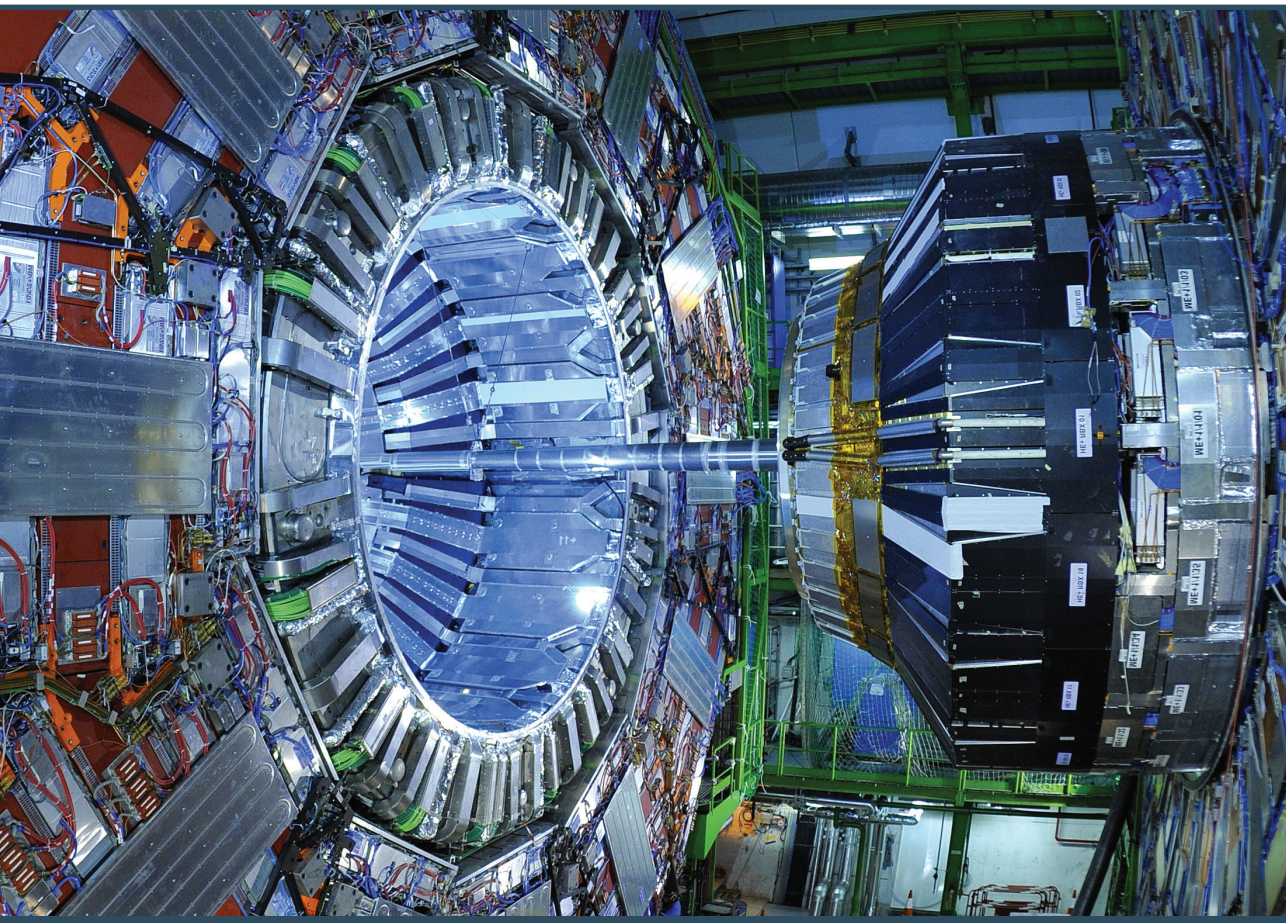
RIGA TECHNICAL UNIVERSITY

Andris Potrebko

THE MEASUREMENT OF THE MASS DIFFERENCE BETWEEN THE TOP QUARK AND ANTIQUARK AT 13 TEV PROTON-PROTON COLLISIONS USING CMS DETECTOR

Summary of the Doctoral Thesis



RTU Press
Riga 2025

RIGA TECHNICAL UNIVERSITY
Faculty of Natural Sciences and Technology
Institute of Particle Physics and Accelerator Technologies

Andris Potrebko
Doctoral Student of the Study Programme
"Particle Physics and Accelerator Technologies"

Measurement of the mass difference between the top quark and antiquark at 13 TeV proton-proton collisions using the CMS detector

Summary of the Doctoral Thesis

Scientific supervisors:

Associate Professor Dr. phys.
KĀRLIS DREIMANIS

Dr. phys.
MARTIJN MULDERS (CERN)

Scientific advisor:
Associate Professor Dr. phys.
MARKUS SEIDEL

RTU Press
Riga 2025

Potrebko, A. Measurement of the Mass Difference between the Top Quark and Antiquark in 13 TeV Proton-Proton Collisions Using the CMS Detector. Summary of the Doctoral Thesis. – Riga: RTU Press, 2025. – 66 p.

Published in accordance with the decision of RTU IP-PAT attestation committee session of 29 January 2025, Minutes No. 1/2025.

The research was carried out within the framework and with financial support of the following projects:

- State Research Programme (SRP) "High-energy physics and accelerator technologies" projects VPP-IZM-CERN-2020/1-0002; VPP-IZM-CERN-2022/1-0001.

Cover picture from www.shutterstock.com.

https://doi.org/10.7250/9789934372346
ISBN 978-9934-37-234-6 (pdf)

DOCTORAL THESIS PROPOSED TO RIGA TECHNICAL UNIVERSITY FOR PROMOTION TO THE SCIENTIFIC DEGREE OF DOCTOR OF SCIENCE

To be granted the scientific degree of Doctor of Science (PhD), the present Doctoral Thesis has been submitted for defence at the open meeting of RTU Promotion Council on December 11, 2025 at 11.00 at the Faculty of Natural Sciences and Technology of Riga Technical University, Āzenes street 6, Conference Hall on the 11th Floor and Zoom online https://cern.zoom.us/j/69802317118?pwd=hkNGtWOXDjQCt95b8wuEoS3WWuRuNc.1.

OFFICIAL REVIEWERS

Professor Dr. phys. Mark Owen,
University of Glasgow, UK

Associate professor Dr. phys. Philip James Ilten,
University of Cincinnati, USA

Dr. phys. James Keaveney,
University of Cape Town, South Africa

DECLARATION OF ACADEMIC INTEGRITY

I hereby declare that the Doctoral Thesis submitted for review to Riga Technical University for promotion to the scientific degree of Doctor of Science (PhD) is my own. I confirm that this Doctoral Thesis has not been submitted to any other university for promotion to a scientific degree.

Andris Potrebko ...../signature/
Date .....

The Doctoral Thesis has been written in English. It consists of an Introduction, 6 chapters, Conclusions, 85 figures, 11 tables, and 7 appendices; the total number of pages is 168, not including appendices. The Bibliography contains 348 titles.

ABSTRACT

The Standard Model (SM) of particle physics has been remarkably successful in describing the fundamental forces and particles observed in experiments such as those at the Large Hadron Collider (LHC). However, it fails to explain several phenomena, including the matter-antimatter asymmetry of the universe, the nature of dark matter, and neutrino oscillations. The SM is built on a rich structure of symmetries, including the internal gauge symmetry $SU(3) \times SU(2) \times U(1)$, which governs the strong, weak, and electromagnetic interactions. Although strong and electromagnetic forces respect the parity (P), charge conjugation (C), and time reversal (T) symmetries, the weak force is known to violate C, P, T and the combined CP symmetry. Nevertheless, CPT symmetry has withstood all the experimental tests to date.

A discovery of even a slight CPT violation would fundamentally challenge our understanding of particle physics, requiring a significant revision of the SM. One of the key predictions of the CPT symmetry is the exact equality of particles and antiparticles. The top quark, as the heaviest known elementary particle, provides a unique opportunity to probe CPT symmetry at the highest available energy scales.

In this Thesis, the measurement of the mass difference between the top quark and antiquark Δm_t is presented using data collected by the CMS experiment at the LHC. The analysis is based on proton-proton collisions at a centre-of-mass energy of 13 TeV corresponding to an integrated luminosity of $137.63\,\mathrm{fb}^{-1}$. The measured value, $\Delta m_t = 139 \pm 67\,\mathrm{MeV}$, represents the most precise determination of Δm_t to date.

Contents

Acronyms

CMSSW CMS software 23

BSM beyond the standard model 9

CERN European Organization for Nuclear Research 27

CHS charged hadron subtraction 32, 43

CKM Cabbibo–Kabayashi–Maskawa 9

CMS Compact Muon Solenoid 11, 12, 16, 27

ECAL electronic calorimeter 28, 30, 40

EFT effective field theory 10

EW electroweak 23

FSR final-state radiation 21

HCAL hadronic calorimeter 27–30, 40

HEP high-energy physics 21

ISR initial-state radiation 21

JEC jet energy corrections 31

JER jet energy resolution 32

JES jet energy scale 32, 56

LEP Large Electron–Positron Collider 27

LHC Large Hadron Collider 9, 17, 27

LO leading order 22

MC Monte Carlo 12, 21, 30–32

ME matrix element 23

ML machine learning 55

MPI multi-parton interactions 22

PDF parton distribution function 21

PDG Particle Data Group 11

PF particle flow 30

PU pileup 31, 32

PV primary vertex 30

QFT quantum field theory 9, 18, 20

SF scale factor 30

SM standard model of particle physics 9, 10, 12, 16–20, 56

SPS super proton synchrotron 27

SUSY supersymmetry 9

SV secondary vertex 31

UE underlying event 22, 23, 31, 36

WM-DTM `WMassDeltaTopMass` 46, 48

WP working point 31, 43

1. INTRODUCTION

Since the discovery of the Higgs boson, the standard model of particle physics (SM) has been a consistent framework that describes the interactions of elementary particles, such as those produced at the Large Hadron Collider (LHC). So far no significant deviation from the theoretical predictions has been observed [1]–[3]. However, the SM is known to fail to account for the observed matter-antimatter asymmetry in the universe [4], describe phenomena such as neutrino oscillations [5], [6], dark matter [7], [8], and does not in a consistent way include the force of gravity [9], [10], motivating the search for extensions of the SM, called beyond the standard model (BSM) physics. The search for many theoretical ideas, such as supersymmetry (SUSY) [11], leptoquarks [12], axion-like particles [13], [14], has not been successful. Recently, the search for BSM physics has shifted towards precision tests of the SM, where increasingly precise measurements are compared with increasingly precise predictions.

The SM is a quantum field theory (QFT) which describes three out of four fundamental forces, electromagnetism, weak and strong interactions, but not gravity. The electromagnetic and strong interactions are symmetric under parity, P, charge conjugation, C, and time reversal, T, operations. In contrast, the weak force has been found to maximally violate P and partly the combined CP symmetry. However, it is predicted that the combination of the three discrete symmetries CPT will be completely preserved by any Lorentz-invariant and local quantum field theory, such as the SM [15]–[17]. Therefore, SM extensions violating CPT have to violate either locality or Lorentz invariance. It has been shown that using string theory, which is inherently non-local, it is possible to consistently implement CPT violating effects in the SM [18]. In addition, it has been shown that locality and Lorentz invariance violating effects can be added in the neutrino sector through a mechanism called ghost condensation [19], [20]. Nevertheless, based on the previous success of unexpected discoveries of C and CP violations, it is worthwhile to test CPT violation in the SM.

1.1. Previous tests of CPT symmetry

CPT symmetry predicts the equality of particles and their corresponding antiparticles. Tests of CPT symmetry include the measurements of mass, decay width, magnetic moments, production cross sections, and charges [21]. Currently, the best constraint comes from neutral kaon K^0 and antikaon $\overline{K}^0$ measurements and is [22]

$$2\frac{|m_{K^0} - m_{\overline{K^0}}|}{(m_{K^0} + m_{\overline{K^0}})} < 6 \times 10^{-19}, \qquad 2\frac{|\Gamma_{K^0} - \Gamma_{\overline{K^0}}|}{(\Gamma_{K^0} + \Gamma_{\overline{K^0}})} < (8 \pm 8) \times 10^{-18}, \tag{1.1}$$

where m_{K^0} and Γ_{K^0} are kaon mass and decay width. Although the constraint is very significant, the result is obtained indirectly from the Cabbibo–Kabayashi–Maskawa (CKM)

mixing matrix parameters. This and other significant measurements of the particle-antiparticle mass asymmetry are shown in Fig. 1.1, comparing their precision and the scale at which the effect could take place. It shows that while the precision of the measurements at low scales of $\mathcal{O}(1\,\mathrm{GeV})$ is high, the region around the scale of the top quark mass is less explored. Recently, contributions have been made to assess the CPT violation systematically using the effective field theory (EFT) approach [23], which can assume a different sensitivity to the CPT violation for each quark generation. In addition to these searches, the Lorentz symmetry has been tested in the top-quark sector, and no deviation from the full symmetry has been found [24].

1.2. Previous measurements of top quark mass difference

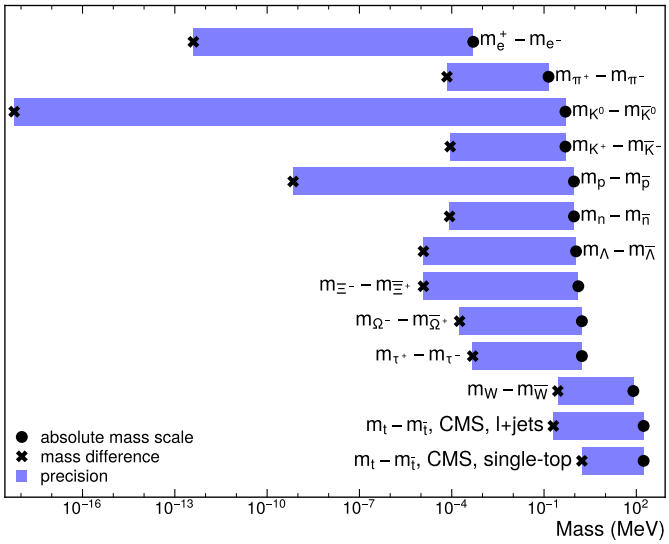


Fig. 1.1. The world best measurements of the mass difference between particles and antiparticles. The black markers represent the scale at which the measurement is made, i.e. the average mass of the particles. The crosses show the measured upper limit on the measured mass difference. The length of the bars corresponds to the precision of the measurement. Data from [21], [25]–[29] are used.

The top quark, due to its short lifetime, decays weakly before hadronization. This allows direct measurements of its mass m_t, unlike for other quarks whose mass has to be deduced from hadron properties. Thus, it also allows for a precise measurement of the mass difference between the top quark and antiquark, $\Delta m_t = m_t - m_{\bar{t}}$. The current results of the measurements of Δm_t are shown in Fig. 1.2, where good consistency with the SM over all the measurements is seen. The most precise measurements of Δm_t are

performed in the top-quark pair process ($t\bar{t}$) in the lepton+jets decay channel, while a complementary measurement from single-top events provides a consistency check of the measurements. The different precision in the measurements is explained by different approaches to the measurement. In the measurements of the Compact Muon Solenoid (CMS) collaboration [30], [31] Δm_t was obtained from two separate datasets, for the top quark and the antiquark, only using the hadronically decaying top quark for the mass measurement and discarding the leptonically decaying top quark. On the other hand, in the other measurements, Δm_t was obtained for each event from both the hadronic and leptonic decays. The latter approach requires estimating the momentum of the lepton neutrino using the missing transverse momentum $p_{\mathrm{T}}^{\mathrm{miss}}$, which degrades the resolution of the measurement.

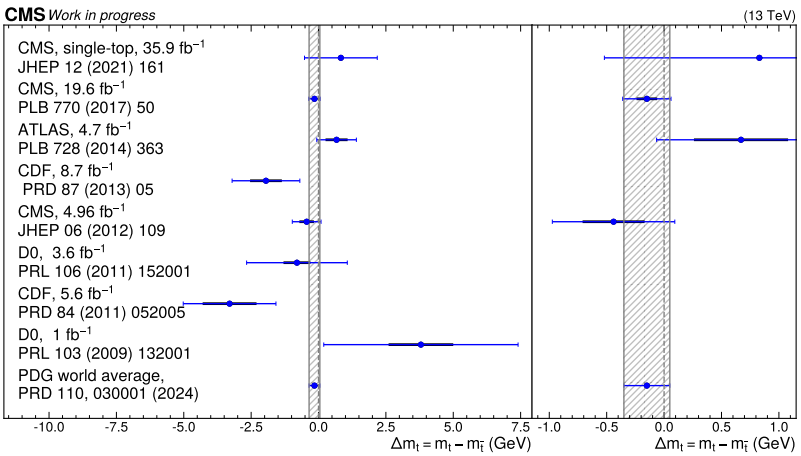


Fig. 1.2. The different measurements of the mass difference between the top quark and antiquark. The thick black bars show the systematic uncertainty, and the blue bars show the total uncertainty. The gray hatched bar shows the world average with the uncertainty obtained by the Particle Data Group (PDG) [21]. The right panel shows a narrower range in Δm_t than the left pane. The data from [21], [31]–[37] are shown.

1.3. Motivation for repeating the measurement in Run 2

The current world best measurement was performed by the CMS collaboration [30] using the data collected at the CMS Run 1 at the centre-of-mass energy $\sqrt{s} = 8\,\mathrm{TeV}$ with the integrated luminosity of $19.6\,\mathrm{fb}^{-1}$. The measurement was consistent with $\Delta m_t = 0$, but was statistically limited. Repeating the measurement using the CMS Run 2 dataset with a luminosity of $137.63\,\mathrm{fb}^{-1}$ offers a tenfold increase in the amount of $t\bar{t}$ events for analysis, and therefore, the statistical uncertainty could be reduced by a factor $\approx \sqrt{10} \approx$

3. In addition, an improved strategy for the quark vs. antiquark jet energy response uncertainty estimation provides a decrease in the systematic uncertainty.

This Thesis summary introduces the main results of the Δm_t analysis and the basic theoretical and experimental background. The Thesis starts with Chapter 2, where the main theoretical concepts are introduced, giving great detail to symmetries in SM. Chapter 3 builds on that explaining how the Monte Carlo (MC) simulations for particle physics collisions are obtained. This section also mentions the author's study to obtain a CMS-ATLAS common MC $t\bar{t}$ dataset simulated using the SHERPA event generator. The results of this study are published in a CMS-ATLAS common note [38], [39]. While the dataset is not used for the current measurement, it will be important to understand and reduce the systematic uncertainties of further measurements, such as those due to different hadronization models.

The experimental system, including the collider and the CMS detector used to obtain the data used, is explained in Chapter 4, while the event reconstruction methods are described in Chapter 5. Expanding on that, the study of the author of this Thesis on jet flavour studies and jet energy reconstruction is shown in Chapter 6. The flavour-antiflavor uncertainties necessary for the Δm_t measurement are derived there as well as the flavour uncertainties necessary for the jet reconstruction in the CMS collaboration in general. The results shown in this section are described in an internal CMS note [40]. The measurement technique and results of Δm_t are described in Chapter 7. These results are published in an internal CMS note [41]. Finally, a short summary and outlook is provided in Chapter 8.

Aim of the Doctoral Thesis

To conduct the **most precise to date** measurement of Δm_t, performing the measurement using the full CMS Run 2 dataset selecting $t\bar{t}$ events in the lepton+jets decay channel.

Tasks of the Doctoral Thesis

1. Generate a $t\bar{t}$ dataset using the SHERPA event generator.

2. Obtain flavour-dependent jet energy corrections to correct the peak of the W boson mass in top-quark decays.

3. Obtain improved flavour-dependent jet energy scale uncertainties and quark-antiquark jet energy scale uncertainties comparing the flavour responses of the PYTHIA 8 and HERWIG 7 event generator programmes and thus estimating the uncertainty in the MC generators.

4. Obtain quark-antiquark jet energy scale uncertainties arising due to the disagreement of the simulated detector response with the CMS hadronic calorimeter test beam data.

5. Implement a profile likelihood fit for the Δm_t measurement. Add an estimation of all the relevant systematic uncertainties and evaluate their impact on Δm_t. Perform the fit, and analyse the result.

Thesis statements to be defended

1. The updated method for quark-antiquark jet energy uncertainty in CMS and the implementation of the profile-likelihood fit instead of the Ideogram method **reduces the systematic uncertainty of the Δm_t measurement by a factor of 3**.

2. The usage of the full CMS Run 2 dataset instead of the Run 1 dataset, **reduces the statistical uncertainty by a factor of 3**.

3. The disagreement of the simulated detector response with the CMS test beam data for charged pions **is the leading systematic uncertainty in the measurement**.

4. A $t\bar{t}$ dataset generated using the SHERPA event generator **describes the H_T distribution in data better ($\chi^2/n_{dof} = 0.42$)**than the currently used POWHEG +PYTHIA 8 dataset **($\chi^2/n_{dof} = 1.41$)**.

5. The flavour-dependent **jet energy uncertainty for gluons is reduced by 50 %** when using the CMS Run 2 simulation instead of Run 1.

Scientific novelty

1. This Thesis presents a novel approach for estimating the quark-antiquark jet energy scale uncertainties using the differences in event generators.

2. This Thesis shows the first estimation of the effect of the charged pion detector response mismodelling by the GEANT4 programme on the jet energy scale.

3. The Thesis introduces the first practically usable CMS-ATLAS common SHERPA $t\bar{t}$ dataset.

4. This Thesis presents the first measurement of Δm_t using the CMS Run 2 dataset and results in the world-leading measurement of this observable.

Practical significance

1. The refined flavour-dependent jet energy scale uncertainties developed in this Thesis will benefit the entire CMS collaboration, increasing the precision of key measurements, such as the measurements of the strong coupling constant, top-quark mass, and the cross-section measurements of top-quark pairs associated with additional jets. This will improve our ability to rigorously test the standard model, pushing the limits of current theoretical predictions and potentially revealing signs of new physics.

2. The new common CMS-ATLAS top-pair dataset generated with SHERPA enables a simplified ATLAS-CMS result comparison and validation, and can lead to an improved estimation of the jet flavour uncertainties.

3. A discovery of CPT symmetry violation would revolutionise our understanding of fundamental physics, challenging long-held assumptions and reshaping theoretical frameworks. Such breakthroughs not only expand scientific knowledge but historically have driven technological advancements of substantial practical significance, often in unforeseen ways, ranging from advancements in materials science to quantum computing and beyond.

Approbation of PhD Thesis in Scopus and Web of Science indexed articles

During the doctoral studies, the author of the Thesis co-authored 204 publications, most as a member of the CMS collaboration. Of these, 17 are the publications of the CMS Top Physics Analysis Group, where the author of this Thesis is an active member. Statistics was retrieved on 14.03.2025 from https://inspirehep.net/authors/1829953. The full list is added at the end of the Thesis.

Additionally, the study of the author of this Thesis is presented in the following 4 documents that are internally reviewed have been approved by the CMS collaboration to contain valuable information that will be included in further CMS publications.

[1] *CMS-ATLAS common note (Chapter 3 in the Thesis), public*
The ATLAS and CMS Collaborations, "Improved Common $t\bar{t}$ Monte-Carlo Settings for ATLAS and CMS", technical report CMS-NOTE-2023-004, CERN, Geneva, 2023 cds:2861366.

[2] *CMS internal note (Chapter 6 in the Thesis)*
Andris Potrebko, "Flavor dependent (L5) MC truth jet energy corrections and flavor uncertainties in Run 2", technical report, CERN, Geneva, 2024, CMS AN-23/074.

[3] *CMS internal note (Chapter 7 in the Thesis)*
Andris Potrebko et al., "Measurement of the Mass Difference between the Top Quark and Antiquark in $t\bar{t}$ Events with Lepton+jets Final States in 13 TeV Proton-Proton Collisions", technical report, CERN, Geneva, 2024, CMS AN-25/029.

[4] *Proceedings in the conference Moriond QCD'24 (sections in Chapter 2 in the Thesis)*
Andris Potrebko, "Review of the measurements of the strong coupling constant in CMS at 13 TeV. Contribution to the 2024 QCD session of the 58th Rencontres de Moriond". In: 58th Rencontres de Moriond on QCD and High Energy Interactions. June 2024. arXiv: 2406.01405.

Other publications on the topic that are not included in PhD Thesis

[1] **Andris Potrebko** and Inese Polaka, "Application of Gene Expression Programming in Improving the Event Selection of the Semi-leptonic Top Quark Pair Process". In: *2021 62nd International Scientific Conference on Information Technology and Management Science of Riga Technical University (ITMS)*. IEEE. 2021, pp. 1–6. DOI 10.1109/ITMS52826.2021.9615317.

Dissemination in international scientific conferences

[1] Andris Potrebko, "Flavor-dependent (L5) MC truth jet energy corrections and flavor uncertainties in Run 2". In: 3rd CERN Baltic Conference. Oct. 2023. URL: https://indico.cern.ch/event/1288731/contributions/5585483/.

[2] Andris Potrebko, "Flavor corrections". In: CMS JetMET workshop in Brussels. May 2023. URL: https://indico.cern.ch/event/1230157/contributions/5328141/.

[3] Andris Potrebko, "Common ttbar Monte-Carlo sample for CMS and ATLAS". In: 12th LHC students poster session. Nov. 2022. URL: https://indico.cern.ch/event/1204801/contributions/5136626/.

[4] Andris Potrebko, "Top pair process simulation and jet energy studies". In: 2nd CERN Baltic Conference. Oct. 2022. URL: https://indico.cern.ch/event/1147717/contributions/5082195/.

[5] Viesturs Veckalns and Andris Potrebko, "CMS Latvia Group activities in the CMS Top Physics Analysis Group". In: 1st CERN Baltic Conference. June 2021. URL: https://indico.cern.ch/event/970609/contributions/4415899.

2. THEORETICAL BACKGROUND

2.1. Basic structure of the standard model

In an experiment carried out in 1897 by J. J. Thomson, the first elementary particle, the electron, was discovered [42]. The last particle added to SM, the Higgs boson, was discovered in 2012 by the CMS and ATLAS collaborations [43], [44]. This finding was the final piece necessary to complete the SM. In this section, the author briefly introduces the structure of SM that was gradually assembled on the basis of several scientific breakthroughs between the two mentioned before.

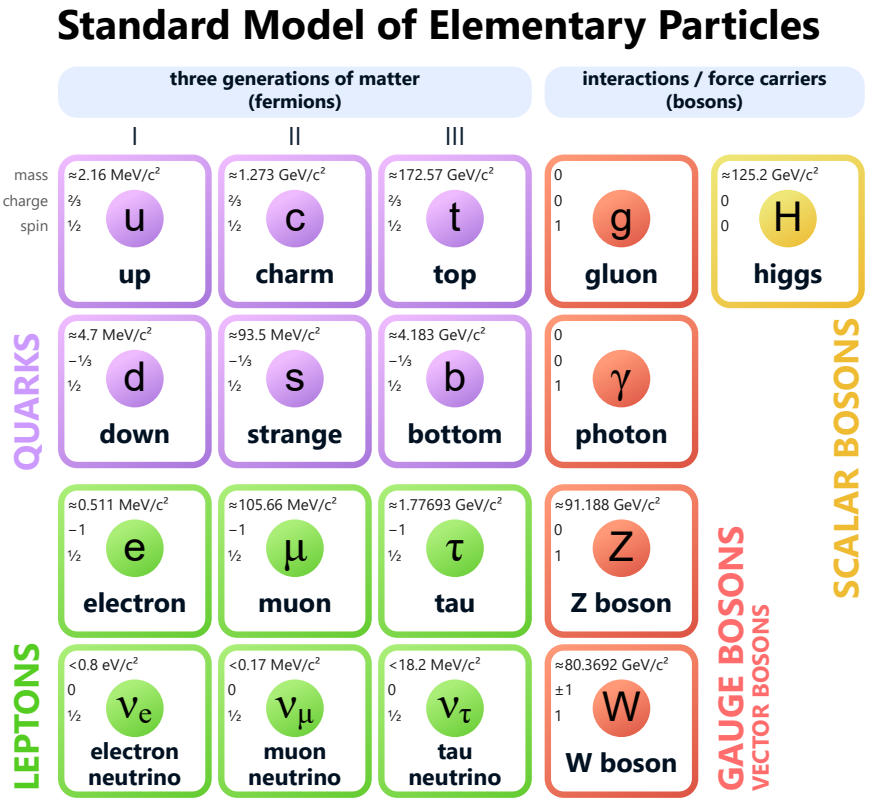


Fig. 2.1. The table of elementary particles constituting the SM.

The overview of currently discovered elementary particles is shown in Fig. 2.1. It consists of spin 1/2 particles, called fermions, spin 1 particles, gauge bosons, and the Higgs boson. Fermions are subdivided into quarks and leptons, where leptons are further subdivided into charged leptons, electron e, muon μ, and tau τ, with electric charge, $Q = -1$ and the corresponding neutrinos with $Q = 0$. In addition to interacting electromagneti-

cally, leptons interact weakly. Quarks interact electromagnetically, weakly, and strongly. They have a subinteger Q, with up-type quarks (u, c, and t, or top quark) having $Q = 2/3$ and down-type quarks (d, s, and b) having $Q = -1/3$. Fermions are divided into three generations, where the second and third generations consist of successively heavier copies of the first-generation particles with the same quantum numbers.

Force is carried by gauge bosons: the electromagnetic force by photons γ, the weak nuclear force by heavy Z and W bosons, and the strong force by eight gluons g. The Higgs boson, H, is the boson corresponding to the Higgs field, which, through the Higgs mechanism, is responsible for the generation of the masses of elementary particles.

All the fermions mentioned above are known as matter particles. In the SM, each fermion has a corresponding antimatter counterpart (antiparticle) with the same properties as the matter particle, except for an opposite charge. Neutral bosons can be considered as their own antiparticles, while W^+ and W^- are the antiparticles of each other. A notable example of antimatter is the positron e^+, the antimatter partner of the electron e^-. Neutrinos ν and antineutrinos $\overline{\nu}$ have the same properties, but they participate in different interactions, as seen in W boson decays $W^+ \to e^+\nu$ and $W^- \to e^-\overline{\nu}$.

Quarks possess an additional quantum number, called colour, which is usually defined as red, blue, and green. Antiquarks have anticolour, i.e. antired, antiblue, and antigreen. Gluons come in eight different types, corresponding to all the possible non-singlet combinations of colour and anticolour. Confinement is an experimental fact that at low energies quarks bind into colour neutral hadrons before reaching the detector. This binding process is called hadronization. The colour-neutral bound states of quarks are called hadrons. Two typical types of hadrons are quark-antiquark bound states, called mesons, and three quark bound states, called baryons. An example of a baryon is the proton, $p = uud$. Pions that can be charged, $\pi^+ = u\overline{d}$, $\pi^- = d\overline{u}$, or neutral, $\pi^0 = \frac{u\overline{u} \pm d\overline{d}}{\sqrt{2}}$ are examples of mesons. Additionally, tetraquarks [45] and pentaquarks [46] that are four- and five-quark bound states have recently been discovered.

2.2. Top quark

The top quark is the heaviest particle in the SM. Due to the large mass, the top quark almost always decays weakly before hadronization, unlike other quarks, which after their production bind into hadrons. The top quark decays almost always take place through a b quark and a W boson, i.e $t \to Wb$.

At the LHC and in pp and e^+e^- colliders in general, the top quark is mostly generated in top quark-antiquark pairs $t\bar{t}$. An example of a $t\bar{t}$ process is depicted in Fig. 2.2, where the red lines show the hard process, and the orange lines show the decays of the top quark. The dominant $t\bar{t}$ production process is the gluon-gluon fusion $gg \to t\bar{t}$, although it can also be produced by other means, like quark-antiquark annihilation.

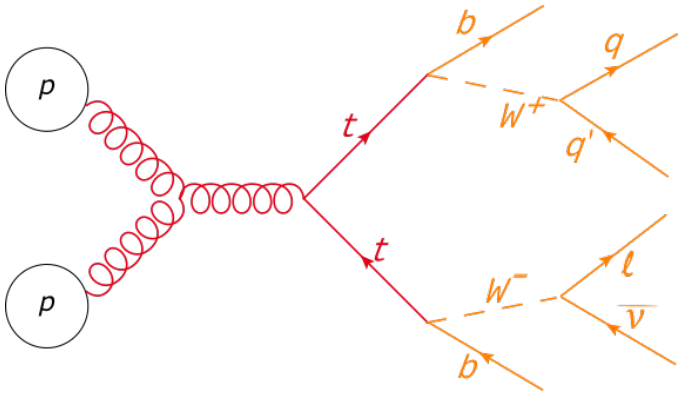


Fig. 2.2. An example of a Feynman diagram showing top quark pair production by gluon fusion and the lepton+jets decay channel.

After the top decay, the W boson can undergo a leptonic $W \to l\nu$ or hadronic decay $W \to qq'$, where the quarks hadronize and produce a collimated spray of particles, called jets. The $t\bar{t}$ events are classified according to the decays of each of the W bosons, into all-hadronic, lepton+jets and dilepton decay channels, if, respectively, both, one or none of the W bosons decay hadronically. The typical decay channel used in precision measurements is the lepton+jets channel. The signature of this channel, namely, the presence of one high-energetic (hard) charged lepton, large missing transverse momentum, $p_{\mathrm{T}}^{\mathrm{miss}}$[1], and at least four jets, out of which two can be further identified as bottom jets (b tagged), ensures good event purity (number of signal events over the total number of events). At the same time, good precision in the measurement can be achieved using the hadronically decaying top quark.

2.3. Symmetries

The early 20th century revealed a remarkable observation – most of the physics governing our world satisfies a vast set of symmetries. Each symmetry is associated with a transformation that leaves the equations of motion unchanged. In this way, for each symmetry, a quantity that is not observable through physics experiments can be defined. Such quantities are effectively "hidden" by the invariance of the laws of physics and have no absolute meaning or relevance. For example, as a consequence of the space transformation, the absolute coordinate is meaningless and the start of the coordinate system can be chosen arbitrarily. In the SM of particle physics, the symmetries are embedded in the physics equations using the framework of quantum field theory (QFT). In QFT, for

[1]The missing transverse momentum vector $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ is defined as the projection onto the plane perpendicular to the beam axis of the negative vector sum of the momenta of all reconstructed the objects in an event. Its magnitude is referred to as $p_{\mathrm{T}}^{\mathrm{miss}}$.

Table 2.1

The symmetries of the SM showing the quantity that cannot be absolutely observed, the symmetry transformation that leaves the Lagrangian invariant and the conserved quantity due to Noether's theorem (for continuous symmetries) or the selection law (for discrete symmetries). Parity and time reversal are broken in weak interactions while the combined $SU_L(2)_L \times U_Y(1)$ is broken into $U_{\mathrm{EM}}(1)$ by the Higgs mechanism, just like $SU(N_f)$ is broken by different couplings of Higgs to each of the generations of the fermions

	Non-observable	Symmetry transformation	Conserved quantity
Poincaré	absolute space position	space translation $\vec{r} \to \vec{r} + \vec{\Delta}$	momentum, $\vec{p}$
	absolute time	time translation $t \to t + \tau$	energy, E
	absolute space direction	rotation $\vec{r} \to R \cdot \vec{r}$	angular momentum, $\vec{L}$
	absolute velocity	Lorentz boost	centre-of-energy, $\vec{K} = E\vec{r}_{cm} - t\vec{P}$
Discrete	absolute right or left	$\vec{r} \to -\vec{r}$	parity, P
	absolute sign of electric charge, Q	$Q \to -Q$	charge conjugation, C
	absolute time direction	$t \to -t$	time reversal, T
Internal	relative phase between states of different charge	$U(1)$ transformation $\psi \to e^{iQ\theta}\psi$	charge, Q
	difference between coherent mixtures of a lepton and neutrino under weak interaction	$SU(2)$ transformation $\psi \to \psi' = \exp^{i\vec{\chi}(x)\cdot\vec{\sigma}/2}\,\psi$	weak isospin/ weak current
	absolute orientation in the colour space	$SU(3)$ transformation $\psi \to \psi' = \exp^{i\vec{\chi}(x)\cdot\vec{\lambda}/2}\,\psi$	total colour charge
	phase between the generations of fermions	$SU(N_f)$	the fermion generations

each particle a separate field spanning the universe is introduced, where the particles are excitations of the field.

An overview of the symmetries present in the SM is provided in Table 2.1. Symmetries are classified as continuous if they involve a smooth, continuous transformation of the system, and discrete otherwise. Internal symmetries are symmetries that act only on fields and not on spacetime points. Noether's theorem [47] states that every continuous symmetry leads to a conservation law associated with a conserved quantity, called a conserved current.

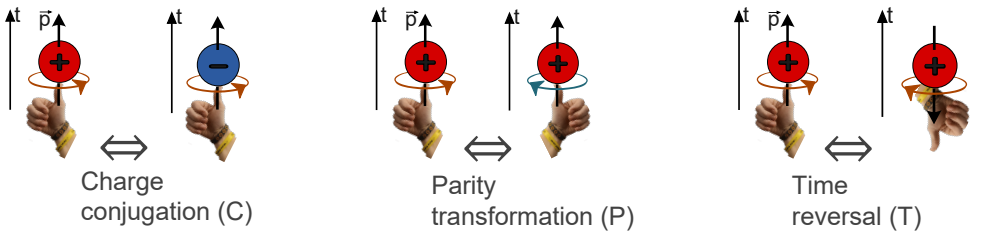


Fig. 2.3. Graphical depictions of the C, P, and T symmetries.

The discrete symmetries in the SM are depicted in Fig. 2.3. These include the invariances under the C, P, and T transformations. The C transformation conjugates all the internal quantum numbers of the particle, including the electromagnetic charge, $Q \to -Q$ and the colour charge. The other two discrete transformations, P and T, act on the spacetime and transform the spacetime as $(t, \vec{x}) \to (t, -\vec{x})$ and as $(t, \vec{x}) \to (-t, \vec{x})$, respectively. Assuming rotational symmetry in the SM, two of the three dimensions in the P transformation can be rotated to the initial position, such that P is analogical to the mirror reflection around one of the axes.

Experimentally, no deviation from the exact C, P, and T symmetries are found in the gravitational, electromagnetic and strong interactions. However, the weak force maximally violates P, as first discovered in the experiment carried out by Chien-Shiung Wu measuring the beta decays of cobalt-60 nuclei [48]. This unexpected discovery showed that only left-handed particles and only right-handed antiparticles interact with the weak force. In other words, the weak interaction discriminates between the interaction and its mirror image.

In 1964, by the observation of oscillations between K_S and K_L, the violation of the CP symmetry was also observed [49]. However, unlike P that is maximally violated, the CP violation is small, characterized by $\mathbf{Re}(\varepsilon'/\varepsilon) = (1.66 \pm 0.23) \times 10^{-3}$.

The CPT theorem states that any local Lorentz invariant QFT must be invariant under the CPT transformation [15]–[17]. Therefore, CP violation also implies T violation. An intuitive analogy of this is that particles can distinguish not only the interaction from its mirror image but also the direction in time. The CPT theorem implies that antiparticles (obtained by the CP transformation) behave like particles moving back in time. Thus, it also implies that observables like masses should be equal for particles and antiparticles.

Despite strong limits on CPT violation, as mentioned in Chapter 1, the violation can be introduced by violating locality [18] or Lorentz symmetry [19], [20], [50]. However, both must be violated only weakly to ensure agreement with current SM measurements.

3. MONTE CARLO SIMULATION

To test our understanding of the SM, experimental results are compared with numerical ones, typically obtained using Monte Carlo (MC) techniques. As the experimental precision improves, the numerical precision must also be improved to achieve sufficient sensitivity for detecting any deviations from the SM or signals of new physics. This drives the continuous development of high-energy physics (HEP) event generators.

In general, the objective of MC generators is to predict an expectation value of an observable, O. Typically, this is a count of events that are observed in a given detector acceptance. Mathematically, it is equivalent to evaluating an integral

$$\langle O \rangle = \int d\phi_n \frac{d\sigma_{A,B\to n}}{d\phi_n} O(\phi_n), \tag{3.1}$$

where the integral runs over, ϕ_n, the phase space of n final-state particles, $\frac{d\sigma_{A,B\to n}}{d\phi_n}$ is the differential cross section for initial state particles A and B to produce final state n, and $O(\phi_n)$ is the value of the observable at the given phase space point.

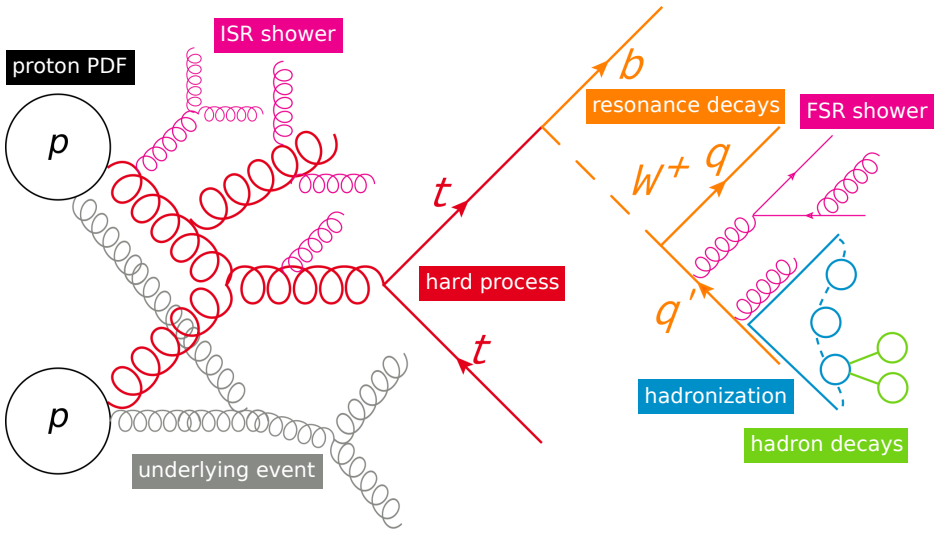


Fig. 3.1. A schematic depiction of the several steps required for a generation of a top quark pair event [51].

The complete generation of a proton-proton collision process involves many steps that can be factored into several stages, as illustrated for the $t\bar{t}$ process in Fig. 3.1. The hard process and subsequent resonance decays are the main components of the large momentum transfer process, involving typically the physics of interest, e.g. a creation of a heavy particle. Parton distribution functions (PDFs) describe the probability of a given parton (quark or gluon) to contain the momentum fraction x of the proton when probed at the given energy. Final-state radiation (FSR) and initial-state radiation (ISR)

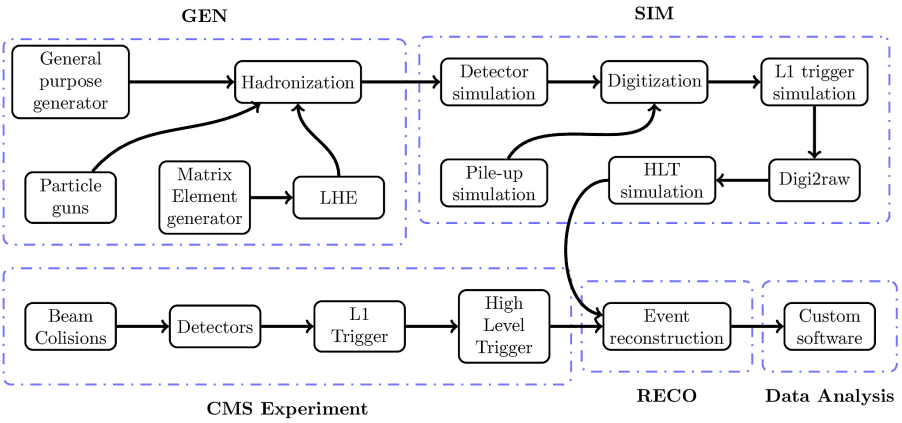


Fig. 3.2. Steps of the event simulation and event reconstruction within the CMSSW framework [60].

account for emissions of soft and collinear partons from the final or initial state, respectively. It is modelled as consecutive $1 \to 2$ and $2 \to 3$ splittings as a Markov chain. Afterwards, hadronization describes the transition from colour-charged partons to colour-neutral hadrons, which subsequently undergo decays. The underlying event (UE) is the activity in the event that is not associated with the hard parton interaction. It consists of the beam remnants and multi-parton interactions (MPI) in the event.

The strong coupling constant α_s decreases with the momentum transfer of the given interaction, the fact called running of α_s. Due to this fact, the large momentum transfer stages of the event generation can be calculated in perturbative series in α_s, while low momentum transfer stages rely on non-perturbative heuristic models. The two regimes are connected using parton showers.

Different general-purpose event generators such as PYTHIA 8 [52], [53], HERWIG 7 [54], [55], and SHERPA [56] are able to perform calculations of the full event generation but consist different formulations of parton showers, hadronization and other processes. Specialized event generators exist, such as POWHEG [57], [58] and MC@NLO [59] that are used to perform the generation of the hard process beyond the leading order (LO) in the strong and electromagnetic coupling constants and consistently merging these hard processes with parton showers without double counting partons.

The chain of physics event generation described ends with stable final-state particles. In the CMS definition, these represent particles that are long-lived enough to reach the first layers of the tracking detector of CMS, i.e. their proper lifetime, τ, is such that $c\tau > 1$ cm. To be able to compare the MC data with the reconstructed collision data on the same footing, MC data have to go through the detector simulation and the same reconstruction that is performed on the real data. The summary of the steps performed on MC and collision data are shown in Fig. 3.2. Event generation does not always have

to start from the general-purpose event generator or matrix element (ME) generator but can also start from a particle gun where single particles of specific energies are shot to the detector. Within the CMS software (CMSSW) framework the output from each event generator is converted into a common data format stored in ROOT [61] files. The SIM step goes through substeps of simulating the particle interactions in the detector using the GEANT4 programme [62]–[64]. Then, the conversion of the simulated hits to the digital signal in the detector is done and the trigger simulation is performed.

3.1. Comparison of the common SHERPA sample with the common POWHEG+PYTHIA 8 sample and with the ATLAS and CMS data

The standard MC samples for $t\bar{t}$ production in both CMS and ATLAS collaborations use POWHEG-Box-v2 implementation of the hvq process [57], [58] interfaced with PYTHIA 8. For PYTHIA 8, the parton shower, hadronization, and UE are tuned by each of the experiments. Some settings, including the α_s value, are different among the nominal ATLAS and CMS samples and lead to differences in the predictions. An effort within the LHC Top Working Group (LHCTopWG) [65] was dedicated to agreeing on common MC settings for both POWHEG+PYTHIA 8 and SHERPA. These settings are to be used in ATLAS+CMS combinations[2] and for comparison of measurements performed with slightly different event cuts, especially if one of the experiments sees a result that is not seen in the other.

The POWHEG+PYTHIA 8 MC sample is beneficial as it offers a small number of negative weights and the generation of $t\bar{t}$ at NLO in α_s. However, SHERPA $t\bar{t}$ MC can be a better choice because it allows one to switch between two common types of hadronization models, i.e. Lund hadronization and cluster-based AHADIC model. This enables a comparison of these models and derivations of uncertainties due to their differences. Additionally, SHERPA allows the generation of a high multiplicity of hard jets and has implemented electroweak (EW) corrections as weights.

Before the study described in this thesis, there had been no centrally generated $t\bar{t}$ SHERPA sample in CMS available for hadronization studies and experimental measurements. In contrast, ATLAS had already produced a SHERPA $t\bar{t}$ sample. The ATLAS settings were deemed a suitable starting option for the common CMS-ATLAS SHERPA sample. The aim of the author of this thesis was to generate a SHERPA sample within CMSSW and validate it with respect to a sample generated using the same settings within the ATLAS software, called ATHENA. In addition, the sample was also validated with respect to particle-level data measured by ATLAS and CMS experiments at $\sqrt{s} = 13\,\text{TeV}$ during Run 2 of the LHC. A comparison was also made with respect to the POWHEG+PYTHIA 8

[2] In the recent LHC top mass combination [66], for example, it was necessary to assume 50 % correlation between the MC samples of CMS and ATLAS.

setup. Although it is possible to run a standalone SHERPA installation within CMSSW, interfacing it to the CMSSW framework has the benefit that further steps of the event simulation can be run easily and consistently.

3.2. SHERPA sample validation

This section explains the results of the CMS-ATLAS MC $t\bar{t}$ sample project to which the author of this thesis was the main contributor from the CMS side, i.e. the CMS-ATLAS common SHERPA sample. The results have been published by the ATLAS and CMS experiments in a common note [38], [39].

It was found impossible to implement the same random number scheme in ATHENA and CMSSW, thus the comparison needed to be made within the statistical uncertainties of each sample. The validation was performed using two routines in the Rivet tool [67]: `MC_TTBAR` (`ONELEP`) and `MC_FSPARTICLES`. The former is particularly used for comparing $t\bar{t}$ predictions, while the latter shows characteristic distributions for all final-state particles with $|\eta| < 5.0$ and $p_T > 500$ MeV. In this way `MC_FSPARTICLES` is suitable for validation of the parton shower that produces most of these multiplicities of particles.

Fig. 3.3 shows observables from these two Rivet analyses. The uncertainty bars show only the statistical uncertainty to facilitate the evaluation of the compatibility between the two samples. Good agreement within the statistical uncertainties is seen in all observables, except N_{ch}, where a slight slope is visible, predicting that CMS SHERPA produces a slightly smaller number of charged particles than the ATLAS sample. A similar trend was also observed in the POWHEG+PYTHIA 8 sample [38], [39], leading to the conclusion that it could be an issue of the CMS and ATLAS frameworks. One possible reason for this could be that the CMS framework loses momentum precision of soft particles when storing them in the data files, e.g. because of the different floating-point precision. In addition, MC generators usually generate the interaction vertex at the same coordinate for all events, which is smeared afterwards by the reconstruction frameworks to match that seen in the data. It was not possible to verify whether this is done in the ATHENA framework at this or at the subsequent stages. The difference is only seen for the softest particles and is not seen to impact any of the relevant distributions.

The common SHERPA dataset was validated by comparing it with data, the POWHEG+PYTHIA 8 sample generated using the common ATLAS-CMS settings as well as the ones generated using the default ATLAS and CMS settings of each generator. The comparison encompassed Rivet routines containing ATLAS and CMS particle-level data obtained at $\sqrt{s} = 13\,\text{TeV}$.

Fig. 3.4 shows a comparison of H_{T} and N_{jets} distributions obtained in the CMS analysis of $t\bar{t}$ differential cross section in the resolved l+jets channel [68]. It is visible that SHERPA tends to produce a smaller number of jets that are also less hard while POWHEG+PYTHIA 8 tends to produce a larger number of jets that are harder. This presents a common trade-of

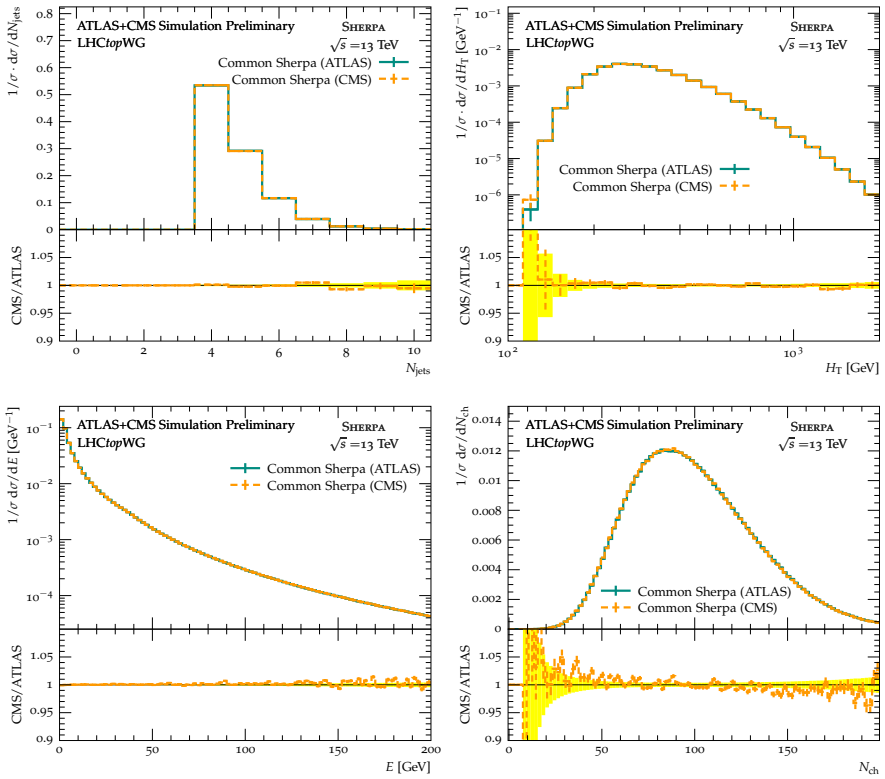


Fig. 3.3. Comparison of events produced with the common SHERPA settings within ATLAS (cyan) and CMS (orange) frameworks. The number of jets (top left), the scalar sum of the jet transverse momentum (top right) is shown, obtained using the Rivet routine `MC_TTBAR` (`ONELEP`), as well as the energy of all the final-state particles with $|\eta| < 5.0$ and $p_T > 500$ MeV (bottom left), and the number of charged particles per event N_{ch} (bottom right) is shown, obtained using the Rivet routine `MC_FSPARTICLES`:. Plots are published in [38], [39].

in tuning these two observables in NLO generators. While SHERPA improves the agreement with the data for H_T, the agreement for N_{jets} becomes slightly worse than for POWHEG+PYTHIA 8. A similar trade-of was observed for other Rivet routines and observables [38], [39].

3.3. Summary

In this section, the author's contribution in generating a common CMS-ATLAS SHERPA $t\bar{t}$ MC sample and the first CMS SHERPA $t\bar{t}$ sample was presented. The SHERPA sample was found to provide an alternative to the commonly used POWHEG+PYTHIA 8 sample, giving better agreement with the unfolded data for distributions such as top and

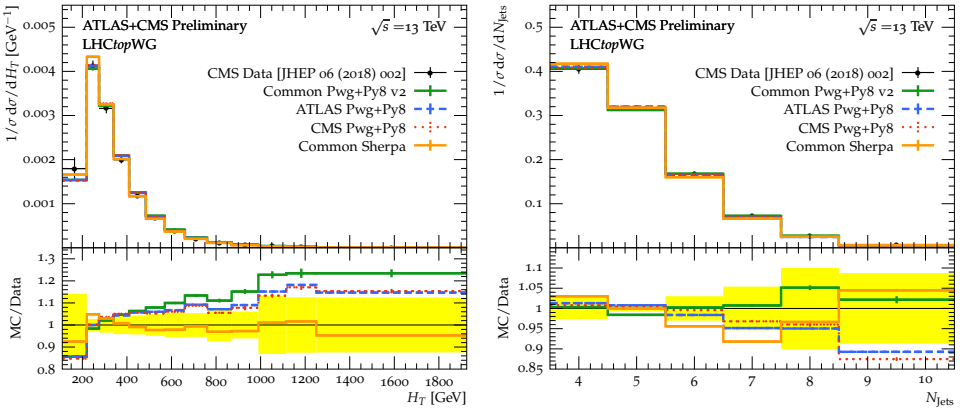


Fig. 3.4. Comparison of the CMS-ATLAS common MC SHERPA settings with the common and default ATLAS and CMS POWHEG+PYTHIA 8 settings. The distribution of the sum of the jet transverse momentum scalar (left) and the number of jets (right) is shown compared to the CMS data. Plots published in [38], [39].

jet p_T, while being slightly worse on others like N_{jet} distributions. In this way, the aim No. 4 is fully fulfilled. It is encouraged by the LHCTopWG to add the predictions of the common SHERPA sample to particle-level distributions of all $t\bar{t}$ analyses in CMS and ATLAS to simplify their comparison between experiments and result combinations.

A potential way to continue this project is to repeat the generation of the common sample after incorporating a new cluster hadronization model tune in SHERPA [69]. Validated SHERPA settings in CMS enable additional studies of the effect of the hadronization model on jets, for example, by performing studies of the jet energy scale as done by ATLAS [70]. Such studies would be an important next step of the study presented later in Chapter 6.

4. EXPERIMENTAL SETUP

4.1. The Large Hadron Collider

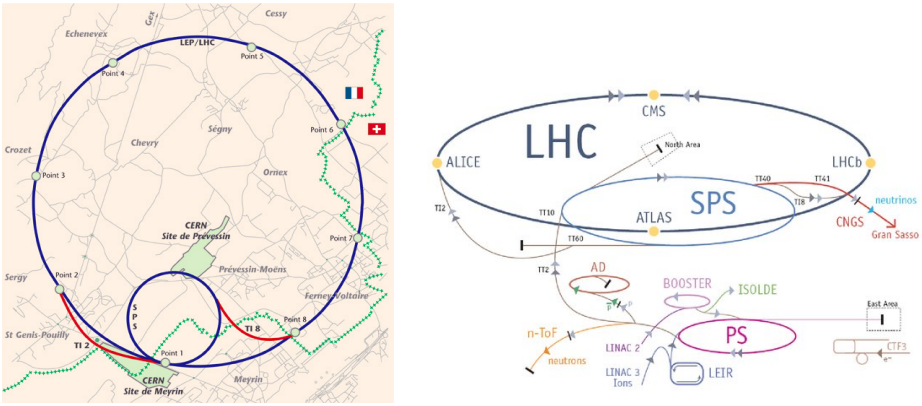


Fig. 4.1. Left: A map of the LHC and super proton synchrotron (SPS) tunnels (in blue) in Geneva together with the transfer lines (red) [71]. The border between France and Switzerland is marked with the green dotted line. The centres of the eight straight sections with access to the accelerator and where the inner and outer beampipes are interchanged are marked by grey points. Right: A scheme of the European Organization for Nuclear Research (CERN) accelerator complex and the main experiments at CERN benefiting from it [72].

The measurement presented in this Thesis was performed using data collected by the CMS experiment at the Large Hadron Collider (LHC). As the world's most powerful particle accelerator, the LHC is capable of accelerating two counter-rotating proton beams to a maximum energy of 7 TeV each and lead ions up to 2.8 TeV [73], [74]. The map of the LHC, and the smaller super proton synchrotron (SPS), tunnel is shown in Fig. 4.1 Left. The LHC was installed in the tunnel formerly built for the Large Electron–Positron Collider (LEP) located at a depth between 45 m and 170 m below ground [75].

The proton acceleration process involves a series of carefully synchronized stages, in which particles pass through multiple accelerators before reaching their final energy. The full CERN accelerator complex, shown in Fig. 4.1 Right, feeds particles into the LHC while also supporting other cutting-edge experiments.

4.2. The Compact Muon Solenoid

CMS is a hermetic multipurpose experiment [77]. Here, hermetic indicates that the detector covers an extensive phase space, as the CMS hadronic calorimeter (HCAL) reaches pseudorapidity of up to $|\eta| = |-\ln[\tan\left(\frac{\theta}{2}\right)]| = 5.2$, where θ-coordinate represents the polar angle from the beam axis. Originally designed to discover the Higgs boson and understand

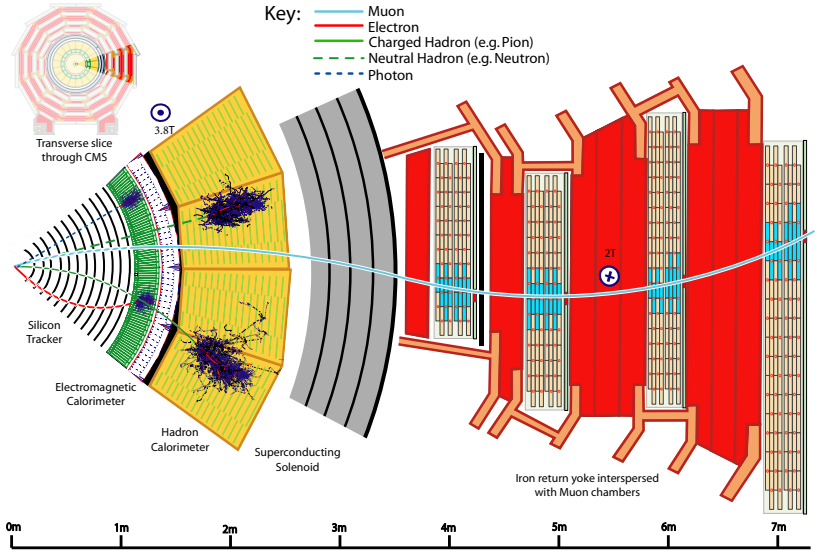


Fig. 4.2. A sector of a cross section of the CMS detector along the x-y plane (i.e. taken perpendicularly to the beamline) [76].

its properties, CMS is well suited to carry out a variety of precision physics measurements, as well as searches for new physics. CMS has a cylindrical shape, is azimuthally symmetric, 21 m long, and 15 m in diameter. A sector of a slice in the transverse direction of the CMS detector is shown in Fig. 4.2, together with example trajectories of particles traversing it. It illustrates the layered structure of CMS with the first layers, pixel and strip tracking detector (tracker) responsible for particle tracking. It is followed by the calorimetric layers, where the electronic calorimeter (ECAL) is responsible for absorbing the energies of photons and electrons, while the HCAL does it for hadrons. After that the magnet is responsible for bending the charged particles to distinguish their charge and p_{T}. Finally, the muon chambers are dedicated to the tracking of muons, which do not get absorbed in the calorimeters and the magnet.

CMS uses a right-handed coordinate system as shown in Fig. 4.3 where the z-axis points in the direction of the counterclockwise beam, the x-axis points to the centre of the LHC, and the y-axis points vertically upwards. In addition, cylindrical coordinates (r, ϕ, z) and spherical coordinates (r, θ, ϕ) are often used, where the r-coordinate represents the radial distance from the z-axis. The ϕ-coordinate represents the azimuthal angle around the beam axis. The polar angle θ is often replaced by η.

The detector is divided into the barrel, endcap, and forward regions. The barrel region contains cylindrical subdetectors and covers up to $|\eta| < 1.3$. The endcap region covers the sides of the CMS cylinder and is divided into a region with the CMS tracker coverage, $1.566 < |\eta| < 2.5$, and with no tracker coverage, $2.5 < |\eta| < 3$. This leaves a

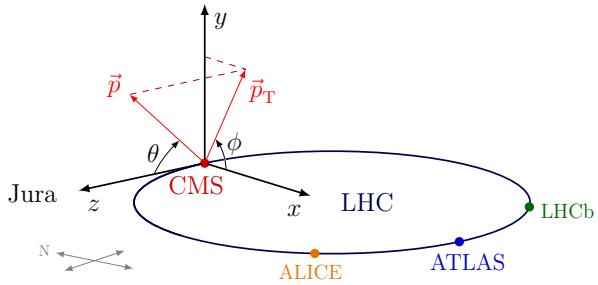


Fig. 4.3. The coordinate system used in CMS [78].

transition region of $1.3 < |\eta| < 1.566$ with a larger material budget due to the cabling and limited detector coverage. The forward region with the forward HCAL is defined as $3 < |\eta| < 5.191$.

5. EVENT RECONSTRUCTION IN CMS

Once a signal is detected, the raw detector response, recorded as hits and deposited energy, has to be transformed into information in terms of physics objects, such as particles, their momenta, trajectories, and types. This is done in the event reconstruction.

The first step in reconstruction is tracking, where individual particle trajectories are identified from hits in the tracking detectors [79]. After tracks are reconstructed, the primary vertices (PVs) are determined by extrapolating the tracks and fitting the position of the track origin. The PV with the highest track p_{T} pointing to it is identified as the leading PV, representing the primary interaction point of the hard scattering process.

Physics objects are reconstructed using the particle flow (PF) algorithm that combines the information from all the subdetectors [76]. Muons are reconstructed using the tracks in the muon chambers and the inner tracker. Electrons are identified as superclusters of bremsstrahlung deposits in the ECAL matched to corresponding tracks, while photons are found by similar superclusters without associated tracks. Hadronic activity is classified based on energy deposits in the HCAL: clusters without associated tracks are tagged as neutral hadrons, while those linked to tracks are identified as charged hadrons.

After individual particles are reconstructed, collimated sprays of hadrons, jets, are clustered from groups of PF candidates. After that, reconstruction continues with the calibration for the energies of the physics objects and improving the resolution. Additionally, data-to-MC scale factors (SFs) accounting for differences between the simulation and data are determined. The following sections introduce jet clustering and energy calibration in more detail.

5.1. Jet clustering and identification

The colour confinement and large logarithms appearing for the matrix elements of soft and collinear emissions of quarks and gluons cause the creation of a large multiplicity of hadrons[3] collimated in narrow cones, as depicted in Fig. 5.1. Precise treatment of such large multiplicities of particles is complicated both theoretically (e.g. due to limited order calculations in coupling constants and parton shower approximation) and experimentally (e.g. due to computational power required to track the large number of particles and the limited detector resolution). This complexity can be efficiently reduced by combining the particles into jets originating from an initial colour-charged particle.

Jet clustering algorithms assign final-state particles to jets. The jet clustering algorithm used in this Thesis is an anti-k_{T} algorithm that creates conical jets of a radius taken as $R = 0.4$ [80], [81]. In the MC simulation, it is possible to assign the given jet a flavour of the parton that initiated it. In this Thesis, it is done using the parton flavour algorithm. This algorithm repeats the jet clustering from a list of hadrons, which is extended

[3] This includes also other particles, like leptons arising from the decays of unstable hadrons.

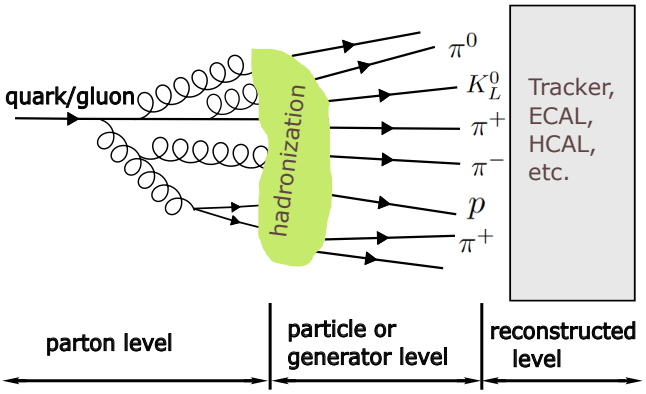


Fig. 5.1. Schematic picture of a jet. MC generators allow to distinguish jets clustered from different types of objects. Parton level denotes jets clustered from partons. Particle-level jets are clustered from stable hadrons. Reconstructed jets are obtained by clustering reconstructed objects like energy deposits in the calorimeters or PF candidates.

by adding all the final partons with their momenta scaled to a negligible value. The jet is assigned a bottom (b) flavour if there is a b parton clustered inside this jet and charm (c) flavour if there is a c parton but no b parton inside. Otherwise, the jet adopts the flavour of the hardest parton clustered inside.

For both MC simulation and data, b and c jet identification (b and c tagging) is possible from the reconstructed track information [82]. This relies on such information as the existence of a secondary vertex (SV) within the jet. Three working points (WPs) are created, "loose", "medium", and "tight", which correspond to the probabilities to misidentify a light-quark jet as a b jet of 10 %, 1 % and 0.1 %, respectively.

5.2. Jet energy calibration

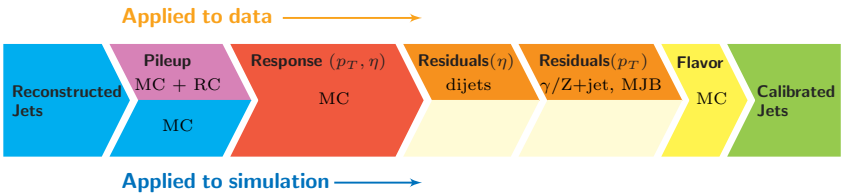


Fig. 5.2. jet energy corrections (JEC) scheme in CMS Runs 1 and 2 [83]. Flavour MCs are optional but used in the Δm_t analysis. The top shows the corrections applied to data, and the bottom shows the corrections applied to MC.

The reconstructed jets require calibration to remove effects of pileup (PU), the losses in the detector, e.g. due to dead material and the detector non-compensation, and UE

contaminating the jet cone. In CMS, the scheme to correct jets in Run 2 is shown in Fig. 5.2. It is factorised in several steps. First, the PU contribution is removed using the charged hadron subtraction (CHS) and hybrid area method. After that, the detector effects are compensated for. These response corrections are first obtained from MC simulation (MC-truth corrections), ensuring that the p_{T} of reconstructed jets matches the p_{T} of particle-level jets. Since $p_{T,\mathrm{reco}}$ is smeared by detector resolution, the jet energy response $R = \frac{p_{T,\mathrm{reco}}}{p_{T,\mathrm{ptcl}}}$ is a distribution similar to the one shown in Fig. 5.3. The MC-truth corrections thus shift the jet energy scale (JES), namely the mean (or median) of this distribution, to 1. After this, the residual data and MC corrections are obtained using data-driven corrections relying on p_{T}-balance in dijet and Z+jets events.

After all the mandatory CMS JECs described above, optional flavour-dependent JEC (flavour corrections, L5 corrections) obtained from the MC can be applied. These take into account the fact that different flavour jets have different fragmentation patterns and behave in the detector differently. In addition, flavour uncertainties account for the mismodelling of each individual jet flavour. These uncertainties are mandatory in CMS for all analyses that use jets. The flavour corrections and flavour uncertainties will be discussed in Chapter 6. All the steps mentioned above come with a separate set of uncertainties.

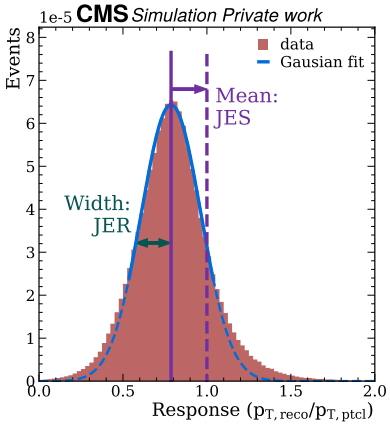


Fig. 5.3. Jet energy response distribution showing its mean, JES, and the width, jet energy resolution (JER).

6. FLAVOUR-DEPENDENT JET ENERGY CORRECTIONS AND QUARK-ANTIQUARK RESPONSE ASYMMETRY

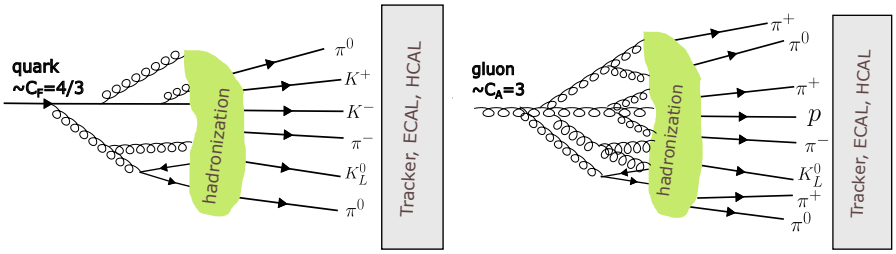


Fig. 6.1. Schematic picture of the differences between a quark jet (left) and a gluon jet (right). Gluon jets have a larger particle multiplicity and a wider jet radius. The hadron content between the jets also differs, for example, jets initiated by strange quarks have a higher probability of producing kaons (strange hadrons).

As described in Section 5.2, jets clustered from particle-level constituents and from reconstructed particles differ. In CMS, these effects are typically accounted for using JEC obtained from the QCD dataset dominated by gluon jets. Nonetheless, the response varies depending on the flavour of the jet due to differences in jet fragmentation [84]. Response-varying effects include hadron content differences [70], the momentum spectrum of the constituent hadrons, and the shape of the jet [85]. The differences in jet flavour are schematically depicted in Fig. 6.1. For instance, light quark jets (u, d) tend to exhibit the highest energy response, while gluon jets present a comparatively lower response. This is due to the larger gluon Casimir colour factor leading to gluons fragmenting into a greater number of softer particles, with a wider spread in the detector. This can also cause the deposits in some of the calorimeter cells to fall below the detector thresholds and thus reduce the energy response. Moreover, the potential mismodelling of the jet fragmentation across different flavours leads to flavour uncertainty.

The necessity for flavour-dependent JEC (flavour corrections) is to facilitate the estimation of the flavour uncertainties and to bring into agreement the gluon-jet-dominated QCD dijet (QCD) and quark-jet-dominated Z+jet samples used to derive the data-driven corrections. In addition, analyses relying on b jet identification, such as the top quark mass measurement, would benefit from the correction of the b jet to light quark jet response. Such a correction is applied in the Δm_t measurement as described in Subsection 7.2.1. On the contrary, the flavour uncertainty is necessary for all the analyses using jets in CMS. The uncertainty associated with b jets has emerged as the dominant uncertainty in some analyses, such as the top-quark mass measurement [86], while the gluon jet uncertainty is one of the leading in measurements of α_s [87]–[90] or cross-section measurements of the $t\bar{t}$ + jet process [91].

Flavour corrections are usually derived using MC-truth methods as an extension of the inclusive JEC derivation procedure described in Section 5.2. The difference is that for flavour corrections, the sample of jets is split according to the jet flavour. Data-driven techniques [92]–[95] also have been developed. Although promising, these methods typically suffer from large systematic uncertainties or require additional tests. As a result, data-driven methods have so far primarily served only as a validation of the MC-driven results.

The previous CMS flavour corrections were obtained using a QCD sample generated by PYTHIA 6 (Z2* tune), while the flavour uncertainties were obtained from the differences against a HERWIG++ (EE3 tune) QCD sample [83]. Additionally, a private set of corrections using HERWIG++ and PYTHIA 8 was derived with early Run 2 samples [96]. In this study, we repeat the work using PYTHIA 8 (CP5 tune) for the flavour corrections and a comparison to HERWIG 7 (CH3 tune [97]) for determining the flavour uncertainties. Instead of only using the QCD sample, we do simultaneous fits of the QCD, DY (or Z+jets), and $t\bar{t}$ samples, in addition to a fit of the $t\bar{t}$ sample individually. Alternative MC-truth methods for the estimation of jet flavour uncertainties exist, for example, by splitting the flavour uncertainty into three components: hadronization, shower and UE+shower as done by ATLAS [98]. In this study, however, we continue the approach previously used in CMS, estimating the flavour uncertainty through differences between HERWIG (angular-ordered shower + cluster hadronization) and PYTHIA (p_{T}-ordered dipole shower + Lund string hadronization).

This section contains the main results obtained by the author of this Thesis. First, the flavour corrections are shown in Section 6.1, then, the flavour uncertainties in Section 6.2, and finally, the flavour-antiflavor uncertainties are obtained as explained in Section 6.3. The results are useful for the entire CMS collaboration, and their application to the Δm_t measurement is shown in Chapter 7. The results are presented in a CMS internal note [40].

6.1. Flavour dependent jet energy corrections

Flavour corrections were obtained from the $t\bar{t}$ and QCD datasets individually and from a simultaneous fit of QCD, $t\bar{t}$ and DY datasets. The corrections derived using the PYTHIA 8 QCD dataset (Summer20) were compared to the corrections used in Run 1 (Winter14) obtained from the QCD sample generated using PYTHIA 6 with the Z2* tune [83]. In addition, an earlier set of corrections using an independent framework was derived in Ref. [96] on the QCD dataset generated using PYTHIA 8 (Autumn18). These corrections rely on an older CMS MC campaign, which contains earlier versions of JEC that were less fine-tuned than in the UL campaign. For this comparison, the Summer20 corrections are derived in the same $|\eta|$ binning as the Autumn18 corrections. The Winter14 fits were derived in a coarser binning corresponding to HCAL sections.

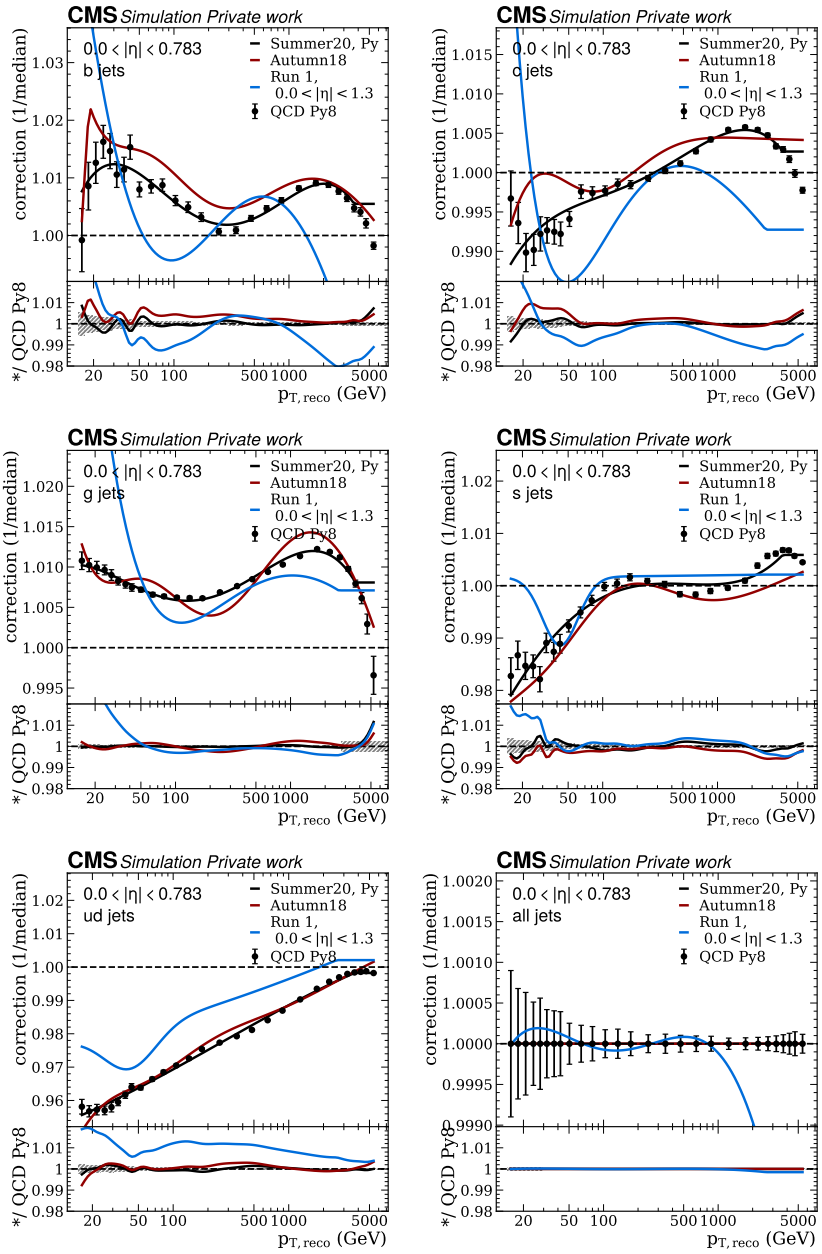


Fig. 6.2. Inverse median of the jet energy response in the $0 < |\eta| < 0.783$ bin for six different jet flavours for the QCD sample, generated using PYTHIA 8 (black markers). The fits of the QCD sample are shown (Summer20, black curve) together with the Autumn18 (early Run 2, red curve) and Winter14 (Run 1, blue curve) corrections.

Fig. 6.2 compares the new Summer20 with the Autumn18 and Winter14 fit results of the JEC factor defined as the inverse median jet energy response. The results are shown in the most central $|\eta|$ bin. The legend displays the corresponding $|\eta|$ bin used for the Winter14 fit. It shows a reasonable agreement between the Summer20 and Autumn18 fits.

The Winter14 corrections show a significant disagreement with the other two fits. The main reason for this is the intrinsic differences between PYTHIA 6 with the Z2* tune and PYTHIA 8 with the CP5 tune, such as UE parameters and the α_s value in the parton shower. Moreover, the number of events for the PYTHIA 6 sample used for the Winter14 fits (around 10 million events) was significantly smaller than in the PYTHIA 8 QCD dataset (around 20 million events). This is also seen by the larger statistical fluctuations in the Winter14 fits. In addition, the Winter14 fits were derived on jets clustered with $R = 0.5$ instead of $R = 0.4$ and using different binning in $|\eta|$ than the Summer20 fits. Finally, the Winter14 fits used a different classification method for parton-level jets, called the physics definition, while in Summer20 and Autumn18, the parton flavour was used. Despite the differences between the PYTHIA 6 and PYTHIA 8 flavour corrections, they show the same global trends. Namely, b and g jets have a lower response, generally $R < 1$, while light jets have a larger response with $R > 1$.

6.2. Flavour uncertainties

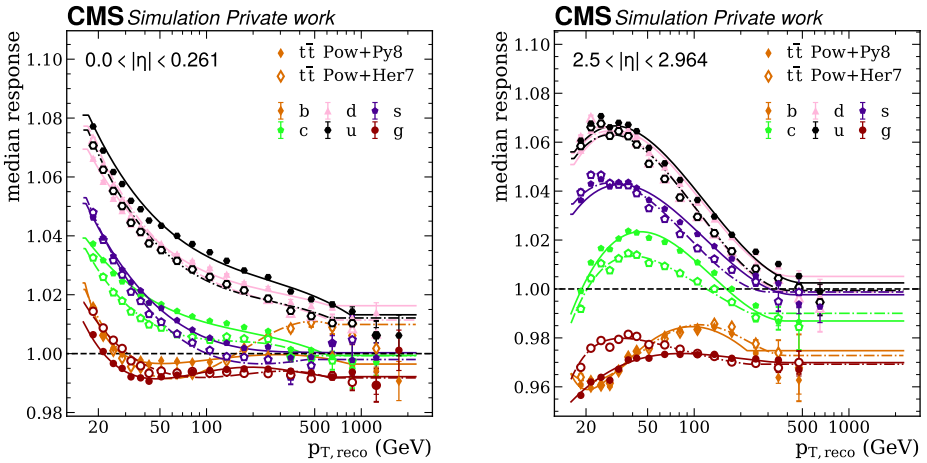


Fig. 6.3. Comparison of the median jet energy response predicted by PYTHIA 8 (closed markers) and HERWIG 7 (open markers) showers for different jet flavours. The inverse of the JEC fits are shown as curves. Results are shown for two different η regions.

In Fig. 6.3 the HERWIG 7 and PYTHIA 8 comparison for the median jet energy responses in the $t\bar{t}$ sample, i.e. the inverse of the correction, is shown. The general trends

are the same as for the results of the QCD fit shown in Fig. 6.2. One can see that light flavours need the largest correction and gluon jets need the smallest, while the rest of the flavours are in-between the two. The ordering and overall trend between different flavours is consistent with what was seen in the Run 1 analysis [83]. Some differences between the PYTHIA 8 and HERWIG 7 datasets are observable. The $t\bar{t}$ flavour uncertainty is proportional to these differences but normalized to a reference point of $p_T = 200\,\text{GeV}$ in the Z+jets flavour composition, to which the data-driven residual corrections are calibrated.

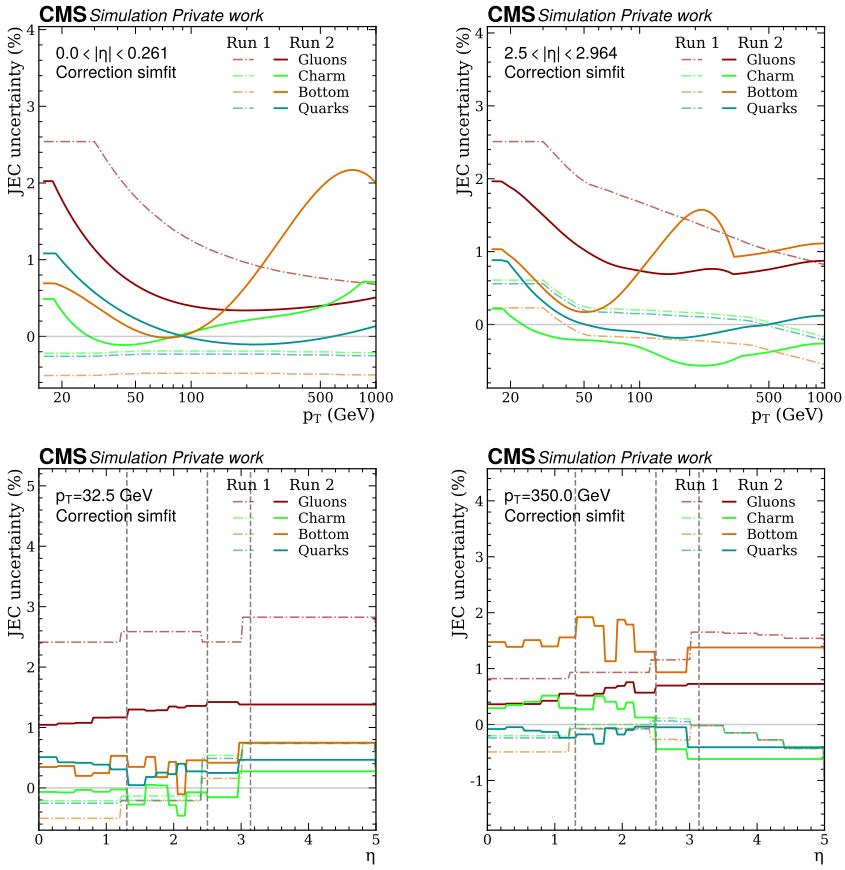


Fig. 6.4. Flavour uncertainty as a function of p_T for two different η regions (top), and as a function of η for two different p_T regions (bottom). The dash-dotted lines show the flavour uncertainties from Run 1. The solid lines show flavour uncertainties for Run 2.

In Fig. 6.4, the flavour uncertainty for different flavours is compared to the flavour uncertainty of Run 1. For gluon jets. the obtained uncertainty is significantly smaller than in Run 1, and for other flavours, the uncertainty in some regions of p_T becomes slightly larger, while decreasing in others. As in Run 1 gluon jets were the dominant source of the uncertainty, and due to the limited number of simulated MC events, the rest of the

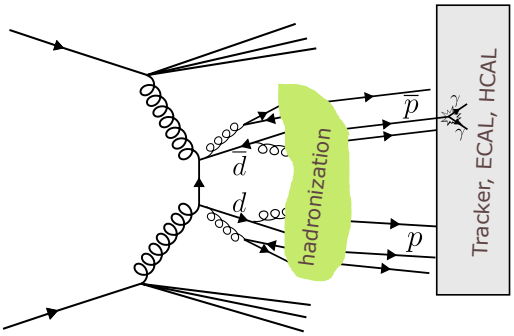


Fig. 6.5. Schematic picture of an event with a quark and antiquark created. The difference in the hadron evolution leaves a different impact in the detector.

flavours in Run 1 were fitted with a very approximate fit, which often was a simple straight line. This often leads to an underestimation of the uncertainty. For Run 2, we now fit all flavours from a larger number of generated MC events, and an improved dependence over p_T is obtained. The uncertainty for b jets reaches 2.5 % at $p_{\mathrm{T}} > 100\,\mathrm{GeV}$, compatible with the 3 % in the Autumn18 PYTHIA 8/HERWIG++ analysis [96]. The potential cause of this increase was found to be differences in the modelling of b hadron decays in HERWIG 7 and PYTHIA 8, as further discussed in the Doctoral Thesis.

6.3. Flavour-antiflavour uncertainty

In precision measurements of potential differences in the properties of matter and antimatter, such as the Δm_t measurement, even small response differences between quark and antiquark jets are important. Such differences can arise from different hadron and antihadron content in flavour and antiflavour jets as depicted in Fig. 6.5. These hadrons and antihadrons interact with the detector material that is made out of matter, where the antihadrons lead to a slightly larger energy deposit in the calorimeters. In this section, the author presents two sources for flavour-antiflavour uncertainties.

The first source of the uncertainty, shown in Subsection 6.3.1, covers potential mismodelling of the flavour vs antiflavour jet content and fragmentation and decays by the parton showers. This is obtained by a comparison of the HERWIG 7 and PYTHIA 8 parton showers in a similar way as the flavour uncertainties were obtained. The second uncertainty, presented in Subsection 6.3.2, is the mismodelling of the single-pion responses in the GEANT4 simulation in comparison to the HCAL test beam data.

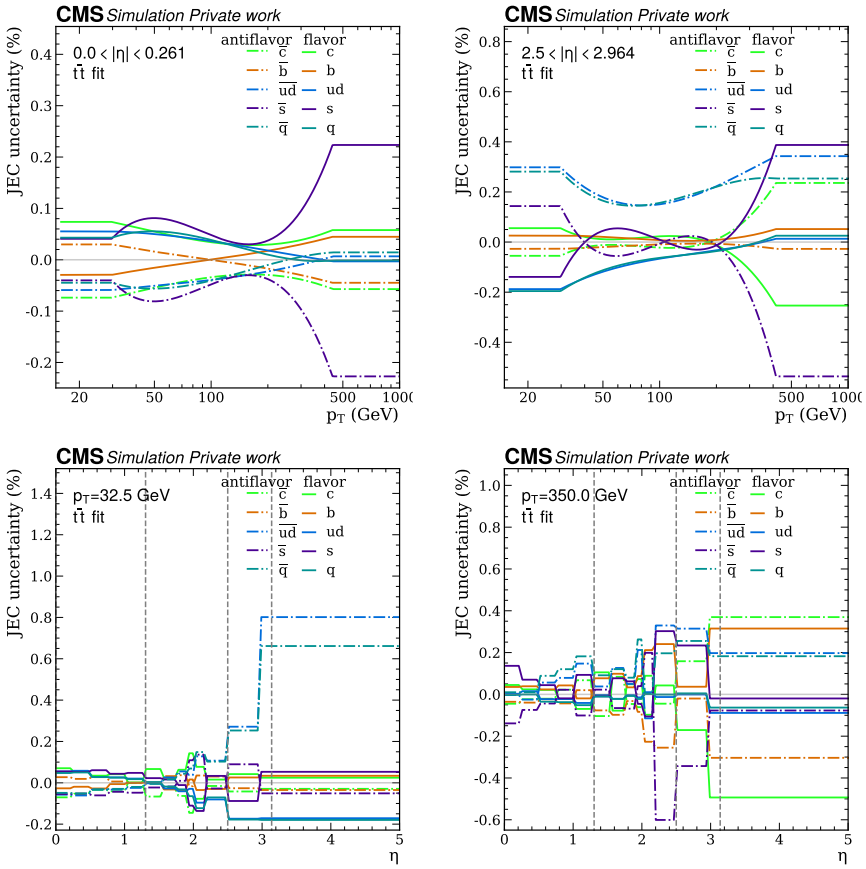


Fig. 6.6. Flavour vs antiflavour uncertainty as a function of p_{T} for two different η regions (top) and as a function of η for two different p_{T} regions (bottom).

6.3.1. Flavour-antiflavour uncertainties from the parton shower modelling

The flavour uncertainties were estimated by subtracting the PYTHIA 8 fit of the median jet energy response from the HERWIG 7 fit for each quark-antiquark combination. The resulting uncertainty curves can be seen in Fig. 6.6. The dominating uncertainty, especially at the lowest η, is the s vs $\overline{s}$ uncertainty. Due to the u/d flavour fraction dominating over $\overline{u}/\overline{d}$ at large p_T and η, the $\overline{u}/\overline{d}$ uncertainty is much larger than the u/d uncertainty. Thus, at large p_T and $|\eta|$, where the fraction of $\overline{u}/\overline{d}$ jets is small, the uncertainty for $\overline{u}/\overline{d}$ becomes larger than the u/d uncertainty. The b vs $\overline{b}$ uncertainty changes sign with increasing p_T and $|\eta|$, and is below 0.03 % in the range of $50 < p_{\mathrm{T}} < 100\,\mathrm{GeV}$, where the largest amount of b jets in $t\bar{t}$ events are created. The results explained stem from a large disagreement between the predictions of the s jet response by HERWIG 7 and PYTHIA 8 showers and a small disagreement in the predictions of the b jet response.

For the Run 1 measurement of Δm_t [30], the b vs $\bar{b}$ uncertainty was estimated by taking the total difference in responses between b and $\bar{b}$ in the same shower, which at that time was PYTHIA 6. A constant shift of 0.078 % was observed. However, a good MC generator should model the actual b and $\bar{b}$ response differences as in data, while the potential mismodelling of the different flavour vs antiflavour responses can cause a bias in the Δm_t measurement. Thus, this analysis using the predictions of HERWIG 7 and PYTHIA 8 presents a more proper estimate of the flavour-antiflavour uncertainty. Note, how the flavour-antiflavour uncertainty is around 10 times smaller than the inclusive flavour uncertainty shown in Fig. 6.4.

6.3.2. Flavour-antiflavour uncertainties from simulated hadron response mismodelling

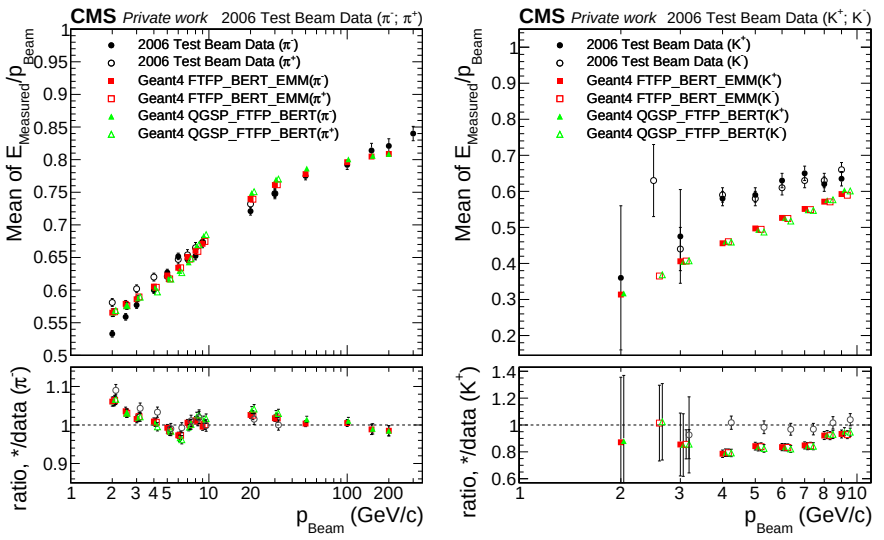


Fig. 6.7. The total ECAL and HCAL response from the test beam and the simulated response using two different physics lists: left – the results for $\pi^{\pm}$, and right – for $K^{\pm}$. Data are obtained from [99]. The particle gun simulation is performed using GEANT4 v.10.4.p03.

Before mounting the barrel HCAL and ECAL (EB, HB, and HO) into CMS, a test beam was performed to measure the properties of these subdetectors [99]. Responses to beams of $\pi^{\pm}$, p, $\overline{p}$, $K^{\pm}$, $e^{\pm}$, and $\mu^{\pm}$ were measured. The results for the test beam for $\pi^{\pm}$ and $K^{\pm}$ are shown in Fig. 6.7, together with a simulation of the detector response by GEANT4. In this study, we use GEANT4 version 10.4.p03 [62]–[64] together with the physics model configuration (physics list) `FTFP_BERT_EMM`, as used for MC simulation used in the Δm_t measurement. An alternative physics list `QGSP_FTFP_BERT_EMM` is used for comparison. The initial HCAL calibration had been done with 50 GeV electrons.

The results show a larger response for π^+ than for π^-. At beam energy $p_{\text{Beam}} = 2\,\text{GeV}$, this difference is almost 10 % and slowly decreases with increasing p_{Beam}. This difference is not described by GEANT4. Both of the physics lists used in this study predict approximately the same response of π^+ and π^-, agreeing with the π^+ data. This difference in data could be explained by different probabilities to create π^0 through charge-exchange reactions, $\pi^+ + n \to \pi^0 + p$ (I) and $\pi^- + p \to \pi^0 + n$ (II), taking place in the ECAL. π^0 quickly decays to two photons and in CMS calorimeters leaves around 20 % more signal than $\pi^\pm$. Since $PbWO_4$, the scintillator material in the ECAL consists of 42 % more neutrons than protons, the reaction (I) has a higher probability of taking place, explaining the increased response for π^+.

Another notable disagreement between the test beam results and the simulation of GEANT4 is in kaons. The response for both K^+ and K^- is larger than predicted. The agreement with π^- was found to improve towards GEANT4 version 10.7p01 used for the early Run 3 simulation [100]. On the other hand, the agreement with π^+ does not seem to improve but even deteriorates. Regarding kaons, it was shown that the agreement between the simulation and the particle gun data improves only slightly, and there is a remaining disagreement of up to 20 % in the response for $p_{\text{Beam}} > 3\,\text{GeV}$.

To account for the impact of the π^- and π^+ response mismodelling on the flavour-antiflavour JEC, the response of the PF candidates formed from negatively charged tracks was scaled down by a factor obtained from a ratio seen in Fig. 6.7. The effect of the response shifts on JEC is seen in Fig. 6.8, where the ratio of the jet energy response of the dataset corrected for the π^- mismodelling and the central dataset is shown. An overall impact of up to 0.30 % can be seen important for precise inclusive jet energy calibration, while for flavour and antiflavour, the impact is similar. Statistical uncertainties are reduced by a factor of 100 due to the correlations of the scaled and central responses. At low p_{T}, π^- scaling impacts s jets more than $\overline{s}$ jets, while at large p_{T} it is the other way around. A similar but opposite trend can be seen for u/d and $\overline{u}/\overline{d}$ jets. For b and c jets, no clear trend of the quark vs antiquark jet relation over p_{T} is visible. Therefore, it was decided to merge all the p_{T} bins and obtain the uncertainty as an overall scale factor for each of the four $|\eta|$ bins. The uncertainty is obtained as the difference of the central shift of the jet energy response for quark jets and for antiquark jets.

6.4. Summary and conclusions

The derivation of jet flavour corrections and uncertainties was presented. Unlike in Run 1, they are obtained from QCD and $t\bar{t}$ samples individually, and from a combined fit of QCD, $t\bar{t}$, and DY samples. The newly derived flavour corrections are derived in a finer binning than the Run 1 corrections and have lower statistical fluctuations. The differences are up to 2 % (for light quark, u/d, jets).

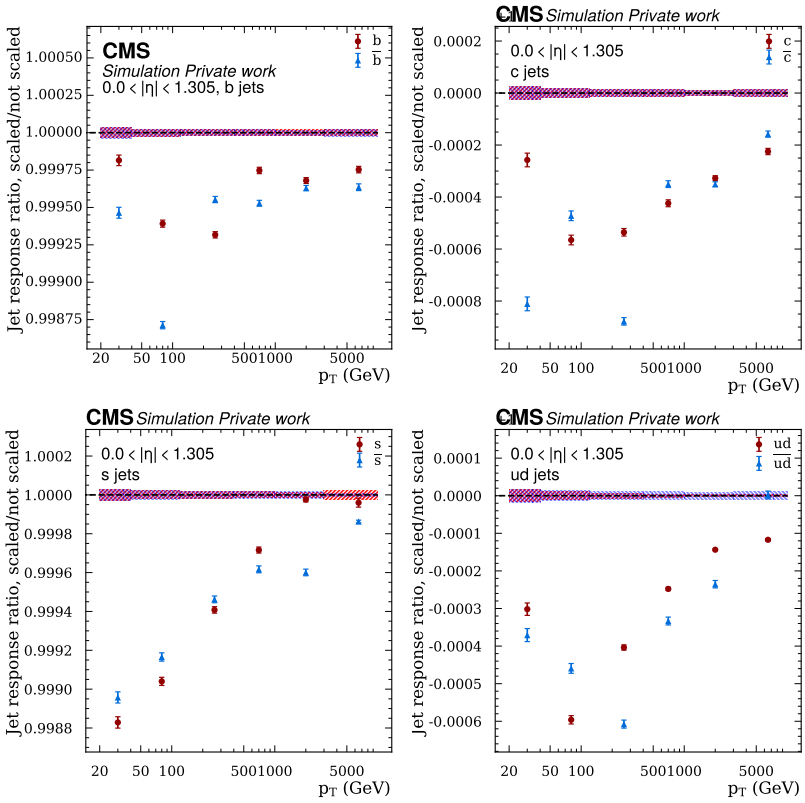


Fig. 6.8. Ratio of the jet energy response scaled by the π^- shift and the central simulation. The impact of the shift of the corresponding flavour and antiflavour are overlaid.

Compared to Run 1, flavour uncertainty for g jets is reduced by up to 50 %, while for other flavours it remained similar. For b jets, the flavour uncertainty has a rapid increase at large momenta, which is possibly due to the mismodelling of long-lived b hadrons. However, the exact reasons for the issue have yet to be determined. The flavour-antiflavour uncertainty takes into account two different effects:

1) the MC modelling of the flavour composition;

2) the effect of the disagreement between the hadron and antihadron responses of the GEANT 4 simulation compared to the test beam data.

7. THE MEASUREMENT OF THE MASS DIFFERENCE BETWEEN TOP AND ANTI-TOP QUARK

In this Chapter, the measurement of the mass difference between the top quark and antiquark, $\Delta m_t = m_t - m_{\bar{t}}$, using $t\bar{t}$ events in the lepton+jets decay channel is described.

7.1. Event selection

Table 7.1

$|\eta|$ and p_T requirements for electrons, muons, and jets

	Electrons	Muons	Jets
Max $\|\eta\|$ (2016)	2.4	2.4	2.4
Max $\|\eta\|$ $(2017 - 2018)$	2.5	2.5	2.5
Min p_T [GeV] (2016)	29	26	30
Min p_T [GeV] $(2017 - 2018)$	35	29	30

The event selection is targeting the $t\bar{t}$ lepton+jets decay channel and requires events to contain:

1) exactly one isolated lepton (electron or muon) with p_T and $|\eta|$ selection requirements, as shown in Table 7.1;

2) at least four CHS jets clustered with the anti-k_{T} algorithm with $R = 0.4$ and with selection requirements, as shown in Table 7.1;

3) at least two of the jets have to be classified (tagged) as b jets using the DeepJet algorithm, using the Medium WP.

The requirement of at least 2 b tagged jets is different from the Run 1 Δm_t analysis, where at least 1 b tagged jet was required. This stricter requirement improves the signal-to-background ratio and was made feasible by the increased amount of data in Run 2 compared to Run 1.

An event display of an example event that passes the event selections is shown in Fig. 7.1. The event is obtained from simulated $t\bar{t}$ events and shows four jets, two of which are b tagged, an isolated muon, and a large $p_{\mathrm{T}}^{\mathrm{miss}}$.

7.2. Event reconstruction

The purpose of event reconstruction is to match the reconstructed objects to the $t\bar{t}$ event hypothesis, correct the event kinematic variables, and select events with a higher signal-to-background ratio. First, reconstructed objects, that is, the lepton, $p_{\mathrm{T}}^{\mathrm{miss}}$, two b-tagged and two light jets are matched to the parton-level objects in the $t\bar{t}$ event shown

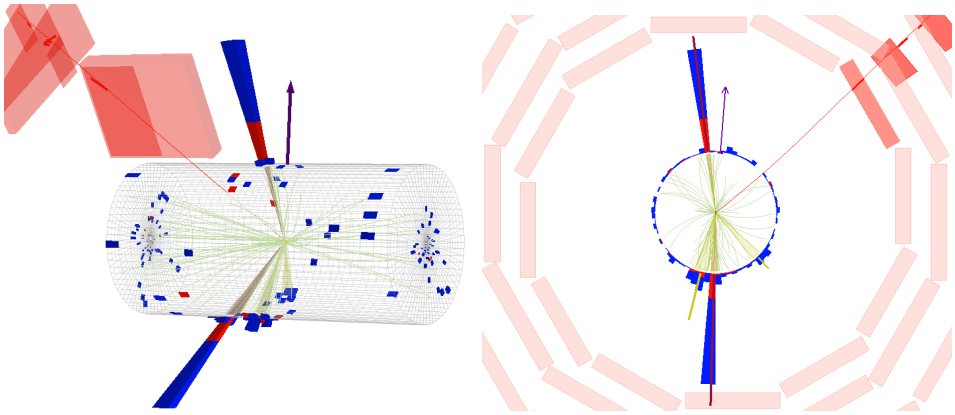


Fig. 7.1. An event display of a $t\bar{t}$ event candidate in MC simulation. The view from an arbitrary plane (left) and in the $\rho - \phi$ plane (right) is shown. The yellow (dark red) cones with same colour lines pointing out represent light (b tagged) jets. The red track represents muon going through the muon chambers. The red (blue) stacked blocks represent the energy deposits in the ECAL (HCAL). The violet arrow shows the direction of the $p_{\mathrm{T}}^{\mathrm{miss}}$. The green lines show the trajectories of the PF candidates.

in Fig. 2.2. The two leading light jets form the W boson candidate, while the b-tagged jets are assigned as b quark candidates.

There are two possible permutations of the b jet to b quark assignment. Based on MC truth information, the jet permutation is classified as "correct", "wrong" and "no match". Due to the small mistag rate of the b-tagging algorithm, the "wrong" category arises mostly from the cases where the four jets are selected correctly but the two selected b jets are assigned to the wrong b parton. The "no match" categorises permutations where the four jets do not match the partons arising from the $t\bar{t}$ event, which happens in cases where, for example, an ISR jet is hard enough to appear among the leading light jets. It should be noted that "wrong" and "no match" permutations are also sensitive to the generated m_t and thus are useful for the fit, though not in a straightforward way as the correct permutations.

Finally, $p_{\mathrm{T}}^{\mathrm{miss}}$ is assigned to the neutrino. The $|p_z|$ of the neutrino can be found by solving for $(p_\nu + p_l)^2 = (m_{\mathrm{W}}^{\mathrm{pdg}})^2$. This quadratic constraint can lead to two solutions, and an assumption of choosing one of the solutions has to be taken. Reasonable options include taking both solutions, taking the solution with the smallest $|p_z|$, the neutrino solution that is closer to the lepton $|p_z|$, the solution that maximises the cosine of the angle between the lepton and the W, or not solving and keeping $|p_z| = 0$. In this thesis, the solution that minimises the neutrino $|p_z|$ is chosen, that is, the most central one.

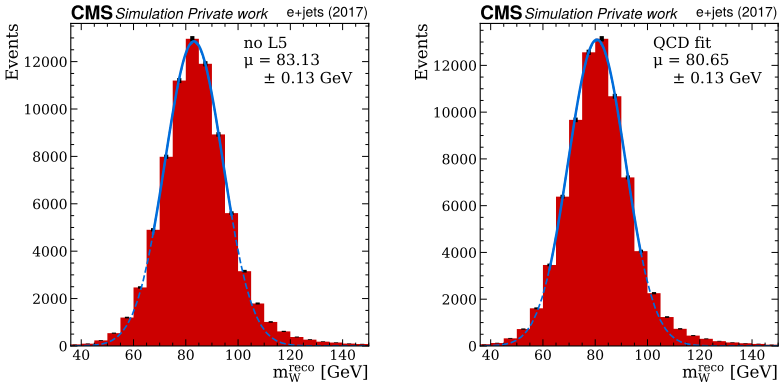


Fig. 7.2. Distribution of the hadronically decaying reconstructed W boson mass, m_W^{reco}, for the "correct" jet match of the $t\bar{t}$ lepton+jets events. Left: the result without the flavour corrections, right: after applying the flavour corrections from the $t\bar{t}$ dataset. The blue line shows the result of the Gaussian fit in the range where the fit was conducted and the dashed line shows the whole range of the fit result. Results are shown for the 2017 data in the e+jets channel.

7.2.1. Application of flavour-dependent JEC

Fig. 7.2 shows the m_W^{reco} distribution from correctly assigned $t\bar{t}$ lepton+jets permutations before and after applying the flavour corrections. The effect of the corrections brings the peak of $m_{\text{W}}^{\text{reco}}$ from $83.13 \pm 0.13\,\text{GeV}$ to $80.41 \pm 0.14\,\text{GeV}$ for the e+jet events, thus bringing it into agreement with $m_{\text{W}}^{\text{pdg}} = 80.4\,\text{GeV}$. The effect is explained by the light jets having a response $R > 1$ for the entire range of p_{T} and the jet $|\eta|$ and, so, the flavour correction lowers the p_{T} of the light jets forming the W boson candidate. The flavour corrections derived from the QCD dataset were also applied, but showed a slightly worse central value $m_{\text{W}}^{\text{pdg}} = 80.65 \pm 0.13\,\text{GeV}$, explained by differences in the event topologies. This shows that although significantly improving the p_{T} spectra of jets of different flavours, one set of flavour corrections may not be ideal for all analyses.

7.2.2. The reconstruction algorithm, WMassDeltaTopMass

Typically, for $t\bar{t}$ event reconstruction in the lepton+jets and all-jets channels, a kinematic fit is employed, which is a χ^2 fit of the kinematic properties of the four jets, lepton and neutrino momenta with several constraints [86], [101]. One of the constraints is typically $m_{t,\text{had}}^{\text{reco}} = m_{t,\text{lep}}^{\text{reco}}$, i.e. that the hadronically decaying top-quark mass should match the leptonically decaying top-quark mass. The equal-mass constraint corrects the b jet momenta to accommodate a smaller $m_{t,\text{had}}^{\text{reco}} - m_{t,\text{lep}}^{\text{reco}}$ and thus also biases the Δm_t measurement. In attempts using kinematic fit without the equal mass constraint, around 35 % of the jet combinations in $t\bar{t}$ events suffered non-convergence.

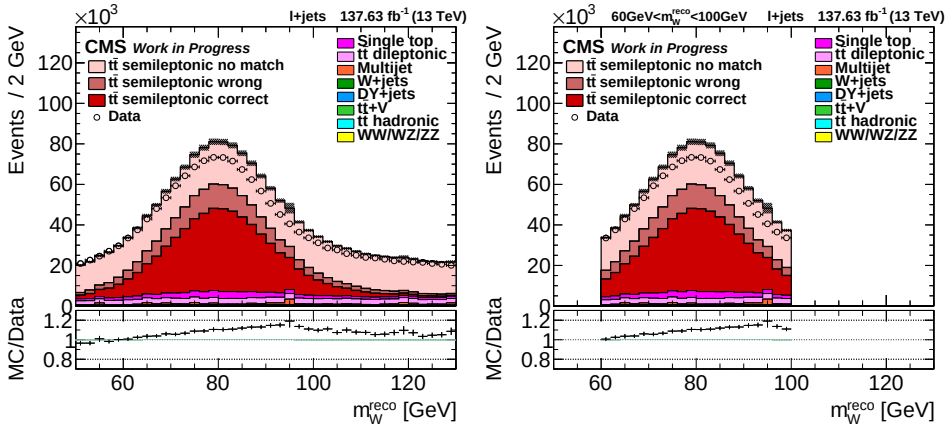


Fig. 7.3. The distribution of the reconstructed hadronic W mass, m_W^{reco}, before applying the $m_{\text{W}}^{\text{reco}}$ requirement (left) and after (right). Results are shown for the whole Run 2.

Instead of a kinematic fit, a cut-based reconstruction algorithm, called `WMassDeltaTopMass` (WM-DTM), was used. The aim of the algorithm is to obtain the best possible purity of the correct permutations and top mass peak resolution without introducing a bias in Δm_t. The analysis procedure is described in the following.

For each event, the two leading light jets are assigned to the W boson candidate. To reduce the detector effects on the light jets, the light-jet four-momenta are scaled, so that their invariant mass is equal to $m_{\text{W}}^{\text{pdg}}$ using

$$\begin{aligned} p_{q_1} &\to p_{q_1} \cdot \frac{\sqrt{(p_{q_1} + p_{q_2})^2}}{m_{\text{W}}^{\text{pdg}}} \\ p_{q_2} &\to p_{q_2} \cdot \frac{\sqrt{(p_{q_1} + p_{q_2})^2}}{m_{\text{W}}^{\text{pdg}}}. \end{aligned} \tag{7.1}$$

In the following, observables with superscript "fit" denote observables obtained after the scaling in Eq. (7.1) and choosing the neutrino solution as described in Section 7.2, while "reco" denotes observables before these corrections. Afterwards, the two permutations of the b jet assignment to the two top-quark candidates are tested, and the one yielding the smallest $|\Delta m_t^{\text{reco}}| = |m_{t,\text{had}}^{\text{reco}} - m_{t,\text{lep}}^{\text{reco}}|$ is stored.

To increase the ratio of correct permutations with regard to the total MC, a condition on m_W^{reco} was added such that $60\,\text{GeV} < m_{\text{W}}^{\text{reco}} < 100\,\text{GeV}$ ($m_{\text{W}}^{\text{reco}}$ requirement), mimicking the constraint of $m_{\text{W}}^{\text{reco}}$ in the kinematic fit. Note that with the light jet scaling in Eq. (7.1) the $m_{\text{W}}^{\text{fit}}$ distribution is just a delta function, while $m_{\text{W}}^{\text{reco}}$ stays intact and can be used for the $m_{\text{W}}^{\text{reco}}$ requirement. The distribution of $m_{\text{W}}^{\text{reco}}$ before and after the $m_{\text{W}}^{\text{reco}}$ requirement is shown in Fig. 7.3. It it possible to see how the $m_{\text{W}}^{\text{reco}}$ requirement removes the tails of the distribution consisting mostly of wrong b jet permutations and backgrounds. Note

that the peak of the distribution centres well around 80 GeV due to the application of the flavour corrections (compare with, e.g. [86], [102]).

7.2.3. The performance of the WMassDeltaTopMass algorithm

Table 7.2

Event yields for all the channels and all years with and without the $m_{\mathrm{W}}^{\mathrm{reco}}$ requirement

	No $m_{\mathrm{W}}^{\mathrm{reco}}$ requirement				With the $m_{\mathrm{W}}^{\mathrm{reco}}$ requirement			
	l^-+jets Run 2		l^++jets Run 2		l^-+jets Run 2		l^++jets Run 2	
Dataset	Events [k]	Ratio to total MC, %	Events [k]	Ratio to total MC, %	Events [k]	Ratio to total MC, %	Events [k]	Ratio to total MC, %
$t\bar{t}$ l+jets total	1239.7	84.0	1240.6	84.5	530.6	89.4	531.6	89.5
l+jets correct	289.1	19.6	289.5	19.7	255.9	43.1	256.3	43.1
l+jets wrong	115.7	7.8	115.8	7.9	83.0	14.0	83.0	14.0
l+jets no match	834.8	56.6	835.3	56.9	191.7	32.3	192.3	32.4
$t\bar{t}$ dilepton	113.3	7.7	113.4	7.7	26.3	4.4	26.4	4.4
$t\bar{t}$ all hadronic	1.2	0.1	1.3	0.1	0.4	0.1	0.4	0.1
$t\bar{t}$+V	3.1	0.2	2.7	0.2	0.9	0.2	0.8	0.1
single-top	70.3	4.8	61.0	4.2	25.4	4.3	23.3	3.9
DY+jets	4.2	0.3	4.1	0.3	1.0	0.2	1.0	0.2
W+jets	23.4	1.6	17.8	1.2	4.7	0.8	3.7	0.6
VV	1.0	0.1	0.8	0.1	0.3	0.0	0.3	0.0
Multijet	18.8	1.3	26.3	1.8	4.0	0.7	6.8	1.1
Simulation total	1474.9	100.0	1468.0	100.0	593.5	100.0	594.2	100.0
Data	1390.6	94.3	1378.7	93.9	542.8	91.5	540.3	90.9

In Table 7.2 the number of events obtained for each of the MC samples and data is compared with and without the $m_{\mathrm{W}}^{\mathrm{reco}}$ requirement. A charge asymmetry in the background MC samples can be seen, in particular in the single-top and W+jets samples, which contain more events in l^-+jets events (i.e. in events adding to the hadronic top quark distribution) than in l^++jets events (i.e. in events adding to the hadronic top antiquark distribution). The opposite trend is seen for multijet events. After the $m_{\mathrm{W}}^{\mathrm{reco}}$ requirement, the asymmetry becomes smaller for samples such as single-top and W +jets, while it increases for the multijet sample. The asymmetry is slightly larger in the data than in the simulation.

The $m_{\mathrm{W}}^{\mathrm{reco}}$ requirement is seen to improve the correct permutation purity to 43 %, which is almost as good as the kinematic fit with the P_{gof} cut, where it was found to be 47 % [86]. The MC count exceeds the data count by approximately 6 %, which can be covered by the MC uncertainties, as seen in previous analyses [86]. The MC-to-data disagreement increases to around 9 % after the $m_{\mathrm{W}}^{\mathrm{reco}}$ requirement.

The distributions of m_t^{reco} and m_t^{fit} after the $m_{\mathrm{W}}^{\mathrm{reco}}$ requirement are shown in Fig. 7.4. The condition in Eq. (7.1) can be seen to significantly reduce the width of the top-quark mass distribution. The $m_{\mathrm{W}}^{\mathrm{reco}}$ requirement is seen to remove the amount of backgrounds significantly, showing a sharp m_t^{fit} peak.

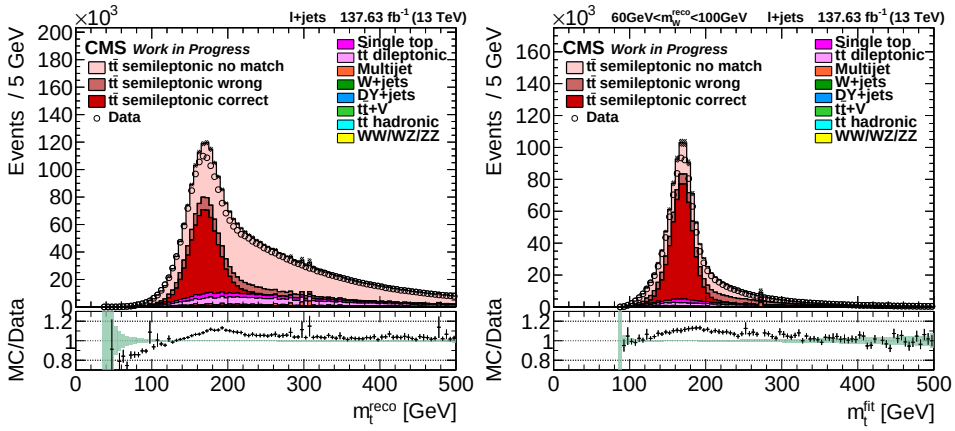


Fig. 7.4. The distribution of the reconstructed top-quark mass m_t^{reco} (left) and the fit top-quark mass m_t^{fit} after the $m_{\mathrm{W}}^{\mathrm{reco}}$ requirement. Distributions are shown for the whole Run 2 dataset.

The good efficiency and resolution of m_t^{fit} illustrate that WM-DTM serves as a more lightweight cut-based alternative to the kinematic fit while maintaining analogous characteristics. The selection of the $t\bar{t}$ events with WM-DTM leads to a smaller bias for the Δm_t measurement than with the kinematic fit, as there is no explicit shift of the b jet momenta matching the top-quark masses. Selecting only the first permutation does not have a significant impact on Δm_t. The usage of the second permutation would increase the number of events and thus might improve the sensitivity but also worsen the fraction of correctly matched events with respect to all the events. In future revisions of this analysis, the impact of adding the second permutation could be tested.

7.2.4. Reweighed datasets with non-zero top quark mass difference

As no MC samples were available with $|\Delta m_t| > 0$, the mass difference between top quarks and antiquarks was incorporated by adding weights to the simulated events shifting the position of the peak for the top quark and antiquark mass. In this way, events that have mass at the peak can be given a smaller weight than events at a new peak. The event reweighing was done using ratios of Breit-Wigner distributions $B(m_t^{\mathrm{gen}}, m_t, \Gamma)$ and is given as

$$w(m_t^{\mathrm{gen}}, m_{\bar{t}}^{\mathrm{gen}}, m_{t,\mathrm{new}}, m_{\bar{t},\mathrm{new}}, m_{t,\mathrm{old}}, \Gamma) = \left(\frac{B(m_t^{\mathrm{gen}}, m_{t,\mathrm{new}}, \Gamma)}{B(m_t^{\mathrm{gen}}, m_{t,\mathrm{old}}, \Gamma)}\right) \cdot \left(\frac{B(m_{\bar{t}}^{\mathrm{gen}}, M_{\bar{t},\mathrm{new}}, \Gamma)}{B(m_{\bar{t}}^{\mathrm{gen}}, M_{t,\mathrm{old}}, \Gamma)}\right), \tag{7.2}$$

where $m_{t,\mathrm{old}} = 172.5\,\mathrm{GeV}$ and $\Gamma = 1.31\,\mathrm{GeV}$ are the top quark mass and width parameters used for the original event generation, which are the same for top quark and top antiquark; $m_{t,\mathrm{new}}$ and $m_{\bar{t},\mathrm{new}}$ are the values of the top quark and top antiquark mass parameters to

which the reweighing should be performed; while m_t^{gen} and $m_{\bar{t}}^{\text{gen}}$ are the generated top-quark and antiquark masses in a specific event. In other words, the values of m_t^{gen} and $m_{\bar{t}}^{\text{gen}}$ are unique for each top quark and are distributed with a peak at the corresponding m_t and width Γ. The first term on the right-hand side in Eq. (7.2) reweighs the top quark, and the second term reweighs the top antiquark. The Breit-Wigner distribution is given as

$$B(m_t^{\text{gen}}, m_t, \Gamma) = \frac{k}{((m_t^{\text{gen}})^2 - m_t^2)^2 + m_t^2\Gamma^2}, \tag{7.3}$$

where

$$k = \frac{2\sqrt{2}m_t\Gamma\gamma}{\pi\sqrt{m_t^2+\gamma}} \approx \frac{m_t\Gamma\gamma}{\sqrt{m_t^2+\gamma}} \tag{7.4}$$

and

$$\gamma = \sqrt{m_t^2(m_t^2+\Gamma^2)}. \tag{7.5}$$

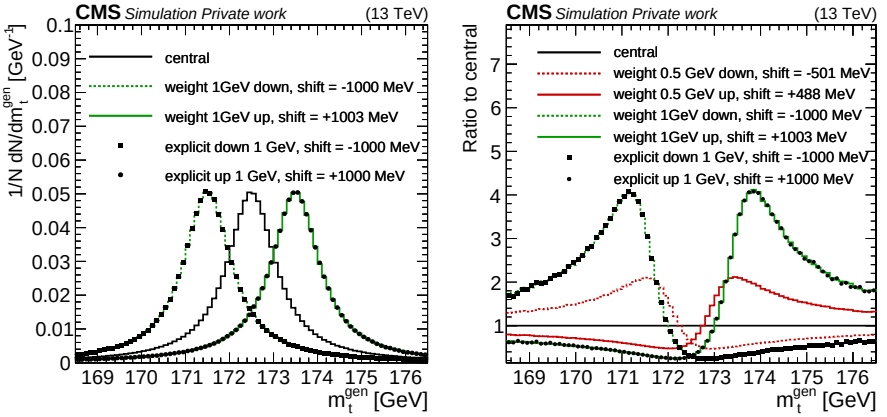


Fig. 7.5. Left: The comparison of the reweighed m_t^{gen} distribution, where m_t values were reweighed to 171.5 GeV and 173.5 GeV (in green) with the sample where the peak was explicitly generated at $m_t = 171.5$ GeV and at $m_t = 173.5$ GeV (the black markers). In addition, the central unreweighed sample generated with $m_t = 172.5$ GeV. Fit mean obtained by fitting distributions with a Voigt distributions is shown in the legend. Right: The ratios of all the samples shown on the left with respect to the central sample. In addition, here also the reweighed distribution to $m_t = 172$ GeV and $m_t = 173$ GeV is also shown (in red).

Reweighing is validated by comparing the m_t^{gen} and $m_{\bar{t}}^{\text{gen}}$ distributions, for samples where the initial $m_t = 172.5$ is reweighed to $m_t = 172.5 \pm 1$ GeV with the sample generated specifically with the same m_t values. The results are shown in Fig. 7.5 for the m_t^{gen} distribution. No visible difference is seen between the two samples. A similar agreement was obtained when comparing the $m_{\bar{t}}$ distributions.

For the profile likelihood fit, we use reweighing to $\Delta m_t = \pm 400$ MeV, thus each m_t and $m_{\bar{t}}$ is reweighed by ± 200 MeV. This covers the 68 % confidence interval of the Run 1 analysis, which measured $|\Delta m_t| + |\Delta(\Delta m_t)| = 360$ MeV [30]. Although reweighing

can increase the statistical uncertainties in the dataset, the loss of statistical accuracy is negligible for variations that are small compared to Γ.

7.3. Systematic uncertainties

The impact of all systematic uncertainties recommended in CMS is estimated. In addition, two sources of flavour-antiflavour uncertainty are estimated as derived in Section 6.3. The systematic uncertainties are divided into 3 categories: experimental uncertainties and signal normalisation (I), modelling uncertainties applied as event weights (II), and modelling uncertainties applied as additional uncorrelated samples (III). For the modelling uncertainties, the usage of weights as in type (II) is preferable to applying them as additional samples as in type (III) because of a perfect correlation among the variation if applying weights. Since the statistical uncertainties of the type (II) variations are almost fully correlated, the impact of the variation can be seen even if it occurs within the statistical uncertainties for bins with small number of events. In addition, when running the MC event generation, the weights are calculated on the fly, which does not increase the computation time as much as generating a new sample. Thus category (III) uncertainties are used only when an uncertainty using weights is not possible.

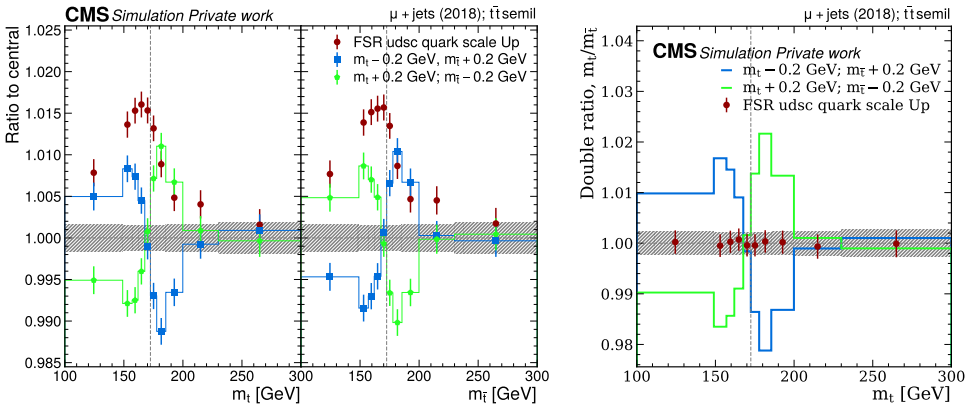


Fig. 7.6. Left: The ratio of the bin count in the FSR light quark scale up variation with respect to the central $t\bar{t}$ lepton+jets dataset. The left panel shows the results for the μ^-+jets sample, while the right panel for the μ^++jets sample. The ratio of the reweighed samples with $\Delta m_t \pm 400\,\text{MeV}$ is overlaid (up/plus variation in green and down/minus variation in blue). The uncertainty band shows the statistical uncertainty for the central sample, and the error bars show the statistical uncertainties for each of the other datasets. The dashed vertical line shows the generated top-quark mass value, $m_t^{\text{gen}} = 172.5\,\text{GeV}$. Results are shown for the 2018 μ+jets channel.
Right: The double ratio $(\text{var}(m_t)/\text{central}(m_t)\ )/(\text{var}(m_{\bar{t}})/\text{central}(m_{\bar{t}}))$, i.e. the ratio of the left panel with respect to the right panel of the left plot.

Most of the usual systematic uncertainties considered in measurements of m_t are expected to impact m_t and $m_{\bar{t}}$ in a correlated way and thus cancel in the Δm_t measurement. Nevertheless, these uncertainties are kept for completeness and since the profile-likelihood approach allows a simple implementation of these uncertainties.

Fig. 7.6 shows the ratio of the bin count in the $t\bar{t}$ lepton+jets up variation of the FSR light quark scale with the central dataset. Only the $t\bar{t}$ lepton+jets signal sample is shown, as the contribution of the other samples is small. The ratio is compared with the datasets reweighed to $\Delta m_t = \pm 400\,\text{MeV}$. $\Delta m_t = +400\,\text{MeV}$ corresponds to $m_t \to m_t + 200\,\text{MeV}$ and $m_{\bar{t}} \to m_{\bar{t}} - 200\,\text{MeV}$, while $\Delta m_t = -400\,\text{MeV}$ corresponds to the opposite effect. Note that the reweighing shifts m_t and $m_{\bar{t}}$ peaks in the opposite directions. A significant impact is seen on both m_t and $m_{\bar{t}}$, but it is very similar in both cases.

To visually compare the differences on how each variation impacts Δm_t, a double ratio of the left plot in Fig. 7.6 is constructed, where the values for m_t are divided by the values for $m_{\bar{t}}$. As the effect of the variation is correlated for m_t and $m_{\bar{t}}$, the double ratio shows only a small impact on Δm_t. That is, the double ratio for the variation overlaps almost completely with 1.00. Note that since the variation sample is obtained using weights on the central sample, both samples are correlated, and the statistical error bars are larger than the exact variation. A small non-cancellation is seen, potentially due to effects like charge asymmetry effects $t\bar{t}$ events, i.e. different η and p_T distributions for top quarks and antiquarks in the LHC [103]. In general, the η distribution is expected to be broader for top quarks than for top antiquarks. This could lead to slightly different jet dynamics after applying the FSR variation. In addition, the deviation from unity can also be a small effect of non-correlation between the central and variation sample, caused by the change in the events passing the m_W^reco requirement and due to some events failing the analysis pipeline.

An example of a variation that has an anticorrelated effect for m_t and $m_{\bar{t}}$ is the b jet flavour-antiflavour uncertainty, as shown in Fig. 7.7. Although the impact on the overall m_t and $m_{\bar{t}}$ distributions is more than $\mathcal{O}(10)$ smaller than for the FSR light quark scale variation, the double ratio shows a small effect of the variation on Δm_t. Moreover, unlike the double ratio in Fig. 7.6, the values in Fig. 7.7 are smoother and thus are more consistent with one of the Δm_t variations, in this case, with the Δm_t up variation. The minimum ratio of the b vs $\bar{b}$ variation is around 0.9995, compared to the effect of Δm_t, which has a minimum of 0.98, and thus is expected to have no more than a $400/40 = 10\,\text{MeV}$ impact on Δm_t. This can be compared to the 51 MeV impact on the Run 1 measurement [30].

7.4. Profile likelihood fit result

Fig. 7.8 shows the ratio of the m_t^{fit} and $m_{\bar{t}}^{fit}$ distributions, i.e. distributions in Fig. 7.4 Right, but split into positive and negative lepton charges. Note that since this is a single ratio, not a double ratio as in the previous section, the central prediction (gray line)

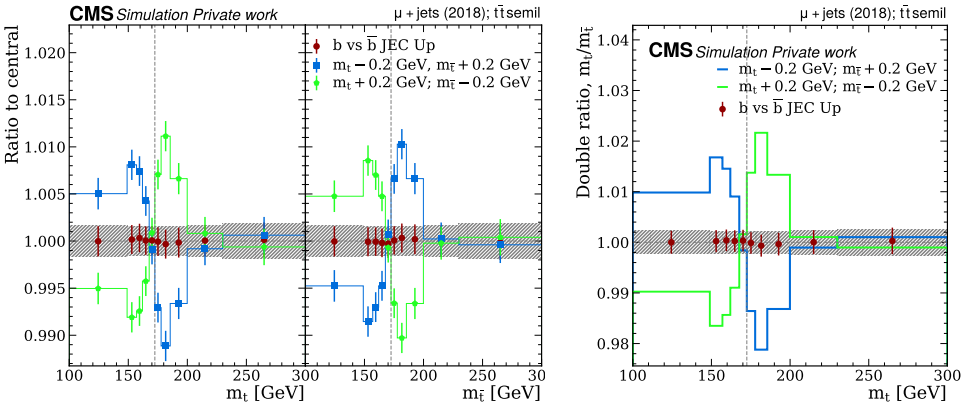


Fig. 7.7. Left: The ratio of the bin count in the b vs $\bar{b}$ up variation with respect to the central $t\bar{t}$ lepton+jets sample for μ^-+jets sample (left) and for μ^++jets sample (right). The legend explanation is the same as in Fig. 7.6.
Right: The double ratio (var(m_t)/central(m_t))/(var($m_{\bar{t}}$)/central($m_{\bar{t}}$)).

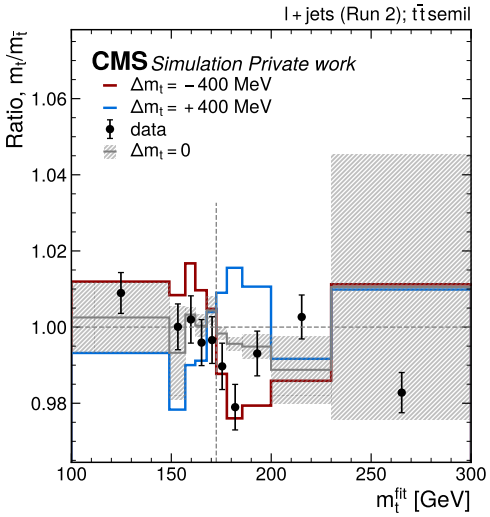


Fig. 7.8. The ratio of the m_t^{fit} and $m_{\bar{t}}^{fit}$ distributions. The gray line shows the ratio with $\Delta m_t = 0$, with the shaded boxes corresponding to the statistical uncertainty of the simulation. Blue and red lines show the $\Delta m_t \pm 400\,\text{MeV}$, respectively. The black markers show the result in data.

does not have to coincide with 1. The central prediction shows the expected small (up to 1 %) asymmetry in $t\bar{t}$ events with positive and negative charges with $\Delta m_t = 0$. The asymmetry is predominantly caused by background events, such as single top, W+jets, and QCD events, as shown in Table 7.2. The systematic uncertainties are approximately equal for all the data points because of approximately equal amount of data chosen for

each bin. Conversely, in the MC events, the uncertainty bars differ bin to bin, because of some background processes contributing more at the tails with larger weights. Comparing the data with the MC, it can be seen that some points, e.g. those around 175 GeV and 180 GeV are shifted towards $\Delta m_t = 400$ MeV and do not agree with the central simulation with the error bars, while some lower points are shifted towards $\Delta m_t = -400$ MeV.

A binned profile likelihood fit approach is used to perform the statistical inference. It is performed using the CMS COMBINE tool [104], which has the benefit of ensuring that the results can later be combined with other measurements by correctly calculating the correlations between the systematic uncertainties. The sensitivity to Δm_t was ensured by splitting the events according to the lepton charge and constructing the $(m_t^{\text{fit}},\ q^\ell)$ distribution, i.e. the m_t^{fit} distribution for events with positive and negative lepton charge, $q^\ell = \pm 1$. Events with $q^\ell = -1$ are sensitive to the top quark mass, m_t^{fit}, and events with $q^\ell = +1$ to the top antiquark mass, $m_{\bar{t}}^{\text{fit}}$.

Two parameters of interest are defined, m_t and Δm_t, and a nuisance parameter for each of the variations. The COMBINE fit obtains the best fit values by comparing the data with MC templates with different m_t and Δm_t. Three MC samples are used with m_t values of 171.5 GeV, 172.5 GeV and 173.5 GeV (all $\Delta m_t = 0$) to optimise m_t and samples with Δm_t values of -400 MeV, 0 and $+400$ MeV (all with $m_t = 172.5$ GeV) to optimise Δm_t. The m_t variations are obtained from separate MC datasets, while the Δm_t variations are obtained from event reweighing, as shown in Subsection 7.2.4.

The fit was performed in two steps. First, it was performed on the Asimov dataset, namely, taking the nominal MC result as the data, normalising the event yields to data. In other words, the values of the parameters of interest and nuisance parameters in the Asimov dataset are equal to the values used to generate the dataset. This gives the result consistent with $\Delta m_t = 0$, but enables the validation of the fit and the estimation of statistical and systematic uncertainties associated with the result without unblinding, i.e. revealing the actual result in the data. After that, a fit was performed on the collision data to obtain the final result. The fit was performed for Δm_t as the parameter of interest, letting m_t float freely. The whole Run 2 MC dataset was used for the fit.

The impact of each nuisance parameter is obtained by

$$\Delta(\hat{\Delta m_t}) = \hat{\Delta m_t}(\hat{\theta_k} + \Delta^{\pm}\theta_k) - \Delta m_t, \tag{7.6}$$

where $\hat{\Delta m_t}(\hat{\theta_k} + \Delta^{\pm}\theta_k)$ is the value of m_t obtained by the likelihood fit when the nuisance parameter θ_k is shifted to $\hat{\theta_k} + \Delta^{\pm}\theta_k$. The impact distribution for the main nuisance parameters for the measurement is shown in Fig. 7.9 together with the constraints and pulls on the nuisance parameters, relative to their prefit uncertainties, $(\hat{\theta}_i - \theta_i)/\sigma_i$, where σ_i is the standard deviation of the nuisance parameter before the fit. The expected results contain no pulls because the Asimov dataset is used for the fit. The leading uncertainties are the two b vs $\bar{b}$ JES sources, where the effect is greater for b vs $\bar{b}$ from the pion scaling

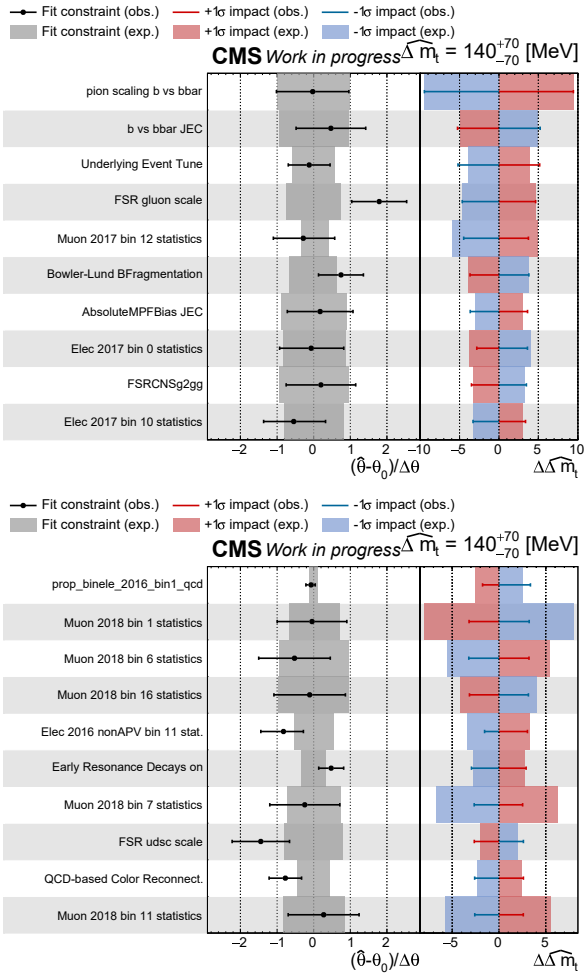


Fig. 7.9. The constraints and pulls of the main nuisance parameters (middle panel) and the nuisance parameter impacts on Δm_t (right panel) from the profile likelihood fit on the full Run 2 dataset. The shaded boxes show the results performed on the Asimov dataset (expected results), while the error bars show the results on the data (observed results). The red (blue) shaded boxes and error bars correspond to the positive (negative) impacts.

and is approximately 10 MeV. Further important uncertainties include variations obtained from independent datasets that possibly contain large statistical fluctuations.

Multiple variations are associated with the MC statistical uncertainty, indicated as `<sample> bin <bin_number> statistics`. This shows that the analysis, similarly to the Run 1 analysis, is dominated by statistical uncertainty. The greatest impact comes from the MC statistical uncertainty and the QCD background statistical uncertainty. Finally, there is a large impact from the FSR variation, where FSR for gluons is pulled towards larger values. Pulls for FSR variations have also been observed in the previous CMS m_t

measurement [86]. The nuisance parameter related to the variation of the POWHEG parameter *hdamp* was removed from the fit, because the tests where this nuisance parameter was included, it was pulled from the prefit value by 1.5σ creating a non-physically large pull. The reason for this pull is probably that the variation was obtained from a dataset with a limited number of events. The next revision of this analysis should include the estimation of the *hdamp* uncertainty using machine learning (ML) event reweighing, as in [105], [106]. The impact of this variation is not expected to be larger than a few MeV.

Table 7.3

Measured value, as well as the statistical, systematic and total uncertainty of the Δm_t measurement for each of the data taking period and channel

Period	e+jets [MeV]				μ+jets [MeV]				l+jets [MeV]			
	Δm_t	stat.	syst.	tot.	Δm_t	stat.	syst.	tot.	Δm_t	stat.	syst.	tot.
2016APV	350	$^{+262}_{-262}$	$^{+175}_{-176}$	$^{+315}_{-315}$	276	$^{+200}_{-201}$	$^{+186}_{-183}$	$^{+273}_{-271}$	281	$^{+159}_{-159}$	$^{+128}_{-127}$	$^{+204}_{-204}$
2016	-355	$^{+280}_{-280}$	$^{+183}_{-183}$	$^{+335}_{-335}$	303	$^{+219}_{-219}$	$^{+426}_{-424}$	$^{+479}_{-477}$	1	$^{+174}_{-174}$	$^{+191}_{-192}$	$^{+258}_{-259}$
2017	-111	$^{+184}_{-184}$	$^{+135}_{-135}$	$^{+228}_{-228}$	-68	$^{+147}_{-147}$	$^{+135}_{-144}$	$^{+199}_{-205}$	-62	$^{+115}_{-115}$	$^{+84}_{-87}$	$^{+142}_{-144}$
2018	444	$^{+153}_{-153}$	$^{+117}_{-117}$	$^{+193}_{-193}$	75	$^{+120}_{-120}$	$^{+150}_{-149}$	$^{+193}_{-192}$	198	$^{+94}_{-94}$	$^{+61}_{-61}$	$^{+112}_{-112}$
Full Run 2	125	$^{+100}_{-100}$	$^{+54}_{-54}$	$^{+113}_{-113}$	110	$^{+79}_{-79}$	$^{+53}_{-51}$	$^{+95}_{-94}$	139	$^{+62}_{-62}$	$^{+25}_{-25}$	$^{+67}_{-67}$

The final result of the measurement is $\Delta m_t = 139^{+67}_{-67}$ MeV. The breakdown of the uncertainty into statistical and systematic uncertainty is shown in Table 7.3, where a comparison between the individual data periods and the full Run 2 is shown. When compared to the Run 1 19.6 fb^{-1} result, the statistical uncertainty is reduced from 190 MeV to 62 MeV, that is, more than by a factor of 3. This is slightly better than predicted from the integrated luminosity increase but can be explained by the improved b-tagging efficiency and the usage of a profile-likelihood fit for the statistical inference instead of the Ideogram method. The systematic uncertainty is reduced from 90 MeV to 25 MeV, mostly due to the improved flavour vs antiflavour JES uncertainty. Both the statistical and systematic uncertainty are roughly proportional to the luminosity in each data-taking period. This comes from the fact that COMBINE is able to better constrain each variation with an increasing amount of data.

7.5. Summary

In this section, the measurement of Δm_t has been described. The measurement technique yielded a result with a total uncertainty of ± 67 MeV, which is an improvement by a factor of more than 3 with respect to the Run 1 measurement of the same parameter. This leads to the world-leading Δm_t measurement to date. The measurement excludes the effect of the ME-PS scale, which should be included in subsequent iterations using event reweighing based on ML.

8. CONCLUSION AND OUTLOOK

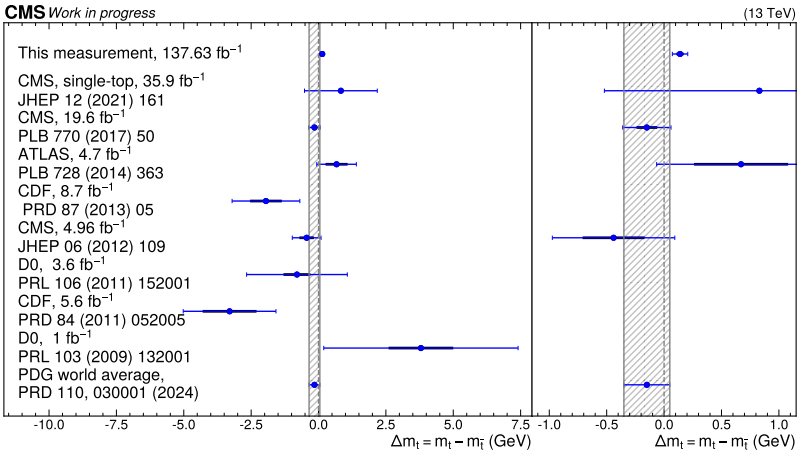


Fig. 8.1. The different measurements of the mass difference between the top quark and antiquark, including the result of this Thesis. The legend explanation is the same as for Fig. 1.2.

In this Thesis, the measurement of the mass difference between the top quark and antiquark was presented, yielding a result of $\Delta m_t = 139 \pm 67\,\text{MeV} = 139 \pm 62\ (\text{stat.}) \pm 25\ (\text{syst.})\,\text{MeV}$. This presents the most precise Δm_t measurement to date. The comparison of this measurement with other measurements of Δm_t is shown in Fig. 8.1. There is a slight disagreement with the CMS measurement at $\sqrt{s} = 8\,\text{TeV}$ and with the world average. Otherwise, the measurements exhibit general consistency. The result deviates from the SM prediction of $\Delta m_t = 0$ by 2.1σ, which is lower than 3σ, which is the accepted minimal threshold in particle physics for substantiating evidence of a discovery, and much less than 5σ required for a discovery. Thus, the measurement does not prove the violation of CPT symmetry.

The precision of this measurement relies on the improved prescription for the flavour-antiflavour JES uncertainty. Due to improved technique for the systematic uncertainties, the measurement is statistically limited just like the CMS Run 1 measurement. Adding the CMS Run 3 dataset would improve the statistical uncertainty by around a factor $\sqrt{2}$ due to the double luminosity (2023 data-taking period has $29\,\text{fb}^{-1}$, 2022 has recorded $35\,\text{fb}^{-1}$ luminosity, while in the 2024 data-taking period CMS has recorded $113.32\,\text{fb}^{-1}$ luminosity). The leading uncertainty is the b vs $\bar{b}$ uncertainty due to pion scaling, which estimates the mismodelling of GEANT 4 hadron interactions. The uncertainty could be reduced by repeating the study shown in Subsection 6.3.2 with a larger MC dataset, enabling the variation to be jet p_T dependent instead of only being jet $|\eta|$ dependent.

ACKNOWLEDGMENTS

My gratitude is overflowing to all the people who contributed directly or indirectly to this Thesis and my doctoral studies. Naming everyone individually would exceed the limits of this page.

First and foremost, I am thankful to my RTU supervisor Kārlis Dreimanis for his mentoring, guidance, and ensuring I had all the necessary resources throughout my research. I am equally grateful to my CERN supervisor Martijn Mulders for his support, especially for guiding me through the world of the CERN CMS group humming with cutting-edge physics ideas. Heartfelt thanks to my scientific advisor Markus Seidel. Whether it was scientific advice, practical help with some analysis tools, or proofreading my texts, he was always generous. Markus was unfailingly patient when discussing meticulous details of my work, including when I was often mistakenly sending him wrong links to results or messing up the plot legends.

I warmly thank my family, mom, dad, and brother, for being patient through the times of my absence, even when living in the same city. My father's unceasing support and encouragement in all the stages of my studies laid the foundation for where I stand today.

My gratitude goes to all my colleagues, but particularly to Antra, Normunds, and Robert. Each at a different spacetime point was interacting with me strongly, where the interaction was mediated by common complaints and frustrations, meaningful lunch breaks, or just silence.

I extend my deep appreciation to everyone who helped me to adjust and survive in different countries and environments. In a gray February day of 2022, with me hiding in my room due to COVID positivity and when my Doctoral studies were hanging on a tiny thread, our 7-room shared house was accompanied by Ana María. Out of nowhere it felt she came, and at the right moment had good talks and encouragement to look for help, advice, and prayer. Thanks to God for being faithful when we are not.

Thanks to Sandijs for being available for spontaneous calls even after getting married (I am so happy). Thanks to Paula for bringing joy into the writing process and teaching me so much. Thanks to St. Matthew's church home group for renewing my goals and purpose. To Agape student organization for doing fun things with me at the uni, and Crossroads and Communitas international churches for welcoming me and being ready to listen and pray for me.

I am grateful to my outdoor friends, especially Andrew for making me renew my promise of love to mountains and getting me out of my house almost more than my body could sustain. Thanks to Juan and Tom for the continuation of this wholesome work.

Thanks to my electric guitar, Nelly (Epiphone ES-335), for always sounding so great when fingers and ears collaborate. And though unexpected, thanks also to COVID for teaching me valuable life lessons.

References

[1] C. E. Gerber, "LHC Highlights and Prospects", doi:10.23730/CYRSP-2021-002.161, arXiv:1909.10919. 38 pages, 31 figures, Lecture notes for the 10th CERN Latin-American School of High-Energy Physics (CLASHEP2019), 13 to 26 March 2019 in Villa General Belgrano, Cordoba Province, Argentina.

[2] CMS Collaboration, "Stairway to discovery: A report on the CMS programme of cross section measurements from millibarns to femtobarns", May 2024, doi:10.48550/arXiv.2405.18661.

[3] ATLAS Collaboration, "Summary plots from the ATLAS Standard Model physics group", 2024, https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CombinedSummaryPlots/SM/index.html.

[4] L. Canetti, M. Drewes, and M. Shaposhnikov, "Matter and Antimatter in the Universe", *New J. Phys.*, **14** (September 2012) 095012, doi:10.1088/1367-2630/14/9/095012, arXiv:1204.4186.

[5] SNO Collaboration et al., "Direct Evidence for Neutrino Flavor Transformation from Neutral-Current Interactions in the Sudbury Neutrino Observatory", *Phys. Rev. Lett.*, **89** (June 2002) 011301, doi:10.1103/PhysRevLett.89.011301.

[6] The Super-Kamiokande Collaboration, "Evidence for oscillation of atmospheric neutrinos", *Phys. Rev. Lett.*, **81** (August 1998) 1562–1567, doi:10.1103/PhysRevLett.81.1562, arXiv:hep-ex/9807003.

[7] G. Bertone, "History of dark matter", *Rev. Mod. Phys.*, **90** (2018),, no. 4, doi:10.1103/RevModPhys.90.045002.

[8] J. G. de Swart, G. Bertone, and J. van Dongen, "How dark matter came to matter", *Nat Astron*, **1** (March 2017) 1–9, doi:10.1038/s41550-017-0059.

[9] C. Rovelli, "Notes for a brief history of quantum gravity", January 2001, doi:10.48550/arXiv.gr-qc/0006061.

[10] A. Zee, "Quantum Field Theory in a Nutshell". Princeton University Press, second edition edition, 2010. ISBN 978-0-691-14034-6, https://www.amazon.com/Quantum-Field-Theory-Nutshell-nutshell/dp/0691140340.

[11] ATLAS Collaboration, "ATLAS Run 1 searches for direct pair production of third-generation squarks at the Large Hadron Collider", *Eur. Phys. J. C*, **75** (October 2015) 510, doi:10.1140/epjc/s10052-015-3726-9, arXiv:1506.08616.

[12] CMS Collaboration, "Search for single production of scalar leptoquarks in proton-proton collisions at $\sqrt{s} = 8\ TeV$", *Phys. Rev. D*, **93** (2016)„ no. 3, 032005, doi:10.1103/PhysRevD.93.032005, arXiv:1509.03750. [Erratum: Phys.Rev.D 95, 039906 (2017)].

[13] C. P. Salemi et al., "Search for Low-Mass Axion Dark Matter with ABRACADABRA-10 cm", *Phys. Rev. Lett.*, **127** (2021)„ no. 8, 081801, doi:10.1103/PhysRevLett.127.081801, arXiv:2102.06722.

[14] A. Gómez-Bañón, K. Bartnick, K. Springmann, and J. A. Pons, "Constraining light QCD axions with isolated neutron star cooling", arXiv:2408.07740.

[15] J. S. Bell, "Time reversal in field theory", *Proc. R. Soc. Lond.*, **A** (1955)„ no. 231, 479–495.

[16] R. F. Streater and A. S. Wightman, "PCT, Spin and Statistics, and All That". Princeton University Pressrsity, 2001. ISBN 978-0-691-07062-9, https://press.princeton.edu/books/paperback/9780691070629/pct-spin-and-statistics-and-all-that.

[17] V. N. Gribov and J. Nyiri, "Quantum electrodynamics: Gribov lectures on theoretical physics", volume 13. Cambridge University Press, 7 2005. ISBN 978-0-521-67569-7.

[18] D. Colladay and V. A. Kostelecky, "CPT violation and the standard model", *Phys. Rev. D*, **55** (1997), 6760–6774, doi:10.1103/PhysRevD.55.6760.

[19] S. Mukohyama and S. C. Park, "Ghost condensation and *CPT* violation in neutrino sector", *Physics Letters B*, **696** (February 2011) 505–508, doi:10.1016/j.physletb.2011.01.014.

[20] G. Barenboim and J. Lykken, "A model of CPT violation for neutrinos", *Physics Letters B*, **554** (February 2003) 73–80, doi:10.1016/S0370-2693(02)03262-8, arXiv:hep-ph/0210411.

[21] Particle Data Group Collaboration, "Review of particle physics", *Phys. Rev. D*, **110** (2024)„ no. 3, 030001, doi:10.1103/PhysRevD.110.030001.

[22] CPLEAR Collaboration, "K0 anti-K0 mass and decay width differences: CPLEAR evaluation", *Phys. Lett. B*, **471** (1999), 332–338, doi:10.1016/S0370-2693(99)01333-7.

[23] A. Belyaev et al., "Probing CPT invariance with top quarks at the LHC", November 2024, doi:10.48550/arXiv.2405.12162.

[24] CMS Collaboration, "Searches for violation of Lorentz invariance in top quark pair production using dilepton events in 13 TeV proton-proton collisions", *Phys. Lett. B*, **857** (2024), 138979, doi:10.1016/j.physletb.2024.138979, arXiv:2405.14757.

[25] M. S. Fee et al., "Measurement of the positronium 1 S-31-2 S-31 interval by continuous-wave two-photon excitation", *Phys. Rev. A*, **48** (1993), 192–219, doi:10.1103/PhysRevA.48.192.

[26] Belle Collaboration, "Measurement of the mass of the tau-lepton and an upper limit on the mass difference between tau+ and tau-", *Phys. Rev. Lett.*, **99** (2007), 011801, doi:10.1103/PhysRevLett.99.011801, arXiv:hep-ex/0608046.

[27] D. S. Ayers et al., "Measurements of the lifetimes of positive and negative pions", *Phys. Rev. D*, **3** (1971), 1051–1063, doi:10.1103/PhysRevD.3.1051.

[28] DELPHI Collaboration, "Masses, Lifetimes and Production Rates of Xi- and anti-Xi+ at LEP 1", *Phys. Lett. B*, **639** (2006), 179–191, doi:10.1016/j.physletb.2006.06.029, arXiv:hep-ex/0606030.

[29] E756 Collaboration, "Measurement of the Properties of the $\bar{\Omega}^+$ and Ω^- hyperons", *Phys. Rev. D*, **58** (1998), 072002, doi:10.1103/PhysRevD.58.072002.

[30] CMS Collaboration, "Measurement of the mass difference between top quark and antiquark in pp collisions at $\sqrt{s} = 8$ TeV", *Phys. Lett. B*, **770** (2017), 50–71, doi:10.1016/j.physletb.2017.04.028, arXiv:1610.09551.

[31] CMS Collaboration, "Measurement of the Mass Difference between Top and Antitop Quarks", *JHEP*, **06** (2012), 109, doi:10.1007/JHEP06(2012)109, arXiv:1204.2807.

[32] CMS Collaboration, "Measurement of the top quark mass using events with a single reconstructed top quark in pp collisions at $\sqrt{s} = 13$ TeV", *JHEP*, **12** (2021), 161, doi:10.1007/JHEP12(2021)161, arXiv:2108.10407.

[33] ATLAS Collaboration, "Measurement of the mass difference between top and anti-top quarks in pp collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector", *Phys. Lett. B*, **728** (2014), 363–379, doi:10.1016/j.physletb.2013.12.010, arXiv:1310.6527.

[34] CDF Collaboration, "Measurement of the Mass Difference between Top and Anti-top Quarks", *Phys. Rev. D*, **87** (2013)„ no. 5, 052013, doi:10.1103/PhysRevD.87.052013, arXiv:1210.6131.

[35] D0 Collaboration, "Direct measurement of the mass difference between top and antitop quarks", *Phys. Rev. D*, **84** (2011), 052005, doi:10.1103/PhysRevD.84.052005, arXiv:1106.2063.

[36] CDF Collaboration, "Measurement of the Mass Difference between t and $\bar{t}$ Quarks", *Phys. Rev. Lett.*, **106** (2011), 152001, doi:10.1103/PhysRevLett.106.152001, arXiv:1103.2782.

[37] D0 Collaboration, "Direct measurement of the mass difference between top and antitop quarks", *Phys. Rev. Lett.*, **103** (2009), 132001, doi:10.1103/PhysRevLett.103.132001, arXiv:0906.1172.

[38] ATLAS and CMS Collaboration, "Improved Common $t\bar{t}$ Monte-Carlo Settings for ATLAS and CMS", technical report, CERN, Geneva, 2023, cds:2861366.

[39] ATLAS Collaboration, "Improved Common $t\bar{t}$ Monte-Carlo Settings for ATLAS and CMS", technical report, CERN, Geneva, 2023, cds:2862524.

[40] A. Potrebko, "Flavor dependent (L5) MC truth jet energy corrections and flavor uncertainties in Run 2", Technical Report CMS AN-2023/074, CERN, 2024.

[41] A. Potrebko, M. Seidel, M. Mulders, and K. Dreimanis, "Measurement of the mass difference between the top quark and antiquark in $t\bar{t}$ events with lepton+jets final states in 13 TeV proton-proton collisions", Technical Report CMS AN-2023/074, 2025.

[42] J. J. Thomson, "XL. Cathode Rays", *Lond. Edinb. Dublin Philos. Mag. J. Sci.*, **44** (October 1897) 293–316, doi:10.1080/14786449708621070.

[43] 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*, **716** (2012), 1–29, doi:10.1016/j.physletb.2012.08.020, arXiv:1207.7214.

[44] CMS Collaboration, "Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC", *Phys. Lett. B*, **716** (2012), 30–61, doi:10.1016/j.physletb.2012.08.021, arXiv:1207.7235.

[45] LHCb Collaboration, "Observation of the resonant character of the $Z(4430)^-$ state", *Phys. Rev. Lett.*, **112** (2014),, no. 22, 222002, doi:10.1103/PhysRevLett.112.222002, arXiv:1404.1903.

[46] LHCb Collaboration, "Observation of $J/\psi p$ Resonances Consistent with Pentaquark States in $\Lambda_b^0 \to J/\psi K^- p$ Decays", *Phys. Rev. Lett.*, **115** (2015), 072001, doi:10.1103/PhysRevLett.115.072001, arXiv:1507.03414.

[47] E. Noether, "Invariante variationsprobleme", *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse*, **1918** (1918), 235–257.

[48] C. S. Wu et al., "Experimental Test of Parity Conservation in Beta Decay", *Phys. Rev.*, **105** (February 1957) 1413–1415, doi:10.1103/PhysRev.105.1413.

[49] J. H. Christenson, J. W. Cronin, V. L. Fitch, and R. Turlay, "Evidence for the 2π Decay of the K_2^0 Meson", *Phys. Rev. Lett.*, **13** (1964), 138–140, doi:10.1103/PhysRevLett.13.138.

[50] D. Colladay and A. Kostelecky, "Lorentz-Violating Extension of the Standard Model", *Phys. Rev. D*, **58** (October 1998) 116002, doi:10.1103/PhysRevD.58.116002, arXiv:hep-ph/9809521.

[51] M. Seidel, "Tuning of generators", in *QCD@LHC2022*. October 2022.

[52] T. Sjöstrand, S. Mrenna, and P. Skands, "A Brief Introduction to PYTHIA 8.1", *Computer Physics Communications*, **178** (June 2008) 852–867, doi:10.1016/j.cpc.2008.01.036, arXiv:0710.3820.

[53] T. Sjöstrand et al., "An Introduction to PYTHIA 8.2", *Computer Physics Communications*, **191** (June 2015) 159–177, doi:10.1016/j.cpc.2015.01.024, arXiv:1410.3012.

[54] J. Bellm et al., "Herwig 7.0/Herwig++ 3.0 release note", *Eur. Phys. J. C*, **76** (April 2016) 196, doi:10.1140/epjc/s10052-016-4018-8.

[55] J. Bellm et al., "Herwig 7.1 Release Note", May 2017, doi:10.48550/arXiv.1705.06919.

[56] E. Bothmann et al., "Event Generation with Sherpa 2.2", *SciPost Phys.*, **7** (September 2019) 034, doi:10.21468/SciPostPhys.7.3.034, arXiv:1905.09127.

[57] S. Frixione, G. Ridolfi, and P. Nason, "A positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction", *J. High Energy Phys.*, **2007** (September 2007) 126, doi:10.1088/1126-6708/2007/09/126.

[58] S. Alioli, P. Nason, C. Oleari, and E. Re, "A general framework for implementing NLO calculations in shower Monte Carlo programs: The POWHEG BOX", *J. High Energ. Phys.*, **2010** (June 2010) 43, doi:10.1007/JHEP06(2010)043, arXiv:1002.2581.

[59] 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",

J. High Energ. Phys., **2014** (July 2014) 79, doi:10.1007/JHEP07(2014)079, arXiv:1405.0301.

[60] CMS Collaboration, "CMS Open data docmentation". http://opendata.cern.ch/docs/cms-guide-event-production, Dec 2023.

[61] R. Brun and F. Rademakers, "ROOT – An object oriented data analysis framework", *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, **389** (April 1997) 81–86, doi:10.1016/S0168-9002(97)00048-X.

[62] S. Agostinelli et al., "Geant4—a simulation toolkit", *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, **506** (July 2003) 250–303, doi:10.1016/S0168-9002(03)01368-8.

[63] J. Allison et al., "Recent developments in Geant4", *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, **835** (November 2016) 186–225, doi:10.1016/j.nima.2016.06.125.

[64] J. Allison et al., "Geant4 developments and applications", *IEEE Trans. Nucl. Sci.*, **53** (February 2006) 270–278, doi:10.1109/TNS.2006.869826.

[65] "LHC TOP WG: WG on Top Physics at the LHC | LPCC". https://lpcc.web.cern.ch/content/lhc-top-wg.

[66] ATLAS, CMS Collaboration, "Combination of Measurements of the Top Quark Mass from Data Collected by the ATLAS and CMS Experiments at s=7 and 8 TeV", *Phys. Rev. Lett.*, **132** (2024),, no. 26, 261902, doi:10.1103/PhysRevLett.132.261902, arXiv:2402.08713.

[67] A. Buckley et al., "Rivet user manual", *Computer Physics Communications*, **184** (December 2013) 2803–2819, doi:10.1016/j.cpc.2013.05.021, arXiv:1003.0694.

[68] CMS Collaboration, "Measurements of differential cross sections of top quark pair production as a function of kinematic event variables in proton-proton collisions at $\sqrt{s} = 13$ TeV", *JHEP*, **06** (2018), 002, doi:10.1007/JHEP06(2018)002, arXiv:1803.03991.

[69] G. S. Chahal and F. Krauss, "Cluster Hadronisation in Sherpa", *SciPost Phys.*, **13** (August 2022) 019, doi:10.21468/SciPostPhys.13.2.019.

[70] ATLAS Collaboration, "Dependence of the Jet Energy Scale on the Particle Content of Hadronic Jets in the ATLAS Detector Simulation", technical report, CERN, Geneva, 2022, cds:2808016.

[71] Laurent Guiraud, "Plan of sps to lhc transfer tunnels. plan des accélérateurs lhc et sps ainsi que le plan des deux nouveaux tunnels de transfert ti2 et ti8", technical report, CERN, Aug 2001, cds:42387.

[72] CERN, "LHC Guide", Mar 2017, cds:2255762.

[73] O. S. Brüning et al., "LHC Design Report", technical report, CERN, Geneva, 2004, doi:10.5170/CERN-2004-003-V-2.

[74] L. Evans and P. Bryant, "LHC Machine", *J. Instrum.*, **3** (2008)„ no. 08, S08001–S08001, doi:10.1088/1748-0221/3/08/s08001.

[75] CERN, "LEP design report", technical report, CERN, Geneva, 1984, cds:102083. Copies shelved as reports in LEP, PS and SPS libraries.

[76] CMS Collaboration, "Particle-flow reconstruction and global event description with the CMS detector", *J. Instrum.*, **12** (October 2017) P10003–P10003, doi:10.1088/1748-0221/12/10/P10003, arXiv:1706.04965.

[77] The CMS Collaboration et al., "The CMS experiment at the CERN LHC", *J. Inst.*, **3** (August 2008) S08004, doi:10.1088/1748-0221/3/08/S08004.

[78] I. Neutelings, "CMS coordinate system", July 2024, https://tikz.net/axis3d_cms/.

[79] CMS Collaboration, "Description and performance of track and primary-vertex reconstruction with the CMS tracker", *J. Inst.*, **9** (October 2014) P10009–P10009, doi:10.1088/1748-0221/9/10/P10009, arXiv:1405.6569.

[80] Y. L. Dokshitzer, G. D. Leder, S. Moretti, and B. R. Webber, "Better jet clustering algorithms", *JHEP*, **08** (1997), 001, doi:10.1088/1126-6708/1997/08/001.

[81] M. Cacciari, Gavin P. Salam, and Gregory Soyez, "The anti-kt jet clustering algorithm", *J. High Energy Phys.*, **2008** (April 2008) 063, doi:10.1088/1126-6708/2008/04/063.

[82] CMS Collaboration, "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV", *J. Instrum.*, **13** (2018)„ no. CMS-BTV-16-002, CERN-EP-2017-326, P05011, doi:10.1088/1748-0221/13/05/P05011, arXiv:1712.07158.

[83] V. Khachatryan et al., "Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV", *J. Inst.*, **12** (February 2017) P02014, doi:10.1088/1748-0221/12/02/P02014.

[84] J. Gallicchio and M. D. Schwartz, "Quark and gluon jet substructure", *J. High Energ. Phys.*, **2013** (April 2013) 90, doi:10.1007/JHEP04(2013)090.

[85] A. Tumasyan et al., "Study of quark and gluon jet substructure in Z+jet and dijet events from pp collisions", *J. High Energ. Phys.*, **2022** (January 2022) 188, doi:10.1007/JHEP01(2022)188.

[86] CMS Collaboration, "Measurement of the top quark mass using a profile likelihood approach with the lepton + jets final states in proton–proton collisions at $\sqrt{s} = 13\,\text{TeV}$", *Eur. Phys. J. C*, **83** (2023)„ no. 10, 963, doi:10.1140/epjc/s10052-023-12050-4, arXiv:2302.01967.

[87] CMS Collaboration, "Measurement of multidifferential cross sections for dijet production in proton-proton collisions at $\sqrt{s} = 13$ TeV", arXiv:2312.16669.

[88] CMS Collaboration, "Measurement and QCD analysis of double-differential inclusive jet cross sections in proton-proton collisions at $\sqrt{s} = 13$ TeV", *JHEP*, **02** (2022), 142, doi:10.1007/JHEP02(2022)142, arXiv:2111.10431. [Addendum: JHEP 12, 035 (2022)].

[89] CMS Collaboration, "Measurement of Energy Correlators inside Jets and Determination of the Strong Coupling αS(mZ)", *Phys. Rev. Lett.*, **133** (2024)„ no. 7, 071903, doi:10.1103/PhysRevLett.133.071903, arXiv:2402.13864.

[90] CMS Collaboration, "Measurement of the Ratio of the Inclusive 3-Jet Cross Section to the Inclusive 2-Jet Cross Section in pp Collisions at $\sqrt{s} = 7$ TeV and First Determination of the Strong Coupling Constant in the TeV Range", *Eur. Phys. J. C*, **73** (2013)„ no. 10, 2604, doi:10.1140/epjc/s10052-013-2604-6, arXiv:1304.7498.

[91] CMS Collaboration, "Measurement of the cross section for $\mathrm{t\bar{t}}$ production with additional jets and b jets in pp collisions at $\sqrt{s} = 13$ TeV", *JHEP*, **07** (2020), 125, doi:10.1007/JHEP07(2020)125, arXiv:2003.06467.

[92] CMS Collaboration, "Calculation of Residual Energy Correction for b Jets Using Z+b Events in 8 TeV pp Collisions", technical report, CERN, Geneva, 2014, cds:1951028.

[93] Viola Sordini, Hugues Lattaud, and Viola Sordini, "Energy scale studies for b-jets with γ + jet events in Run2016 data.", Technical Report CMS AN-2017/174, CERN.

[94] K. Kallonen, "Quantitative Comparison of Deep Neural Networks for Quark/Gluon Jet Discrimination". Master's thesis, Helsinki U., 2019.

[95] N. T. Toikka, "Relative response measurement between light quark and gluon jets at the CMS experiment". Master's thesis, Helsinki U., 2023.

[96] Robin Cameron Aggleton, Anastasia Karavdina, and Andreas Hinzmann, "Flavour-dependent jet energy corrections for 2016 Monte Carlo", Technical Report AN-2018/109, CERN, May 2018.

[97] CMS Collaboration, "Development and validation of HERWIG 7 tunes from CMS underlying-event measurements", April 2021, doi:10.1140/epjc/s10052-021-08949-5.

[98] ATLAS Collaboration, "Jet Energy Scale uncertainties (R=0.4 PFlow) with updated flavour treatment", August 2022, https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/JETM-2022-005/.

[99] CMS HCAL/ECAL Collaborations et al., "The CMS barrel calorimeter response to particle beams from 2 to 350 GeV/c", *Eur. Phys. J. C*, **60** (April 2009) 359–373, doi:10.1140/epjc/s10052-009-0959-5.

[100] S. Banerjee and V. Ivanchenko, "Validation of Physics Models of Geant4 Versions 10.4.p03, 10.6.p02 and 10.7.p01 using Data from the CMS Experiment", *EPJ Web Conf.*, **251** (2021), 03010, doi:10.1051/epjconf/202125103010.

[101] A. Hayrapetyan et al., "Review of top quark mass measurements in CMS",.

[102] H. Siikonen, "Measurement of the Top Quark Mass with the CMS Experiment". PhD thesis, Helsinki U., 2022.

[103] CMS Collaboration, "Measurement of the $t\bar{t}$ charge asymmetry in events with highly Lorentz-boosted top quarks in pp collisions at $\sqrt{s} = 13\,\text{TeV}$", *Phys. Lett. B*, **846** (2023), 137703, doi:10.1016/j.physletb.2023.137703, arXiv:2208.02751.

[104] CMS Collaboration, "The CMS statistical analysis and combination tool: COMBINE", April 2024, doi:10.48550/arXiv.2404.06614.

[105] E. Bothmann et al., "Exploring phase space with Neural Importance Sampling", *SciPost Physics*, **8** (April 2020) 069, doi:10.21468/SciPostPhys.8.4.069.

[106] N. Deutschmann and N. Götz, "Accelerating HEP simulations with Neural Importance Sampling", *J. High Energ. Phys.*, **2024** (March 2024) 83, doi:10.1007/JHEP03(2024)083.

Andris Potrebko was born in 1996 in Riga. He obtained a Bachelor's degree in Physics (2018) from the University of Latvia and a Master's degree in Physics (2020) from Lund University, Sweden.
His interest in particle physics was sparked when he participated in the CERN summer student programme in 2018. In 2020, he was a scientific intern at Mainz University in Germany, where he worked on lattice quantum chromodynamics. Presently, he is a scientific assistant at the Institute of Particle Physics and Accelerator Technologies of RTU. From December 2025, he will take a researcher position at Université libre de Bruxelles (ULB) in Belgium.
His research interests include top quark, jet and standard model physics as well as computational physics. He is also passionate about machine learning. Since 2020, he has been a member of the CMS collaboration at CERN. During his PhD studies, he was an active guide of the CMS experiment and took detector control shifts during the experiment operations. In addition, he was teaching a physics course for high school students at RTU and at the University for Future Engineers (UNI). Andris is passionate about mountain climbing and hiking in his free time.