <https://inspirehep.net/record/796887>
- [73] P. Grafström and W. Kozanecki, “Luminosity determination at proton colliders,” *Prog. Part. Nucl. Phys.*, vol. 81, p. 97, 2015 doi:10.1016/j.pnpnp.2014.11.002. [Available Online]: <https://inspirehep.net/record/1357777>

- [74] CMS Collaboration, “Measurement of the  $t\bar{t}$  production cross section, the top quark mass, and the strong coupling constant using dilepton events in pp collisions at  $\sqrt{s} = 13\text{TeV}$ ,” 2018, submitted to *Eur. Phys. J. C*. [Available Online]: <https://inspirehep.net/record/1711626>
- [75] A. M. Sirunyan *et al.* (CMS Collaboration), “Measurement of the  $t\bar{t}$  production cross section using events with one lepton and at least one jet in pp collisions at  $\sqrt{s} = 13\text{TeV}$ ,” *JHEP*, vol. 09, p. 051, 2017 doi:10.1007/JHEP09(2017)051. [Available Online]: <https://inspirehep.net/record/1510260>
- [76] F. Zimmermann, “Luminosity limitations at hadron colliders,” in *Proc. of the 18th International Conference on High-Energy Accelerators (HEACC 2001)*, no. CERN-SL-2001-009-AP, pp. TU-08, Tsukuba, Japan, Mar. 2001. [Available Online]: <https://inspirehep.net/record/557207>
- [77] T. Pettersson and P. Lefèvre, Eds., *The Large Hadron Collider: Conceptual design*. CERN, 1995. [Available Online]: <https://inspirehep.net/record/402898>
- [78] The High Luminosity LHC (HL-LHC) project. [Available Online]: <https://hilumilhc.web.cern.ch>
- [79] eRHIC: An electron-ion collider at BNL. [Available Online]: [https://wiki.bnl.gov/eic/index.php/Main\\_Page](https://wiki.bnl.gov/eic/index.php/Main_Page)
- [80] The Future Circular Collider (FCC) study. [Available Online]: <https://fcc.web.cern.ch/Pages/default.aspx>
- [81] Summaries of CMS cross section measurements. [Available Online]: <https://cern.ch/go/pNj7>
- [82] G.K. Krintiras, “Higher-order azimuthal anisotropy of  $\Lambda + \bar{\Lambda}$  hyperons in PbPb collisions at  $\sqrt{s_{NN}} = 2.76\text{TeV}$  measured by ALICE at LHC,” Master’s thesis, University of Amsterdam, 2015. [Available Online]: <https://esc.fnwi.uva.nl/thesis/centraal/files/f1500767949.pdf>
- [83] F. Karsch, “The phase transition to the quark gluon plasma: Recent results from lattice calculations,” *Nucl. Phys. A*, vol. 590, p. 367, 1995 doi:10.1016/0375-9474(95)00248-Y. [Available Online]: <https://inspirehep.net/record/393363>
- [84] J. Bjorken, “Highly relativistic nucleus-nucleus collisions: The central rapidity region,” *Phys. Rev. D*, vol. 27, p. 140, 1983 doi:10.1103/PhysRevD.27.140. [Available Online]: <https://inspirehep.net/record/178853>
- [85] J. J. Aubert *et al.* (EMC Collaboration), “The ratio of the nucleon structure functions  $F_2^N$  for iron and deuterium,” *Phys. Lett. B*, vol. 123, p. 275, 1983 doi:10.1016/0370-2693(83)90437-9. [Available Online]: <https://inspirehep.net/record/188925>

- [86] K. J. Eskola, H. Paukkunen, and C. A. Salgado, “EPS09: A new generation of NLO and LO nuclear parton distribution functions,” *JHEP*, vol. 04, p. 065, 2009 doi:10.1088/1126-6708/2009/04/065. [Available Online]: <https://inspirehep.net/record/814088>
- [87] D. d’Enterria, K. Krajczar, and H. Paukkunen, “Top quark production in proton-nucleus and nucleus-nucleus collisions at LHC energies and beyond,” *Phys. Lett. B*, vol. 746, p. 64, 2015 doi:10.1016/j.physletb.2015.04.044. [Available Online]: <https://inspirehep.net/record/1340855>
- [88] K. J. Eskola, P. Paakkinen, H. Paukkunen, and C. A. Salgado, “EPPS16: Nuclear parton distributions with LHC data,” *Eur. Phys. J. C*, vol. 77, p. 163, 2017 doi:10.1140/epjc/s10052-017-4725-9. [Available Online]: <https://inspirehep.net/record/1504944>
- [89] M. Hirai, S. Kumano, and T.-H. Nagai, “Determination of nuclear parton distribution functions and their uncertainties at next-to-leading order,” *Phys. Rev. C*, vol. 76, p. 065207, 2007 doi:10.1103/PhysRevC.76.065207. [Available Online]: <https://inspirehep.net/record/761288>
- [90] K. Aleksander, J. P. Lansberg, I. Schienbein, and H.-S. Shao, “Gluon shadowing in heavy-flavor production at the LHC,” *Phys. Rev. Lett.*, vol. 121, p. 052004, 2018 doi:10.1103/PhysRevLett.121.052004. [Available Online]: <https://inspirehep.net/record/1644102>
- [91] A. M. Sirunyan *et al.* (CMS Collaboration), “Constraining gluon distributions in nuclei using dijets in proton-proton and proton-lead collisions at  $\sqrt{s_{NN}} = 5.02$  TeV,” *Phys. Rev. Lett.*, vol. 121, p. 062002, 2018 doi:10.1103/PhysRevLett.121.062002. [Available Online]: <https://inspirehep.net/record/1672941>
- [92] K. J. Eskola, P. Paakkinen, and H. Hannu Paukkunen, “Impact of CMS dijets in 5.02 TeV pPb and pp collisions on EPPS16 nuclear PDFs,” in *Proc. of the International Conference on Hard and Electromagnetic Probes of High-Energy Nuclear Collisions (Hard Probes 2018)*, Aix-Les-Bains, France, Sep. 2018. [Available Online]: <https://inspirehep.net/record/1708894>
- [93] L. Apolinário *et al.*, “Probing the time structure of the quark-gluon plasma with top quarks,” *Phys. Rev. Lett.*, vol. 120, p. 232301, 2018 doi:10.1103/PhysRevLett.120.232301. [Available Online]: <https://inspirehep.net/record/1635270>
- [94] A. M. Sirunyan *et al.* (CMS Collaboration), “Measurement of the top quark mass with lepton+jets final states using pp collisions at  $\sqrt{s} = 13$  TeV,” *Eur. Phys. J. C*, vol. 891, p. 78, 2018 doi:10.1140/epjc/s10052-018-6332-9. [Available Online]: <https://inspirehep.net/record/1671499>

- [95] S. Chatrchyan *et al.* (CMS Collaboration), “Observation and studies of jet quenching in PbPb collisions at  $\sqrt{s_{\text{NN}}} = 2.76\text{ TeV}$ ,” *Phys. Rev. C*, vol. 84, p. 024906, 2011 doi:10.1103/PhysRevC.84.024906. [Available Online]: <https://inspirehep.net/record/889010>
- [96] CMS Collaboration, “Measurement of nuclear modification factors of  $Y(1S)$ ,  $Y(2S)$ , and  $Y(3S)$  mesons in PbPb collisions at  $\sqrt{s_{\text{NN}}} = 5.02\text{ TeV}$ ,” 2018, submitted to *Phys. Lett. B*. [Available Online]: <https://inspirehep.net/record/1674529>
- [97] H. Wiedemann, *Particle Accelerator Physics*, 4th ed. Springer International Publishing, 2015. ISBN 978-3-319-18317-6 doi:10.1007/978-3-319-18317-6
- [98] D. W. Kerst *et al.*, “Attainment of very high energy by means of intersecting beams of particles,” *Phys. Rev.*, vol. 102, p. 590, 1956 doi:10.1103/PhysRev.102.590
- [99] T. Schörner-Sadenius, “Background and history,” in *The Large Hadron Collider*, 1st ed., T. Schörner-Sadenius, Ed. Springer International Publishing, 2015, p. 1. ISBN 978-3-319-15001-7 doi:10.1007/978-3-319-15001-7\_1
- [100] CERN Site Management and Buildings Department, CERN Geographic Information System. [Available Online]: <https://maps.cern.ch/>
- [101] A. Asner *et al.*, “Large Hadron Collider in the LEP tunnel: A feasibility study of possible options,” in *Proc. of the 1st ICFA Seminar on Future Perspectives in High Energy Physics*, no. CERN-DIR-TECH-84-01, p. 207, Tsukuba, Japan, May 1984. [Available Online]: <https://inspirehep.net/record/202859>
- [102] O. Brüning *et al.*, Eds., *LHC Design Report*, v.1: The LHC main ring ed. CERN, 2004. ISBN 978-9-290-83224-9 doi:10.5170/CERN-2004-003-V-1. [Available Online]: <https://inspirehep.net/record/656250>
- [103] G. Fraser, “Evian debut,” *CERN Courier*, vol. 32, no. 4, p. 1, 1992. [Available Online]: <https://cds.cern.ch/record/1732019>
- [104] K. Ehret, “Funding of high energy physics,” in *Physics at the Terascale*, 1st ed., I. Brock and T. Schörner-Sadenius, Eds. Wiley, 2011, pp. 415–430. ISBN 978-3-527-63497-2 doi:10.1002/9783527634965
- [105] O. Brüning, H. Burkhardt, and S. Myers, “The Large Hadron Collider,” *Prog. Part. Nucl. Phys.*, vol. 67, p. 705, 2012 doi:10.1016/j.pnpnp.2012.03.001. [Available Online]: <https://inspirehep.net/record/1119913>
- [106] M. Schaumann *et al.*, “First xenon-xenon collisions at LHC,” in *Proc. of the 9th International Particle Accelerator Conference (IPAC'18)*, p. MOPMF039, Vancouver, BC, Canada, 2018. [Available Online]: <https://inspirehep.net/record/1691330>
- [107] G. Arduini *et al.*, “First run of the LHC as a heavy ion collider,” in *Proc. of the 2nd International Particle Accelerator Conference (IPAC'11)*, p. TUPZ016, San Sebastián, Spain, 2011. [Available Online]: <https://inspirehep.net/record/1183213>



- [108] N. Fuster-Martínez *et al.*, “Cleaning performance of the collimation system with Xe beams at the LHC,” in *Proc. of the 9th International Particle Accelerator Conference (IPAC’18)*, p. MOPMF038, Vancouver, BC, Canada, 2018. [Available Online]: <https://inspirehep.net/record/1691329>
- [109] R. Rossi, “Experimental assessment of crystal collimation at the Large Hadron Collider,” Ph.D. dissertation, Sapienza University of Rome, 2016, CERN-THESIS-2017-424. [Available Online]: <https://cds.cern.ch/record/2644175>
- [110] M. Krasny *et al.*, “The CERN Gamma Factory initiative: An ultra-high intensity gamma source,” in *Proc. of the 9th International Particle Accelerator Conference (IPAC’18)*, p. WEYGBD3, Vancouver, BC, Canada, 2018. [Available Online]: <https://inspirehep.net/record/1690500>
- [111] V. Kain, “Beam dynamics and beam losses: Circular machines,” in *Proc. of the 2014 Joint International Accelerator School: Beam Loss and Accelerator Protection*, p. 21, Newport Beach, USA, Nov. 2014 doi:10.5170/CERN-2016-002.21. [Available Online]: <https://inspirehep.net/record/1480005>
- [112] LHC Program Coordination, Luminosity calculator. [Available Online]: <https://lpc.web.cern.ch/lumi2.html>
- [113] CMS Collaboration, “Emittance scans for CMS luminosity calibration in 2017,” CMS Detector Performance Summary CMS-DP-2018-011, 2018. [Available Online]: <https://cds.cern.ch/record/2306378>
- [114] E. Raka, “RF system considerations for a Large Hadron Collider,” in *Proc. of the 1988 US Particle Accelerator School*, p. 288, New York, USA, 1988. [Available Online]: <https://inspirehep.net/record/271117>
- [115] J. M. Jowett *et al.*, “Heavy ions in 2012 and the program up to 2022,” in *Proc. of Chamonix 2012 workshop on LHC Performance*, p. 200, Chamonix, France, 2012. [Available Online]: <https://cds.cern.ch/record/1492972>
- [116] O. Brüning and P. Collier, “Building a behemoth,” *Nature*, vol. 448, p. 285, 2007 doi:10.1038/nature06077. [Available Online]: <https://inspirehep.net/record/760218>
- [117] R. Schmidt *et al.*, “Protection of the CERN Large Hadron Collider,” *New J. Phys.*, vol. 8, p. 290, 2006 doi:10.1088/1367-2630/8/11/290. [Available Online]: <https://inspirehep.net/record/736682>
- [118] M. Gasior *et al.*, “Introduction to beam instrumentation and diagnostics,” in *Proc. of the CAS-CERN Accelerator School: Advanced Accelerator Physics Course*, Trondheim, Norway, 2014. [Available Online]: <https://inspirehep.net/record/1416004>
- [119] G. Trad *et al.*, “Status of the beam instrumentation at the LHC,” in *Proc. of the 7th Evian Workshop*, Evian Les Bains, France, 2017. [Available Online]: <https://inspirehep.net/record/1635446>

- [120] H. Grote and F. Schmidt, "MAD-X: An upgrade from MAD8," in *Proc. of Particle Accelerator Conference 2003 (PAC03)*, p. 3497, Portland, USA, May 2003. [Available Online]: <https://inspirehep.net/record/624489>
- [121] R. Bailey and P. Collier, "Standard filling schemes for various LHC operation modes," Tech. Rep. LHC-PROJECT-NOTE-323, 2003. [Available Online]: <https://cds.cern.ch/record/691782>
- [122] M. Solfaroli, "LHC nominal cycle," in *Proc. of the 6th Evian Workshop*, p. 45, EVIAN, France, Dec. 2015. [Available Online]: <https://inspirehep.net/record/1637700>
- [123] J. Wenninger *et al.*, "First beam test of a combined ramp and squeeze at LHC," Tech. Rep. CERN-ACC-NOTE-2015-0023, 2015. [Available Online]: <https://cds.cern.ch/record/2049886>
- [124] D. Manglunki *et al.*, "The first LHC pPb run: Performance of the heavy ion production complex," in *Proc. of the 4th International Particle Accelerator Conference (IPAC'13)*, p. WEPEA061, Shanghai, China, May 2013. [Available Online]: <https://inspirehep.net/record/1338947>
- [125] J. Wenninger, "Machine protection and operation for LHC," in *Proc. of the 2014 Joint International Accelerator School: Beam Loss and Accelerator Protection*, p. 377, Newport Beach, USA, 2014. [Available Online]: <https://inspirehep.net/record/1480415>
- [126] P. D. Hermes, "Heavy ion collimation at the Large Hadron Collider: Simulations and measurements," Ph.D. dissertation, University of Münster, 2016, CERN-THESIS-2016-230. [Available Online]: <https://cds.cern.ch/record/2241364>
- [127] R. Assmann *et al.*, "The final collimation system for the LHC," in *Proc. of the 10th European Particle Accelerator Conference (EPAC 2006)*, p. 986, Edinburgh, UK, 2006. [Available Online]: <https://inspirehep.net/record/727805>
- [128] P. D. Hermes *et al.*, "Measured and simulated heavy ion beam loss patterns at the CERN Large Hadron Collider," *Nucl. Instr. Meth. Phys. Res. A*, vol. 819, p. 73, 2016 doi:10.1016/j.nima.2016.02.050. [Available Online]: <https://inspirehep.net/record/1426958>
- [129] R. Bruce *et al.*, "Simulations and measurements of beam loss patterns at the CERN Large Hadron Collider," *Phys. Rev. ST Accel. Beams*, vol. 17, p. 081004, 2014 doi:10.1103/PhysRevSTAB.17.081004. [Available Online]: <https://inspirehep.net/record/1315859>
- [130] J. M. Jowett and C. Carli, "The LHC as a proton-nucleus collider," in *Proc. of the 10th European Particle Accelerator Conference (EPAC 2006)*, p. 550, Edinburgh, UK, June 2006. [Available Online]: <https://inspirehep.net/record/727796>

- [131] C. Salgado *et al.*, “Proton-nucleus collisions at the LHC: Scientific opportunities and requirements,” *J. Phys. G.*, vol. 39, p. 015010, 2012 doi:10.1088/0954-3899/39/1/015010. [Available Online]: <https://inspirehep.net/record/900635>
- [132] R. Alemany *et al.*, “First proton-nucleus collisions at LHC: The pPb pilot physics,” Tech. Rep. CERN-ATS-Note-2012-094 MD, 2012. [Available Online]: <https://cds.cern.ch/record/1496101>
- [133] J. M. Jowett *et al.*, “Proton-nucleus collisions at LHC,” in *Proc. of the 4th International Particle Accelerator Conference (IPAC’13)*, p. MOODB201, Shanghai, China, May 2013. [Available Online]: <https://inspirehep.net/record/1336732>
- [134] R. Alemany-Fernandez, “Plans for the ion run in 2016,” in *Proc. of the LHC Performance Workshop (Chamonix 2016)*, Chamonix, France, Jan. 2016. [Available Online]: <https://chamonix-2016.web.cern.ch>
- [135] A. Fernández *et al.*, “Performance of the CERN injector complex and transmission studies into the LHC during the second proton-lead run,” in *Proc. of the 8th International Particle Accelerator Conference (IPAC 2017)*, p. TUPVA128, Copenhagen, Denmark, May 2017 doi:10.18429/JACoW-IPAC2017-TUPVA128. [Available Online]: <https://inspirehep.net/record/1626325>
- [136] R. Aaij *et al.* (LHCb Collaboration), “Precision luminosity measurements at LHCb,” *JINST*, vol. 9, p. P12005, 2014 doi:10.1088/1748-0221/9/12/P12005. [Available Online]: <https://inspirehep.net/record/1319638>
- [137] J. M. Jowett *et al.*, “The 2016 proton-nucleus run at LHC,” in *Proc. of the 8th International Particle Accelerator Conference (IPAC’17)*, p. TUPVA014, Copenhagen, Denmark, May 2017 doi:10.18429/JACoW-IPAC2017-TUPVA014. [Available Online]: <https://inspirehep.net/record/1626343>
- [138] A. M. Sirunyan *et al.* (CMS Collaboration), “Particle-flow reconstruction and global event description with the CMS detector,” *JINST*, vol. 12, p. P10003, 2017 doi:10.1088/1748-0221/12/10/P10003. [Available Online]: <https://inspirehep.net/record/1605397>
- [139] S. Chatrchyan *et al.* (CMS Collaboration), “Commissioning of the CMS experiment and the Cosmic Run at Four Tesla,” *JINST*, vol. 5, p. T03001, 2010 doi:10.1088/1748-0221/5/03/T03001. [Available Online]: <https://inspirehep.net/record/837869>
- [140] M. Huhtinen, “Optimization of the CMS forward shielding,” CMS Note CMS-NOTE-2000-068, 2007. [Available Online]: <https://inspirehep.net/record/876279>
- [141] Forward and small- $x$  QCD physics publications. [Available Online]: <https://cms-results.web.cern.ch/cms-results/public-results/publications/FSQ/index.html>

- [142] CMS Collaboration, “CMS, the magnet project: Technical design report,” Scientific Committee Paper CERN-LHCC-97-10, 1997. [Available Online]: <https://inspirehep.net/record/455671>
- [143] CMS Collaboration, “The CMS tracker system project: Technical design report,” Scientific Committee Paper CERN-LHCC-98-006, 1998. [Available Online]: <https://inspirehep.net/record/492301>
- [144] CMS Collaboration, “The CMS tracker: Addendum to the technical design report,” Scientific Committee Paper CERN-LHCC-2000-01, 2000. [Available Online]: <https://inspirehep.net/record/1614063>
- [145] S. Chatrchyan *et al.* (CMS Collaboration), “Description and performance of track and primary vertex reconstruction with the CMS tracker,” *JINST*, vol. 9, p. P10009, 2014 doi:10.1088/1748-0221/9/10/P10009. [Available Online]: <https://inspirehep.net/record/1298029>
- [146] A. M. Sirunyan *et al.* (CMS Collaboration), “Precision measurement of the structure of the CMS inner tracking system using nuclear interactions,” *JINST*, vol. 13, p. P10034, 2018 doi:10.1088/1748-0221/13/10/P10034. [Available Online]: <https://inspirehep.net/record/1682069>
- [147] CMS Collaboration, “The CMS electromagnetic calorimeter project: Technical design report,” Scientific Committee Paper CERN-LHCC-97-033, 1997. [Available Online]: <https://inspirehep.net/record/1614057>
- [148] CMS Collaboration, “Addendum to the CMS ECAL technical design report: Changes to the CMS ECAL electronics,” Scientific Committee Paper CERN-LHCC-2002-02, 2002. [Available Online]: <https://cds.cern.ch/record/581342>
- [149] S. Chatrchyan *et al.* (CMS Collaboration), “Energy calibration and resolution of the CMS electromagnetic calorimeter in pp collisions at  $\sqrt{s} = 7\text{TeV}$ ,” *JINST*, vol. 8, p. P09009, 2013 doi:10.1088/1748-0221/8/09/P09009. [Available Online]: <https://inspirehep.net/record/1237915>
- [150] P. Adzic *et al.*, “Energy resolution of the barrel of the CMS electromagnetic calorimeter,” *JINST*, vol. 2, p. P04004, 2007 doi:10.1088/1748-0221/2/04/P04004. [Available Online]: <https://inspirehep.net/record/753174>
- [151] W. Bialas and D. A. Petyt, “Mitigation of anomalous APD signals in the CMS ECAL,” *JINST*, vol. 8, p. C03020, 2013 doi:10.1088/1748-0221/8/03/C03020. [Available Online]: <https://inspirehep.net/record/1228756>
- [152] V. Khachatryan *et al.* (CMS Collaboration), “Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at  $\sqrt{s} = 8\text{TeV}$ ,” *JINST*, vol. 10, p. P06005, 2015 doi:10.1088/1748-0221/10/06/P06005. [Available Online]: <https://inspirehep.net/record/1343791>

- [153] CMS Collaboration, “CMS: The hadron calorimeter technical design report,” Scientific Committee Paper CERN-LHCC-97-31, 1997. [Available Online]: <https://inspirehep.net/record/1614058>
- [154] S. Abdullin *et al.* (CMS HCAL/ECAL Collaboration), “The CMS barrel calorimeter response to particle beams from 2 to 350 GeV,” *Eur. Phys. J. C*, vol. 60, p. 359, 2009 doi:10.1140/epjc/s10052-009-0959-5 Erratum: *Eur. Phys. J. C*, 61 (2009) 353. [Available Online]: <https://inspirehep.net/record/826181>
- [155] S. Chatrchyan *et al.* (CMS Collaboration), “Identification and filtering of uncharacteristic noise in the CMS hadron calorimeter,” *JINST*, vol. 5, p. T03014, 2010 doi:10.1088/1748-0221/5/03/T03014. [Available Online]: <https://inspirehep.net/record/837890>
- [156] A. M. Sirunyan *et al.* (CMS Collaboration), “Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at  $\sqrt{s} = 13$  TeV,” *JINST*, vol. 13, p. P06015, 2018 doi:10.1088/1748-0221/13/06/P06015. [Available Online]: <https://inspirehep.net/record/1667449>
- [157] CMS Collaboration, “CMS: The muon project technical design report,” Scientific Committee Paper CERN-LHCC-97-32, 1997. [Available Online]: <https://inspirehep.net/record/1614056>
- [158] L. Scibile *et al.*, “The LHC Radiation Monitoring System for the Environment and Safety: From design to operation,” in *Proc. of the 10th European Particle Accelerator Conference (EPAC 2008)*, vol. 0806233, p. MOPD030, Genoa, Italy, Jun 2008. [Available Online]: <https://inspirehep.net/record/794553>
- [159] A. Kornmayer, “The CMS Pixel Luminosity Telescope,” in *Proceedings of the 13th Pisa Meeting on Advanced Detectors: Frontier Detectors for Frontier Physics*, vol. 824, p. 304, La Biodola, Italy, May 2016. [Available Online]: <https://inspirehep.net/record/1458032>
- [160] P. Lujan, “Performance of the Pixel Luminosity Telescope for luminosity measurement at CMS during Run 2,” in *Proceedings of the European Physical Society Conference on High Energy Physics, (EPS-HEP 2017)*, vol. 314, p. 504, Venice, Italy, Jul 2017. [Available Online]: <https://inspirehep.net/record/1664774>
- [161] A. A. Zagozdzińska *et al.*, “New Fast Beam Conditions Monitoring (BCM1F) system for CMS,” in *Proceedings of the Topical Workshop on Electronics for Particle Physics (TWEPP15)*, vol. 11, p. C01088, Lisbon, Portugal, Oct 2015. [Available Online]: <https://inspirehep.net/record/1418560>
- [162] M. Hempel, “Development of a novel diamond based detector for machine induced background and luminosity measurements,” Ph.D. dissertation, DESY, 2017, DESY-THESIS-2017-030. [Available Online]: <https://inspirehep.net/record/1611774>

- [163] V. Khachatryan *et al.* (CMS Collaboration), “The CMS trigger system,” *JINST*, vol. 12, p. P01020, 2017 doi:10.1088/1748-0221/12/01/P01020. [Available Online]: <https://inspirehep.net/record/1485699>
- [164] J.-M. Andre, “File-based data flow in the CMS Filter Farm,” in *Proc. of the 21st International Conference on Computing in High Energy and Nuclear Physics (CHEP2015)*, vol. 664, p. 082033, Okinawa, Japan, Apr. 2015 doi:10.1088/1742-6596/664/8/082033. [Available Online]: <https://inspirehep.net/record/1414057>
- [165] A. Tapper and D. Acosta, Eds., *CMS technical design report for the level-1 trigger upgrade*. The LHC experiments Committee, 2013. ISBN 9789290833901. [Available Online]: <https://inspirehep.net/record/1261310>
- [166] N. Dev *et al.*, “The cms level-1 electron and photon trigger: For run 2 of lhc,” *JINST*, vol. 12, p. C02014, 2017 doi:10.1088/1748-0221/12/02/C02014. [Available Online]: <https://inspirehep.net/record/1513693>
- [167] R. Fruehwirth, “Application of Kalman filtering to track and vertex fitting,” *Nucl. Instrum. Meth. A*, vol. 262, p. 444, 1987 doi:10.1016/0168-9002(87)90887-4. [Available Online]: <https://inspirehep.net/record/259509>
- [168] M. Cacciari and G. Salam, “Pileup subtraction using jet areas,” *Phys. Lett. B*, vol. 659, p. 119, 2008 doi:10.1016/j.physletb.2007.09.077. [Available Online]: <https://inspirehep.net/record/755413>
- [169] A. J. Bell, “Beam and radiation monitoring for CMS,” in *Proceedings of the 2008 IEEE Nuclear Science Symposium, Medical Imaging Conference and 16th International Workshop on Room-Temperature Semiconductor X-Ray and Gamma-Ray Detectors (NSS/MIC 2008/RTSD 2008)*, p. 2322, Dresden, Germany, Oct 2008. [Available Online]: <https://inspirehep.net/record/1398863>
- [170] C. Foudas *et al.*, “The CMS tracker readout front-end driver,” *EEE Trans. Nucl. Sci.*, vol. 52, p. 2836, 2005 doi:10.1109/TNS.2005.860173. [Available Online]: <https://inspirehep.net/record/605990>
- [171] Run 2 pixel performance plots for data and simulation. [Available Online]: <https://twiki.cern.ch/twiki/bin/view/CMSPublic/PixelOfflinePlots2016>
- [172] CMS silicon strip performance results 2016. [Available Online]: <https://twiki.cern.ch/twiki/bin/view/CMS/StripsOfflinePlots2016>
- [173] M. J. French *et al.*, “Design and results from the APV25, a deep sub-micron CMOS front-end chip for the CMS tracker,” in *Proc. of the 4th International Symposium on Development and Application of Semiconductor Tracking Detectors*, vol. 466, p. 359, Hiroshima, Japan, Mar 2000 doi:10.1016/S0168-9002(01)00589-7. [Available Online]: <https://inspirehep.net/record/563299>

- [174] R. Corke and T. Sjöstrand, “Interleaved parton showers and tuning prospects,” *JHEP*, vol. 1103, p. 032, 2011 doi:10.1007/JHEP03(2011)032. [Available Online]: <https://inspirehep.net/record/875856>
- [175] D. Contardo *et al.*, “Technical proposal for the phase-II upgrade of the CMS detector,” Tech. Rep. CMS-TDR-15-02, 2015. [Available Online]: <https://inspirehep.net/record/1614097>
- [176] J. Alwall *et al.*, “MADGRAPH5: Going beyond,” *JHEP*, vol. 1106, p. 128, 2011 doi:10.1007/JHEP06(2011)128. [Available Online]: <https://inspirehep.net/record/912611>
- [177] S. Chatrchyan *et al.* (CMS Collaboration), “Study of the underlying event at forward rapidity in pp collisions at  $\sqrt{s} = 0.9, 2.76, \text{ and } 7 \text{ TeV}$ ,” *JHEP*, vol. 04, p. 072, 2013 doi:10.1007/JHEP04(2013)072. [Available Online]: <https://inspirehep.net/record/1218372>
- [178] CMS tracking POG performance plots for 2017 with phase-I pixel detector. [Available Online]: <https://twiki.cern.ch/twiki/bin/view/CMSPublic/TrackingPOGPerformance2017MC>
- [179] L. M. Medrano *et al.*, “Studies on luminous region, pileup and performance for HL-LHC scenarios,” in *Proc. of the 8th International Particle Accelerator Conference (IPAC’17)*, p. TUPIK089, Copenhagen, Denmark, May 2017 doi:10.18429/JACoW-IPAC2017-TUPIK089. [Available Online]: <https://inspirehep.net/record/1626355>
- [180] T. Miao *et al.*, “Beam-position determination using tracks,” CMS Note CMS-NOTE-2007-021, 2007. [Available Online]: <https://inspirehep.net/record/764607>
- [181] W. Ashmanskas *et al.*, “CDF silicon vertex tracker: Tevatron Run 2 preliminary results,” *Part. Nucl. Lett.*, vol. 5, p. 12, 2002. [Available Online]: [https://www1.jinr.ru/Pepan\\_letters/panl\\_5\\_2002.html](https://www1.jinr.ru/Pepan_letters/panl_5_2002.html)
- [182] A. Bardi *et al.*, “The CDF online silicon vertex tracker,” *Nucl. Instrum. Meth. A*, vol. 485, p. 178, 2002 doi:10.1142/9789812776464\_0019. [Available Online]: <https://inspirehep.net/record/568551>
- [183] M. Hostettler and G. Papotti, “Beam-size estimation from luminosity scans at the LHC during 2015 proton physics operation,” in *Proc. of the 7th International Particle Accelerator Conference (IPAC 2016)*, p. MOPMR025, Busan, Korea, May 2016. [Available Online]: <https://inspirehep.net/record/1469586>
- [184] F. Antoniou *et al.*, “LHC luminosity modeling for Run 2,” in *Proc. of the 7th International Particle Accelerator Conference (IPAC 2016)*, p. TUPMW002, Busan, Korea, May 2016. [Available Online]: <https://inspirehep.net/record/1469916>
- [185] CMS Collaboration, “Beamspot measurements in 2016,” CMS Detector Performance Summary CMS-DP-2016-051, 2016. [Available Online]: <https://cds.cern.ch/record/2285428>

- [186] D. Buskulic *et al.* (ALEPH Collaboration), “Performance of the ALEPH detector at LEP,” *Nucl. Instrum. Meth. A*, vol. 360, p. 481, 1995 doi:10.1016/0168-9002(95)00138-7. [Available Online]: <https://inspirehep.net/record/381617>
- [187] V. Khachatryan *et al.* (CMS Collaboration), “Jet energy scale and resolution in the CMS experiment in pp collisions at 8TeV,” *JINST*, vol. 12, p. P02014, 2017 doi:10.1088/1748-0221/12/02/P02014. [Available Online]: <https://inspirehep.net/record/1475475>
- [188] A. M. Sirunyan *et al.* (CMS Collaboration), “Measurement of the inelastic proton-proton cross section at  $\sqrt{s} = 13\text{TeV}$ ,” *JHEP*, vol. 1807, p. 161, 2018 doi:10.1007/JHEP07(2018)161. [Available Online]: <https://inspirehep.net/record/1653948>
- [189] CMS Collaboration, “Performance of missing energy reconstruction in 13TeV pp collision data using the CMS detector,” CMS Physics Analysis Summary CMS-PAS-JME-16-004, 2016. [Available Online]: <https://inspirehep.net/record/1479660>
- [190] D. Bertolini, P. Harris, M. Low, and N. Tran, “Pileup Per Particle Identification,” *JHEP*, vol. 10, p. 059, 2014 doi:10.1007/JHEP10(2014)059. [Available Online]: <https://inspirehep.net/record/1307411>
- [191] CMS Collaboration, “Jet algorithms performance in 13 TeV data,” CMS Physics Analysis Summary CMS-PAS-JME-16-003, 2017. [Available Online]: <https://inspirehep.net/record/1519093>
- [192] T. Pierog *et al.*, “EPOS LHC: Test of collective hadronization with data measured at the CERN Large Hadron Collider,” *Phys. Rev. C*, vol. 92, p. 034906, 2015 doi:10.1103/PhysRevC.92.034906. [Available Online]: <https://inspirehep.net/record/1236629>
- [193] G. Petrucciani, A. Rizzi, and C. Vuosalo, “Mini-AOD: A new analysis data format for CMS,” in *Proc. of the 21st International Conference on Computing in High Energy and Nuclear Physics (CHEP2015)*, p. 072052, Okinawa, Japan, Apr 2015 doi:10.1088/1742-6596/664/7/072052. [Available Online]: <https://inspirehep.net/record/1414028>
- [194] S. Agostinelli *et al.* (GEANT4 Collaboration), “GEANT4 – a simulation toolkit,” p. 250, 2003. [Available Online]: <https://inspirehep.net/record/593382>
- [195] M. Cacciari, G. Salam, and G. Soyez, “The anti- $k_T$  jet clustering algorithm,” *JHEP*, vol. 0804, p. 063, 2008 doi:10.1088/1126-6708/2008/04/063. [Available Online]: <https://inspirehep.net/record/779080>
- [196] M. Cacciari, G. Salam, and G. Soyez, “FASTJET user manual,” *Eur. Phys. J. C*, vol. 72, p. 1896, 2012 doi:10.1140/epjc/s10052-012-1896-2. [Available Online]: <https://inspirehep.net/record/955176>



- [197] M. Cacciari, G. P. Salam, and G. Soyez, “The catchment area of jets,” *JHEP*, vol. 0804, p. 005, 2008 doi:10.1088/1126-6708/2008/04/005. [Available Online]: <https://inspirehep.net/record/779079>
- [198] CMS Collaboration, “Jet energy scale and resolution performance with 13TeV data collected by CMS in 2016,” CMS Detector Performance Summary CMS-DP-2018-02, 2018. [Available Online]: <https://cds.cern.ch/record/2622157>
- [199] A. M. Sirunyan *et al.* (CMS Collaboration), “Constraining gluon distributions in nuclei using dijets in proton-proton and proton-lead collisions at  $\sqrt{s_{NN}} = 5.02$  TeV,” *Phys. Rev. Lett.*, vol. 121, p. 062002, 2018 doi:10.1103/PhysRevLett.121.062002. [Available Online]: <https://inspirehep.net/record/1672941>
- [200] V. Khachatryan *et al.* (CMS Collaboration), “Performance of photon reconstruction and identification with the CMS detector in proton-proton collisions at  $\sqrt{s} = 8$  TeV,” *JINST*, vol. 10, p. P08010, 2015 doi:10.1088/1748-0221/10/08/P08010. [Available Online]: <https://inspirehep.net/record/1343792>
- [201] A. M. Sirunyan *et al.* (CMS Collaboration), “Identification of heavy-flavor jets with the CMS detector in pp collisions at 13TeV,” *JINST*, vol. 13, p. P05011, 2018 doi:10.1088/1748-0221/13/05/P05011. [Available Online]: <https://inspirehep.net/record/1644362>
- [202] S. Chatrchyan *et al.* (CMS Collaboration), “Identification of b quark jets with the CMS experiment,” *JINST*, vol. 8, p. P04013, 2013 doi:10.1088/1748-0221/8/04/P04013. [Available Online]: <https://inspirehep.net/record/1203133>
- [203] V. Khachatryan *et al.* (CMS Collaboration), “Measurement of  $B\bar{B}$  angular correlations based on secondary vertex reconstruction at  $\sqrt{s} = 7$  TeV,” *JINST*, vol. 136, p. 1103, 2011 doi:10.1007/JHEP03(2011)136. [Available Online]: <https://inspirehep.net/record/889807>
- [204] CMS Collaboration, “Constraints on nuclear parton distributions from W boson production in pPb collisions at  $\sqrt{s_{NN}} = 8.16$  TeV,” CMS Physics Analysis Summary CMS-PAS-HIN-17-007, 2018. [Available Online]: <https://inspirehep.net/record/1673108>
- [205] N. Tosi, “The CMS beam halo monitor at the LHC: Implementation and first measurements,” in *Proc. of the 5th International Beam Instrumentation Conference (IBIC 2016)*, p. TUPG20, Barcelona, Spain, Sep. 2017 doi:10.18429/JACoW-IBIC2016-TUPG20. [Available Online]: <https://inspirehep.net/record/1639624>
- [206] L. Bonolis, “Bruno Touschek vs. machine builders: AdA, the first matter-antimatter collider,” *Riv. Nuovo Cim.*, vol. 28N11, p. 1, 2005 doi:10.1393/ncr/i2005-10006-x. [Available Online]: <https://inspirehep.net/record/706626>
- [207] G. Antchev *et al.* (TOTEM Collaboration), “Luminosity-independent measurements of total, elastic and inelastic cross sections at  $\sqrt{s} = 7$  TeV,” *EPL*, vol.

- 101, p. 21004, 2013 doi:10.1209/0295-5075/101/21004. [Available Online]: <https://inspirehep.net/record/1220864>
- [208] G. Abbiendi *et al.* (OPAL Collaboration), “Precision luminosity for  $Z^0$  lineshape measurements with a silicon-tungsten calorimeter,” *Eur. Phys. J. C*, vol. 14, p. 373, 2000 doi:10.1007/s100520000353. [Available Online]: <https://inspirehep.net/record/509391>
- [209] L. Adamczyk *et al.* (ZEUS Collaboration), “Measurement of the luminosity in the ZEUS experiment at HERA II,” *Nucl. Instrum. Meth. A*, vol. 744, p. 80, 2014 doi:10.1016/j.nima.2014.01.053. [Available Online]: <https://inspirehep.net/record/1237232>
- [210] N. Amos *et al.*, “Measurement of small-angle antiproton-proton and proton-proton elastic scattering at the CERN intersecting storage rings,” *Nucl. Phys. B*, vol. 262, p. 689, 1985 doi:10.1016/0550-3213(85)90511-5. [Available Online]: <https://inspirehep.net/record/214689>
- [211] S. van der Meer, “Calibration of the effective beam height in the ISR,” Tech. Rep. CERN-ISR-PO-68-31, 1968. [Available Online]: <https://inspirehep.net/record/1098817>
- [212] B. B. Abelev *et al.* (ALICE Collaboration), “Measurement of visible cross sections in proton-lead collisions at  $\sqrt{s_{NN}} = 5.02$  TeV in van der Meer scans with the ALICE detector,” *JINST*, vol. 9, p. P11003, 2014 doi:10.1088/1748-0221/9/11/P11003. [Available Online]: <https://inspirehep.net/record/1294934/>
- [213] ALICE Collaboration, “ALICE luminosity determination for pp collisions at  $\sqrt{s} = 8$  TeV,” Tech. Rep. ALICE-PUBLIC-2017-002, 2017. [Available Online]: <https://cds.cern.ch/record/2255216>
- [214] G. Aad *et al.* (ATLAS Collaboration), “Improved luminosity determination in pp collisions at  $\sqrt{s} = 7$  TeV using the ATLAS detector at the LHC,” *Eur. Phys. J. C*, vol. 73, p. 2518, 2013 doi:10.1140/epjc/s10052-013-2518-3. [Available Online]: <https://inspirehep.net/record/1219960>
- [215] ALICE Collaboration, “ALICE luminosity determination for pp collisions at  $\sqrt{s} = 13$  TeV,” Tech. Rep. ALICE-PUBLIC-2016-002, 2016. [Available Online]: <https://cds.cern.ch/record/2160174>
- [216] ALICE Collaboration, “ALICE luminosity determination for pp collisions at  $\sqrt{s} = 5$  TeV,” Tech. Rep. ALICE-PUBLIC-2016-005, 2017. [Available Online]: <https://cds.cern.ch/record/2202638>
- [217] CMS Collaboration, “CMS luminosity measurement for the 2015 data taking period,” CMS Physics Analysis Summary CMS-PAS-LUM-15-001, 2015. [Available Online]: <https://inspirehep.net/record/1427340>

- [218] V. Balagura, “Notes on van der Meer scan for absolute luminosity measurement,” *Nucl. Instrum. Meth. A*, vol. 654, p. 634, 2011 doi:10.1016/j.nima.2011.06.007. [Available Online]: <https://inspirehep.net/record/891611>
- [219] S. Papadopoulou *et al.*, “Modeling and measurements of bunch profiles at the LHC,” in *Proc. of the 8th International Particle Accelerator Conference (IPAC’17)*, p. TUPVA044, Copenhagen, Denmark, May 2017 doi:10.1088/1742-6596/874/1/012008. [Available Online]: <https://inspirehep.net/record/1611150>
- [220] F. Antoniou *et al.*, “Can we predict luminosity?” in *Proc. of the 7th Evian Workshop on LHC beam operation*, p. 125, Evian Les Bains, France, Dec. 2016. [Available Online]: <https://inspirehep.net/record/1635435>
- [221] G. Iadarola *et al.*, “MD468: Electron cloud reference fills and dependence on the bunch intensity,” Tech. Rep. CERN-ACC-NOTE-2017-0029, 2007. [Available Online]: <https://cds.cern.ch/record/2260999>
- [222] L. Mether, P. Dijkstalf, G. Iadarola, and G. Rumolo, “Electron cloud in 2016: Cloudy or clear?” in *Proc. of the 7th Evian Workshop on LHC beam operation*, p. 147, Evian Les Bains, France, Dec. 2017. [Available Online]: <https://inspirehep.net/record/1635438>
- [223] C. Barschel *et al.*, “Results of the LHC DCCT calibration studies,” Tech. Rep. CERN-ATS-Note-2012-026 PERF, 2012. [Available Online]: <https://cds.cern.ch/record/1425904>
- [224] D. Belohrad *et al.*, “The LHC Fast BCT system: A comparison of design parameters with initial performance,” CERN, Geneva, Tech. Rep. CERN-BE-2010-010, May 2010. [Available Online]: <https://cds.cern.ch/record/1267400>
- [225] G. Papotti, T. Bohl, F. Follin, and U. Wehrle, “Longitudinal beam measurements at the LHC: The LHC Beam Quality Monitor,” in *Proc. of the 2nd International Particle Accelerator Conference, (IPAC 2011)*, vol. C110904, p. 1852, 2011. [Available Online]: <https://inspirehep.net/record/1183208>
- [226] A. Jeff *et al.*, “Longitudinal density monitor for the LHC,” *Phys. Rev. ST Accel. Beams*, vol. 15, p. 032803, 2012 doi:10.1103/PhysRevSTAB.15.032803. [Available Online]: <https://link.aps.org/doi/10.1103/PhysRevSTAB.15.032803>
- [227] C. Barschel, “Precision luminosity measurement at LHCb with beam-gas imaging,” Ph.D. dissertation, RWTH Aachen University, 2014, CERN-THESIS-2013-301. [Available Online]: <https://inspirehep.net/record/1339684>
- [228] M. Gasior, J. Olexa, and R. Steinhagen, “BPM electronics based on compensated diode detectors: Results from development systems,” *Conf. Proc.*, vol. C1204151, no. CERN-ATS-2012-247, p. MOPG010, Apr 2012. [Available Online]: <https://cds.cern.ch/record/1476070>

- [229] M. Zanetti, “Beams scan based absolute normalization of the CMS luminosity measurement. CMS 2010 luminosity determination,” in *Proceedings of the LHC Lumi Days: LHC Workshop on LHC Luminosity Calibration*, p. 69, Geneva, Switzerland, Jan 2011. [Available Online]: <https://cds.cern.ch/record/1357856>
- [230] “CMS luminosity based on the pixel-cluster counting: Summer 2012 update,” CMS Physics Analysis Summary CMS-PAS-LUM-12-001, 2012. [Available Online]: <https://inspirehep.net/record/1260806>
- [231] M. Rihl, “Imaging the LHC beams with silicon and scintillating fibre vertex detectors,” in *Proc. of the 14th Vienna Conference on Instrumentation (VCI 2016)*, p. 575, Vienna, Austria, Feb. 2015 doi:10.1016/j.nima.2016.04.051. [Available Online]: <https://inspirehep.net/record/1513522>
- [232] A. Alexopoulos *et al.*, “First LHC transverse beam-size measurements with the Beam Gas Vertex detector,” in *Proc. of the 8th International Particle Accelerator Conference (IPAC’17)*, p. TUOAB1, Copenhagen, Denmark, May 2017 doi:10.18429/JACoW-IPAC2017-TUOAB1. [Available Online]: <https://inspirehep.net/record/1611202>
- [233] M. Klute, C. A. Medlock, and J. Salfeld-Nebgen, “Beam-imaging and luminosity calibration,” *JINST*, vol. 12, p. P03018, 2017 doi:10.1088/1748-0221/12/03/P03018. [Available Online]: <https://inspirehep.net/record/1427292>
- [234] W. Kozanecki, T. Pieloni, and J. Wenninger, “Observation of beam-beam deflections with LHC orbit data,” Tech. Rep. CERN-ACC-NOTE-2013-0006, 2013. [Available Online]: <https://cds.cern.ch/record/1581723>
- [235] R. Tomas *et al.*, “CERN Large Hadron Collider optics model, measurements, and corrections,” *Phys. Rev. ST Accel. Beams*, vol. 13, p. 121004, 2010 doi:10.1103/PhysRevSTAB.13.121004. [Available Online]: <https://inspirehep.net/record/882698>
- [236] A. Alici *et al.*, “Study of the LHC ghost charge and satellite bunches for luminosity calibration,” Tech. Rep. CERN-ATS-Note-2012-029 PERF, 2012. [Available Online]: <https://cds.cern.ch/record/1427728>
- [237] M. Rihl, “Development of a nondestructive beam-profile monitor based on scintillating fibre planes and SIPMs,” Ph.D. dissertation, Vienna Technical University, 2018, CERN-THESIS-2018-09. [Available Online]: <https://cds.cern.ch/record/2318275>
- [238] ATLAS Collaboration, “Measurements of W and Z boson production in pp collisions at  $\sqrt{s} = 5.02$  TeV with the ATLAS detector,” 2018, submitted to *Eur. Phys. J. C*. [Available Online]: <https://inspirehep.net/record/1699375>
- [239] ALICE Collaboration, “ALICE luminosity determination for pPb collisions at  $\sqrt{s_{NN}} = 8.16$  TeV,” Tech. Rep. ALICE-PUBLIC-2018-002, 2018. [Available Online]: <https://cds.cern.ch/record/2314660>

- [240] R. Aaij *et al.* (LHCb Collaboration), “Study of  $Y$  production in pPb collisions at  $\sqrt{s_{NN}} = 8.16\text{TeV}$ ,” *JHEP*, vol. 1811, p. 194, 2018 doi:10.1007/JHEP11(2018)194. [Available Online]: <https://inspirehep.net/record/1699106>
- [241] S. Chatrchyan *et al.* (CMS Collaboration), “Determination of the top quark pole mass and strong coupling constant from the  $t\bar{t}$  production cross section in pp collisions at  $\sqrt{s} = 7\text{TeV}$ ,” *Phys. Lett. B*, vol. 728, p. 496, 2013 doi:10.1016/j.physletb.2013.12.009 Erratum: *Phys. Lett. B*, 738 (2014) 526. [Available Online]: <https://inspirehep.net/record/1241819>
- [242] A. M. Sirunyan *et al.* (CMS Collaboration), “Measurement of double-differential cross sections for top quark pair production in pp collisions at  $\sqrt{s} = 8\text{TeV}$  and impact on parton distribution functions,” *Eur. Phys. J. C*, vol. 77, p. 459, 2017 doi:10.1140/epjc/s10052-017-4984-5. [Available Online]: <https://inspirehep.net/record/1516191/>
- [243] M. Czakon, M. Mangano, A. Mitov, and J. Rojo, “Constraints on the gluon PDF from top quark pair production at hadron colliders,” *JHEP*, vol. 1307, p. 167, 2013 doi:10.1007/JHEP07(2013)167. [Available Online]: <https://inspirehep.net/record/1225789>
- [244] G. Aad *et al.* (ATLAS Collaboration), “Measurement of the top quark pair production cross section with ATLAS in the single lepton channel,” *Phys. Lett. B*, vol. 711, p. 244, 2012 doi:10.1016/j.physletb.2012.03.083. [Available Online]: <https://inspirehep.net/record/1083794>
- [245] G. Aad *et al.* (ATLAS Collaboration), “Measurement of the  $t\bar{t}$  production cross section using  $e\mu$  events with b-tagged jets in pp collisions at  $\sqrt{s} = 7$  and 8 TeV with the ATLAS detector,” *Eur. Phys. J. C*, vol. 74, p. 3109, 2014 doi:10.1140/epjc/s10052-014-3109-7. [Available Online]: <https://inspirehep.net/record/1301856>
- [246] V. Khachatryan *et al.* (CMS Collaboration), “Measurement of the  $t\bar{t}$  production cross section in the  $e\mu$  channel in proton-proton collisions at  $\sqrt{s} = 7$  and 8 TeV,” *JHEP*, vol. 08, p. 029, 2016 doi:10.1007/JHEP08(2016)029. [Available Online]: <https://inspirehep.net/record/1426692>
- [247] V. Khachatryan *et al.* (CMS Collaboration), “Measurements of the  $t\bar{t}$  production cross section in lepton+jets final states in pp collisions at 8 TeV and ratio of 8 to 7 TeV cross sections,” *Eur. Phys. J. C*, vol. 77, p. 15, 2017 doi:10.1140/epjc/s10052-016-4504-z. [Available Online]: <https://inspirehep.net/record/1424842>
- [248] M. Aaboud *et al.* (ATLAS Collaboration), “Measurement of the  $t\bar{t}$  production cross section using  $e\mu$  events with b-tagged jets in pp collisions at  $\sqrt{s} = 13\text{TeV}$  with the ATLAS detector,” *Phys. Lett. B*, vol. 761, p. 136, 2016 doi:10.1016/j.physletb.2016.08.019. [Available Online]: <https://inspirehep.net/record/1468168>

- [249] V. Khachatryan *et al.* (CMS Collaboration), “Measurement of the top quark pair production cross section in proton-proton collisions at  $\sqrt{s} = 13$  TeV,” *Phys. Rev. Lett.*, vol. 116, p. 052002, 2016 doi:10.1103/PhysRevLett.116.052002. [Available Online]: <https://inspirehep.net/record/1398582>
- [250] V. Khachatryan *et al.* (CMS Collaboration), “Measurement of the  $t\bar{t}$  production cross section using events in the  $e\mu$  final state in pp collisions at  $\sqrt{s} = 13$  TeV,” *Eur. Phys. J. C*, vol. 77, p. 172, 2017 doi:10.1140/epjc/s10052-017-4718-8. [Available Online]: <https://inspirehep.net/record/1497736/>
- [251] CMS Collaboration, “Projections for heavy ions with HL-LHC,” CMS Physics Analysis Summary CMS-PAS-FTR-13-025, 2013. [Available Online]: <https://inspirehep.net/record/1260902>
- [252] K. Melnikov and F. Petriello, “Electroweak gauge boson production at hadron colliders through  $O(\alpha_s^2)$ ,” *Phys. Rev. D*, vol. 74, p. 114017, 2006 doi:10.1103/PhysRevD.74.114017. [Available Online]: <https://inspirehep.net/record/725573>
- [253] S. Alioli, P. Nason, C. Oleari, and E. Re, “NLO single-top production matched with shower in POWHEG:  $s$ - and  $t$ -channel contributions,” *JHEP*, vol. 09, p. 111, 2009 doi:10.1088/1126-6708/2009/09/111 Erratum: *JHEP*, 09 (2009) 111. [Available Online]: <https://inspirehep.net/record/826654>
- [254] E. Re, “Single-top  $Wt$ -channel production matched with parton showers using the POWHEG method,” *Eur. Phys. J. C*, vol. 71, p. 1547, 2011 doi:10.1140/epjc/s10052-011-1547-z. [Available Online]: <https://inspirehep.net/record/867473>
- [255] N. Kidonakis, “Top quark production,” in *Proc. of the Helmholtz International Summer School on Physics of Heavy Quarks and Hadrons (HQ 2013)*, p. 139, Dubna, Russia, Jul 2013 doi:10.3204/DESY-PROC-2013-03/Kidonakis. [Available Online]: <https://inspirehep.net/record/1263209>
- [256] J. M. Campbell and R. K. Ellis, “MCFM for the Tevatron and the LHC,” *Nucl. Phys. Proc. Suppl.*, vol. 205–206, p. 10, 2010 doi:10.1016/j.nuclphysbps.2010.08.011. [Available Online]: <https://inspirehep.net/record/861954>
- [257] M. Czakon and A. Mitov, “TOP++: A program for the calculation of the top pair cross section at hadron colliders,” *Comput. Phys. Commun.*, vol. 185, p. 2930, 2014 doi:10.1016/j.cpc.2014.06.021. [Available Online]: <https://inspirehep.net/record/1082926>
- [258] M. Czakon, P. Fiedler and A. Mitov, “Total top quark pair production cross section at hadron colliders through  $O(\alpha_s^4)$ ,” *Phys. Rev. Lett.*, vol. 110, p. 252004, 2013 doi:10.1103/PhysRevLett.110.252004. [Available Online]: <https://inspirehep.net/record/1225282>

- [259] E. Todesco and J. Wenninger, “Large Hadron Collider momentum calibration and accuracy,” *Phys. Rev. Accel. Beams*, vol. 20, p. 081003, 2017 doi:10.1103/PhysRevAccelBeams.20.081003. [Available Online]: <https://inspirehep.net/record/1618103>
- [260] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, “Asymptotic formulae for likelihood-based tests of new physics,” *Eur. Phys. J. C*, vol. 71, p. 1554, 2011 doi:10.1140/epjc/s10052-011-1554-0 Erratum: *Eur. Phys. J. C*, 73 (2013) 2501. [Available Online]: <https://inspirehep.net/record/860907>
- [261] L. Lyons, D. Gibaut, and P. Clifford, “How to combine correlated estimates of a single physical quantity,” *Nucl. Instrum. Meth. A*, vol. 270, p. 110, 1988 doi:10.1016/0168-9002(88)90018-6. [Available Online]: <https://inspirehep.net/record/22979>
- [262] A. Valassi and R. Chierici, “Information and treatment of unknown correlations in the combination of measurements using the BLUE method,” *Eur. Phys. J. C*, vol. 74, p. 2717, 2014 doi:10.1140/epjc/s10052-014-2717-6. [Available Online]: <https://inspirehep.net/record/1242645>
- [263] L. Lista, “The bias of the unbiased estimator: A study of the iterative application of the BLUE method,” *Nucl. Instrum. Meth. A*, vol. 764, p. 82, 2014 doi:10.1016/j.nima.2014.07.021 Erratum: *Nucl. Instrum. Meth. A*, 773 (2015) 87. [Available Online]: <https://inspirehep.net/record/1296116>
- [264] V. Khachatryan *et al.* (CMS Collaboration), “Measurement of the differential cross section and charge asymmetry for inclusive  $pp \rightarrow W^\pm + X$  production at  $\sqrt{s} = 8$  TeV,” *Eur. Phys. J. C*, vol. 76, p. 469, 2016 doi:10.1140/epjc/s10052-016-4293-4. [Available Online]: <https://inspirehep.net/record/1426517>
- [265] S. Alekhin *et al.*, “HERAFitter, open source QCD fit project,” *Eur. Phys. J. C*, vol. 75, p. 304, 2015 doi:10.1140/epjc/s10052-015-3480-z. [Available Online]: <https://inspirehep.net/record/1322585>
- [266] <https://www.xfitter.org>. [Available Online]: [www.xfitter.org](http://www.xfitter.org)
- [267] V. N. Gribov and L. N. Lipatov, “Deep inelastic ep scattering in perturbation theory,” *Sov. J. Nucl. Phys.*, vol. 15, p. 438, 1972. [Available Online]: <https://inspirehep.net/record/73449>
- [268] G. Altarelli and G. Parisi, “Asymptotic freedom in parton language,” *Nucl. Phys. B*, vol. 126, p. 298, 1977 doi:10.1016/0550-3213(77)90384-4. [Available Online]: <https://inspirehep.net/record/119585>
- [269] G. Curci, W. Furmanski, and R. Petronzio, “Evolution of parton densities beyond leading order: The non-singlet case,” *Nucl. Phys. B*, vol. 175, p. 27, 1980 doi:10.1016/0550-3213(80)90003-6. [Available Online]: <https://inspirehep.net/record/152873>

- [270] W. Furmanski and R. Petronzio, “Singlet parton densities beyond leading order,” *Phys. Lett. B*, vol. 97, p. 437, 1980 doi:10.1016/0370-2693(80)90636-X. [Available Online]: <https://inspirehep.net/record/155291>
- [271] S. Moch, J. A. M. Vermaseren, and A. Vogt, “The three-loop splitting functions in QCD: The non-singlet case,” *Nucl. Phys. B*, vol. 688, p. 101, 2004 doi:10.1016/j.nuclphysb.2004.03.030. [Available Online]: <https://inspirehep.net/record/646539>
- [272] A. Vogt, S. Moch, and J. A. M. Vermaseren, “The three-loop splitting functions in QCD: The singlet case,” *Nucl. Phys. B*, vol. 691, p. 129, 2004 doi:10.1016/j.nuclphysb.2004.04.024. [Available Online]: <https://inspirehep.net/record/648209>
- [273] M. Botje, “QCDNUM: Fast QCD evolution and convolution,” *Comput. Phys. Commun.*, vol. 182, p. 490, 2011 doi:10.1016/j.cpc.2010.10.020. [Available Online]: <https://inspirehep.net/record/854529>
- [274] M. Aliev, H. Lacker, U. Langenfeld, S. Moch, P. Uwer, and M. Wiedermann, “HATHOR: HAdronic Top and Heavy quarks crOss section calculatoR,” *Comput. Phys. Commun.*, vol. 182, p. 1034, 2011 doi:10.1016/j.cpc.2010.12.040. [Available Online]: <https://inspirehep.net/record/860789>
- [275] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt, “Parton distributions for the LHC,” *Eur. Phys. J. C*, vol. 63, p. 189, 2009 doi:10.1140/epjc/s10052-009-1072-5. [Available Online]: <https://inspirehep.net/record/810127/>
- [276] S. Chatrchyan *et al.* (CMS Collaboration), “Measurement of the muon charge asymmetry in inclusive  $pp \rightarrow W + X$  production at  $\sqrt{s} = 7$  TeV and an improved determination of light parton distribution functions,” *Phys. Rev. D*, vol. 90, p. 032004, 2014 doi:10.1103/PhysRevD.90.032004. [Available Online]: <https://inspirehep.net/record/1273570>
- [277] W. T. Giele and S. Keller, “Implications of hadron collider observables on parton distribution function uncertainties,” *Phys. Rev. D*, vol. 58, p. 094023, 1998 doi:10.1103/PhysRevD.58.094023. [Available Online]: <https://inspirehep.net/record/468349>
- [278] W. T. Giele, S. A. Keller, and D. A. Kosower, “Parton distribution function uncertainties,” 2001. [Available Online]: <https://inspirehep.net/record/555079>
- [279] D. d’Enterria, “Top quark pair production cross sections at NNLO+NNLL in pPb collisions at  $\sqrt{s_{NN}} = 8.16$  TeV,” 2017. [Available Online]: <https://inspirehep.net/record/1608021>
- [280] V. Khachatryan *et al.* (CMS Collaboration), “Study of W boson production in pPb collisions at  $\sqrt{s_{NN}} = 5.02$  TeV,” *Phys. Lett. B*, vol. 750, p. 565, 2015



- doi:10.1016/j.physletb.2015.09.057. [Available Online]: <https://inspirehep.net/record/1353541>
- [281] V. Khachatryan *et al.* (CMS Collaboration), “Study of Z boson production in pPb collisions at  $\sqrt{s_{\text{NN}}} = 5.02 \text{ TeV}$ ,” *Phys. Lett. B*, vol. 759, p. 36, 2016 doi:10.1016/j.physletb.2016.05.044. [Available Online]: <https://inspirehep.net/record/1410832>
- [282] A. Dainese *et al.*, “Heavy ions at the Future Circular Collider,” *CERN Yellow Report*, p. 635, 2017 doi:10.23731/CYRM-2017-003.635. [Available Online]: <https://inspirehep.net/record/1455787>
- [283] M. J. Oreglia, “A study of the reactions  $\psi' \rightarrow \gamma\gamma\psi$ ,” Ph.D. dissertation, Stanford University, SLAC Report SLAC-R-236, see Appendix D. [Available Online]: <https://inspirehep.net/record/158483>
- [284] W. Verkerke and D. P. Kirkby, “The ROOFIT toolkit for data modeling,” in *Proc. of the 13th International Conference on Computing in High Energy and Nuclear Physics (CHEP03)*, p. MOLT007, La Jolla, CA, USA, March. [Available Online]: <https://inspirehep.net/record/621398>
- [285] K. S. Cranmer, “Kernel estimation in high-energy physics,” *Comput. Phys. Commun.*, vol. 136, p. 198, 2001 doi:10.1016/S0010-4655(00)00243-5. [Available Online]: <https://inspirehep.net/record/537082>
- [286] J. Gao *et al.*, “CT10 next-to-next-to-leading order global analysis of QCD,” *Phys. Rev. D*89, vol. 89, p. 033009, 2014 doi:10.1103/PhysRevD.89.033009. [Available Online]: <https://inspirehep.net/record/1221243>
- [287] ATLAS and CMS Collaborations and LHC Higgs Combination Group, “Procedure for the LHC Higgs boson search combination in Summer 2011,” ATLAS/CMS joint note ATL-PHYS-PUB-2011-011, CMS-NOTE-2011-005, 2011. [Available Online]: <https://inspirehep.net/record/1196797>
- [288] Public CMS luminosity information. [Available Online]: [https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults#2017\\_proton\\_proton\\_5\\_02\\_TeV\\_coll](https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults#2017_proton_proton_5_02_TeV_coll)
- [289] CMS Collaboration, “Measurements of normalized multi-differential cross sections for top quark pair production in pp collisions at  $\sqrt{s} = 13 \text{ TeV}$  and simultaneous determination of the strong coupling strength, top quark pole mass, and parton distribution functions,” CMS Physics Analysis Summary CMS-PAS-TOP-18-004, 2018. [Available Online]: <https://inspirehep.net/record/1704287>
- [290] CMS Collaboration, “Projection of measurements of differential  $t\bar{t}$  production cross sections in the  $e/\mu$ +jets channels in pp collisions at the HL-LHC,” CMS Physics Analysis Summary CMS-PAS-FTR-18-015, 2018. [Available Online]: <https://inspirehep.net/record/1708974>

- [291] CMS Collaboration, “Constraining nuclear parton distributions with heavy ion collisions at the HL-LHC with the CMS experiment,” CMS Physics Analysis Summary CMS-PAS-FTR-18-027, 2018. [Available Online]: <https://cds.cern.ch/record/1709472>
- [292] ATLAS and CMS Collaborations, “Report on the physics at the HL-LHC and perspectives for the HE-LHC,” Tech. Rep. CERN-LPCC-2019-01, 2018, submitted as input to the update of the European Particle Physics Strategy (EPPS). [Available Online]: <https://cds.cern.ch/record/2651134/>
- [293] A. Dainese *et al.*, The physics potential of HL-LHC, submitted as input to the update of the European Particle Physics Strategy (EPPS). [Available Online]: <https://twiki.cern.ch/twiki/pub/LHCPhysics/HLHELHCWorkshop/report.pdf>
- [294] A. Dainese *et al.*, The physics potential of HE-LHC, submitted as input to the update of the European Particle Physics Strategy (EPPS). [Available Online]: <https://twiki.cern.ch/twiki/pub/LHCPhysics/HLHELHCWorkshop/HEreport.pdf>
- [295] Citron, Z. and others, “Future physics opportunities for high-density QCD at the LHC with heavy ion and proton beams,” 2018, submitted as input to the update of the European Particle Physics Strategy (EPPS). [Available Online]: <https://inspirehep.net/record/1709331>
- [296] Azzi, P. and others, “Standard model physics at the HL-LHC and HE-LHC,” Tech. Rep. CERN-LPCC-2018-03, 2018, submitted as input to the update of the European Particle Physics Strategy (EPPS). [Available Online]: <https://cds.cern.ch/record/2650160>
- [297] L. Barzè *et al.*, “Neutral current Drell-Yan with combined QCD and electroweak corrections in the POWHEG BOX,” *Eur. Phys. J. C*, vol. 73, p. 2474, 2013 doi:10.1140/epjc/s10052-013-2474-y. [Available Online]: <https://inspirehep.net/record/1220256>
- [298] V. Khachatryan *et al.* (CMS Collaboration), “Measurements of inclusive W and Z cross sections in pp collisions at  $\sqrt{s} = 7\text{TeV}$ ,” *JHEP*, vol. 01, p. 080, 2011 doi:10.1007/JHEP01(2011)080. [Available Online]: <https://inspirehep.net/record/881087>
- [299] CMS Collaboration, “Pileup mitigation at CMS in 13TeV data using the 2016 data set,” 2018, JME-18-001 *in preparation*.
- [300] R. Cousins, J. Linnemann, and T. Jordan, “Evaluation of three methods for calculating statistical significance when incorporating a systematic uncertainty into a test of the background-only hypothesis for a Poisson process,” *Nucl. Instrum. Meth. A*, vol. 595, p. 480, 2008 doi:10.1016/j.nima.2008.07.086. [Available Online]: <https://inspirehep.net/record/1600163>

- [301] T. Schörner-Sadenius, Ed., *The Large Hadron Collider*, 1st ed. Springer International Publishing, 2015. ISBN 978-3-319-15001-7 [doi:10.1007/978-3-319-15001-7](https://doi.org/10.1007/978-3-319-15001-7)



