# New Physics Tools for Discovery, a New Era of Timing Detector, and Lepton Flavor Universality Test at CMS


Thesis by
Olmo Cerri


In Partial Fulfillment of the Requirements for the
Degree of
Doctor of Philosophy

Caltech


CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2023
Defended August 3rd, 2022

Olmo Cerri
ORCID: 0000-0002-2191-0666

*Audentes Fortuna iuvat.*

# ACKNOWLEDGEMENTS

I would like to thank the California Institute of Technology for the inspiration, opportunities, and support provided during all these years. All the means I have been provided with have always been up to and beyond the challenges I faced. Similarly to my second home in Scuola Normale, in my third home here at Caltech I felt part of a collaborative and supportive environment filled with stimulating and brilliant colleagues, which allowed me to grow not only as a physicist and a scientist, but also as a man. I would also like to mention the FNAL and CERN laboratories, which hosted me during the completion of part of this work and whose facilities and support have been crucial to achieve the results presented.

I am deeply grateful to my supervisor Prof. Maria Spiropulu who championed for my presence here at Caltech, welcomed me in her group, and supported me in all the endeavors we embarked on. Maria's mentorship extended well beyond the physics subject boundaries. The freedom, the trust, and the responsibilities she gave me allowed me to grow both on the professional and personal side. It has been a honor and a true pleasure to be in your group and spend here the years of my PhD. A special thanks and acknowledgement goes to Maurizio Pierini, who has been nothing less than a second supervisor to me. Maurizio has been a crucial and invaluable guide and an inspiration from the first work we did together on FCC-ee in 2014, through the multiple machine learning ventures and the flavor physics challenges we are still facing.

I would like to express my deepest gratitude to all the people with whom I have been collaborating and who advised me with fruitful comments and suggestions. In this regard, the first acknowledgment goes to Tony, who in the last year shared with me all the highs and all the lows of the $\mathcal{R}(D^*)$ analysis. The analysis would not have reached its current status without your contribution. I really enjoyed working with you. I would like also to mention Si, Cristian, Justas, Riccardo, Thong, Andrea M., Elisabetta, Andrea R., Michele, and all the others. The long hours spent in discussing ideas and problems are the foundation of this thesis. A sincere thank you to friends and colleagues from Caltech, CMS, and INQNET. I hope that we will have the opportunity to join our efforts once again in the future. I conclude this part of my acknowledgments within the academic community by mentioning the people that allowed all this to begin. I am obliged to my physics high school teacher, Prof. Pascucci for instilling that very first spark of curiosity for a rational description of our Nature that today still burns. I am privileged to have encountered Prof. Gigi Rolandi, who provided an ever-present trustworthy guidance throughout my scholar, research, and life path. I am grateful to prof. Luciano Ristori not only for sharing his wisdom with me, but also for pointing me towards this very one choice of institution for my Ph.D. It is for me an incredible and true honor to be part of the legacy of all these great scientists.

It is with immense pleasure that I thank all the people in my personal life for the extraordinary support, patience, and motivation. My path would not even be comparable without the people I met along the way. A first big thank you goes to a dear friend that over time became much more. From the happy days before the COVID, to the long quarantine months, to the thesis defense, you played the crucial role of being always by my side. Thank you Katrina for your support, your joy, and our happiness. A true thanks goes to my friends, all of them, from the ones that have been sharing with me only a few months, to the ones that have been there since I have memories, to the ones that are not part of my life as much as they used to be. I would like to add a cheer to Ale, Steve, and Leo which, in my Caltech family, I would not hesitate to call the Brothers.

Con particolare coinvolgimento, vorrei inoltre aggiungere una parte in Italiano per ringraziare tutte le persone che, anche senza un aiuto diretto sul lavoro mi hanno permesso di raggiungere i miei obiettivi, spronandomi nelle mie decisioni e aiutandomi a realizzare le miei idee. Fino dalla prima infanzia, la crescita della mia curiosità è stata trattata con particolare cura: avvincenti stimoli e la guida al ragionamento, scientifico senza sapere che lo fosse oserei dire, hanno fondato le basi della persona che sono oggi. Per questo, oltre che per gli innumerevoli aiuti, vorrei ringraizare la mia Famiglia: Babbo, Mamma, Gaia e anche Nonna Paola, Nonna Quirina e Nonno Massimo. Questo traguardo è anche vostro.

At this moment, Future is calling me towards new and different shores.
Nevertheless, the foundations of my identity are, and will remain, deeply rooted in this first third of a century of my life, forged by the people, the events, and the passions I enjoyed so far.

# ABSTRACT

The field of particle and fundamental physics finds itself now in a peculiar situation. The established Standard Model accurately predicts most of the observations, but several compelling reasons motivate a need for an extension of the current theory. In this thesis, I focus my research on facing the current situation of the field in a diversified threefold manner.

First, I develop methods based on physics-driven machine learning algorithms, with a particular focus on developing a model-independent tagger for unexpected events using artificial neural networks. This study shows how model-independent new physics triggers, possibly trained on real data, can select a low rate stream of events able to explore new physics processes up to a 10-100 pb cross section and can create a special dataset of rare unexpected events. Other important results from this body of work include the first application of the proposed anomaly detection strategy to real data, the use of graph neural networks to improve current pileup mitigation algorithms, the development of jet taggers based on the interaction network, and analysis-specific fast simulation.

Second, I focus on the methodological and hardware development of the MIP Timing Layer that is expected to upgrade CMS in preparation for HL-LHC. My seminal study demonstrates the possibility of using time-of-flight information to perform particle identification, which has a significant impact on heavy stable charged particle searches. This work introduces how to measure time-of-flight at CMS, a strategy for particle identification, and an algorithm to locate vertices in space and time. I also participated in the sensor testing and test beam operation. In particular, I conducted a study about the design and prototype of the detector modules' thermal behavior that shows how different geometries could lead to cooling differences of a few K.

Last, I direct my attention towards CMS's first lepton flavor universality tests with B meson decays. Using a dataset acquired thanks to a custom design trigger, I independently develop the measurement of the $\mathcal{R}(D^*)$ ratio, a parameter whose tensions between the predictions and observation have drawn remarkable attentions. I oversaw the complete mature state of the analysis, from the Monte Carlo simulation to the fitting procedure. Further collaboration-wide efforts are still required, but I demonstrate the expected sensitivity of about 15% using an Asimov dataset.

# PUBLISHED CONTENT AND CONTRIBUTIONS

[1] Olmo Cerri et al. "Variational autoencoders for new physics mining at the Large Hadron Collider." In: *J. High Energy Phys.* 2019.5 (2019), p. 36. DOI: `10.1007/JHEP05(2019)036`.
O.C. was the main contributor of the project. He participated in the conception of the seminal idea and developed and bench marked the main model presented in the work.

[2] Olmo Cerri et al. "Identification of long-lived charged particles using time-of-flight systems at the upgraded LHC detectors." In: *J. High Energy Phys.* 2019.4 (2019), p. 37. DOI: `10.1007/JHEP04(2019)037`.
O.C. was the main contributor of the project. He participated in the conception of the seminal idea, carried out the complete development of the work, and was the main writer of the manuscript.

[3] Thong Q. Nguyen et al. "Topology classification with deep learning to improve real-time event selection at the LHC." In: *Comput. Softw. Big Sci.* 3.1 (2019), p. 12. DOI: `10.1007/s41781-019-0028-1`. arXiv: `1807.00083 [hep-ex]`.
O.C. developed the dataset and the physics case.

[4] Oliver Knapp et al. "Adversarially Learned Anomaly Detection on CMS Open Data: Re-discovering the top quark." In: *Eur. Phys. J. Plus* 136.2 (2021), p. 236. DOI: `10.1140/epjp/s13360-021-01109-4`. arXiv: `2005.01598 [hep-ex]`.
O.C. supervised the project and co-directed the study as a followup to previous work.

[5] Arjona Martinez, J. et al. "Pileup mitigation at the Large Hadron Collider with graph neural networks." In: *Eur. Phys. J. Plus* 134.7 (2019), p. 333. DOI: `10.1140/epjp/i2019-12710-3`. URL: `https://doi.org/10.1140/epjp/i2019-12710-3`.
O.C. supervised the project and co-developed work in the study leveraging on its pileup knowledge from the master thesis.

[6] Eric A. Moreno et al. "JEDI-net: A jet identification algorithm based on interaction networks." In: *Eur. Phys. J. C* 80.1 (2020), p. 58. DOI: `10.1140/epjc/s10052-020-7608-4`. arXiv: `1908.05318 [hep-ex]`.
O.C. supervised the project, co-directed the study, and contributed to the algorithm performance bench-marking.

[7] Eric A. Moreno et al. "Interaction networks for the identification of boosted $H \to b\overline{b}$ decays." In: *Phys. Rev. D* 102 (1 2020), p. 012010. DOI: `10.1103/PhysRevD.102.012010`.
O.C. supervised the project, co-directed the study, and contributed to the algorithm performance bench-marking.

[8] Cheng Chen et al. “Analysis-specific fast simulation at the LHC with deep Learning.” In: *Comput. Softw. Big Sci.* 5.1 (2021), p. 15. DOI: 10.1007/s41781-021-00060-4.
O.C. supervised the project, co-directed the study taking care of creating the physics case and validating the results leveraging its knowledge to the application case of the R(D*) analysis.

# NON-JOURNAL PUBLICATIONS AND INTERNAL NOTES

[1] Olmo Cerri et al. "Lepton flavor universality test at CMS with R(D*) measurement in the full leptonic tau final state." In: *CMS AN-2019/162* (2022).
O.C. was the main researcher on this work.

[2] Wozniak, Kinga Anna et al. "New physics agnostic selections for new physics searches." In: *EPJ Web Conf.* 245 (2020), p. 06039. DOI: 10.1051/epjconf/202024506039. URL: https://doi.org/10.1051/epjconf/202024506039.
O.C. contributed to the algorithm development and to the performance benchmarking.

[3] Olmo Cerri et al. "Thermal studies on BTL prototype modules." In: *CMS DN-2020/013* (2020).
O.C. was the main researcher on this work.

[4] Adi Bornheim et al. "Geant4 simulation of the light yield and timing performance for BTL sensors." In: *CMS DN-2018/031* (2018).
O.C. collaborated in the data nalysis of the siumulations output.

[5] Adi Bornheim et al. "BTL sensor performance in testbeam campaigns at FNAL and lab source measurements." In: *CMS DN-2018/032* (2018).
O.C. contributed both to the hardware and software efforts.

[6] Adi Bornheim et al. "Thermal performance of the barrel timing layer for the CMS phase 2 upgrade." In: *CMS DN-2018/039* (2019).
O.C. helped supervise the studies and discuss the results.

[7] Si Xie et al. "Barrel MIP timing detector testbeam setup in 2018 and 2019." In: *CMS DN-2019/042* (2019).
O.C. contributed in several aspects from the hardware assembly, to the complete operation software development and the analysis software improvement.

[8] Olmo Cerri. *CMS precision timing physics impact for the HL-LHC upgrade*. Tech. rep. 2018. arXiv: 1810.00860. URL: http://cds.cern.ch/record/2641474.
O.C. presented this MTD performance overview on the behalf of the CMS collaboration.

# TABLE OF CONTENTS

# NOMENCLATURE

**CMSSW.** Offline software of the CMS experiment (see http://cms-sw.github.io).

**Hadron.** Composite subatomic particles made of bound states with two or more quarks.

**Hadronization.** Process of the formation of hadrons out of quarks and gluons. This non-pertubative process can be emulated in MC simulations using effective parameters fine tuned to experimental observations.

**Luminosity.** Measurement of the collision intensity as the ratio of the number of events in a certain period of time to the cross-section of those events.

*Chapter 1*

# INTRODUCTION AND BACKGROUND

*“Nos esse quasi nanos gigantium humeris insidentes, ut possimus plura eis et remotiora videre, non utique proprii visus acumine, aut eminentia corporis, sed quia in altum subvehimur et extollimur magnitudine gigantea."*
We are like dwarfs seated on the shoulders of giants. If we see more and further than they, it is not due to our own clear eyes or tall bodies, but because we are raised on high and upborne by their gigantic bigness.

— Bernardus Carnotensis, 1115 C.E. ca.

After the discovery of the Higgs boson in 2012, the standard model of particle physics (SM) has been further validated and all its free degrees of freedom have been established. Using the full set of SM parameters, it is thus possible to predict with increasing precision a wide variety of observables which are also measured in experiments. The agreement between measurements and theoretical predictions is a severe consistency test for the model: possible discrepancies would be a clear sign of shortfalls of the SM and the need for new physics to complement the current picture. Fundamental physics is now facing deep questions. On the one hand, the Standard Model of particle physics can predict the results of all the experiments we are doing in our laboratories with great precision. On the other hand, there are several theoretical questions and experimental observations which suggest that this model may only be a low-energy approximation of a more general theory. The Large Hadron Collider (LHC), the largest particle accelerator ever built, is delivering an incredibly large amount of collisions at the highest center of mass energy ever reached. This environment allows direct searches of new physics phenomena at high and unprecedented energy. For the time being, no new or unexpected observation had been made and the quest remains open. Together with direct searches, precise measurements represent the fundamental tools to test the consistency of the standard model and to possibly unveil effects brought by new physics at a larger energy scale.

The field of particle and fundamental physics finds itself now in a situation deeply different from the one experienced during the years between the 1970s and the 2010s when the Standard Model was being validated, and more similar, I believe, to the times right before the $J/\psi$ discovery [1, 2]. On one hand, the established model seems to be completely validated and able to provide solid predictions for most of the observed phenomena. On the other hand, several experimentally and theoretically compelling reasons require an extension of the current theory, but not a well defined model seems to be preferred. Without a clear path forward, the whole field seems to be holding its breath for an upcoming breakthrough able to enlighten a bit of the darkness that is surrounding its frontiers. In this situation, I found it relevant for the scientific quest I endeavored on in my PhD to maintain a strategic and differentiated approach by focusing on three directions: 1) to develop and enhance model-independent methods to look for unexpected evidence of new physics, 2) to upgrade our current detector leveraging on new technologies in order to open the possibility for new research paths, and 3) to focus on precision measurements which could help tighten our grasp of the standard model. During my time here at Caltech, thanks to the great opportunities provided by a diverse and rich environment I found in my supervisors and collaborators, I was able to contribute in all three of those directions. First, I contributed to developing methods based on physics-driven machine learning algorithms. After being involved in existing efforts [3], I focused on a study to develop a model-independent tagger for unexpected events using artificial neural networks for physics anomaly detection [4]. The study shows how variational auto-encoders can be used as model-independent new physics triggers, for which training directly on real data is foreseeable. I also collaborated on the first application of this proposed strategy to real data. In this work [5], the rediscovery of the top quark has been used as proof of the method's effectiveness. In addition, I took part in other studies aimed at expanding the application of machine learning methods to particle physics by direct collaboration or by supervising younger students [6, 7, 8, 9, 10, 11]. Second, I joined the Spiropulu group's involvement in one of the major upgrades that CMS is planning for High Luminosity-LHC (HL-LHC), the MIP Timing Layer (MTD). I took the lead in the seminal study to show the possibility of using time-of-flight information to perform particle identification [12]. As part of that work, we not only successfully introduced the time-of-flight for the first time in CMS and invented a strategy for particle identification, but we also wrote an algorithm to locate proton-proton interaction vertices in space and time. The impact of the work is showcased through proof of application in heavy stable charged particle searches. I have also been involved in the hardware activity related to the MTD sensor developments. I operated in both the group's labs on the Caltech campus and at FNAL, where I took part in four test beam campaigns[13, 14, 15]. Besides participating in the sensor testing, test beam operation, and emergency handling, I was the main developer of the data processing and analysis code which found a collaboration-wide deployment. After the test beam campaigns, I shifted my attention to the design and prototype of the detector modules with a specific focus on the thermal behavior [16, 17]. Finally, I have been involved in CMS's first lepton flavor universality tests with B meson decays. Using a special dataset, enriched with events of B-meson production acquired thanks to a custom design trigger, I played a central role in the CMS collaboration effort to measure the $\mathcal{R}(D^*)$ ratio which has shown tensions between the SM prediction and the current experimental measurement world average. As the only student involved in this analysis, I have been leading the analysis in every aspect: the Monte Carlo (MC) simulation, the event selection and B meson candidate reconstruction, and the observable selection and fitting procedure. Despite the mature state achieved by the analysis, further collaboration-wide efforts are still required to deliver a reliable measurement. However, here an Asimov dataset analysis is used to showcase the analysis's expected sensitivity.

This thesis is divided into five chapters. The remainder of the first chapter presents an overall picture of the theoretical and experimental basis for the work developed in the thesis. Starting from the foundations of the Standard Model, the Higgs mechanism and flavor structure are introduced. Then an overview of the LHC, which is currently operating at CERN, and the Compact Muon Solenoid (CMS) experiment is presented. The main features of the subdetectors are briefly described, together with the reconstruction algorithms. The original work developed during the thesis is discussed in the remaining chapters. The second chapter discusses the contributions to the development of a new machine learning based tool with a particular focus on the work about the variational autoencoders for new physics mining at the LHC. Other contributions are also briefly summarized at the end of the chapter. In the third chapter, the contributions to the methodological and hardware development for the new timing layer expected for the CMS phase II upgrade are presented. The fourth chapter is completely dedicated to an extensive overview of the study performed to deliver a $\mathcal{R}(D^*)$ measurement with the CMS data. Finally, the fifth and last chapter draws the conclusions of this thesis summarizing the main results, underlining the importance of the work, and suggesting possible future developments.

**Document notation**

Throughout this document, the notation commonly used in high energy physics is adopted. The value of the speed of light and the Plank constant is set to $c = 1$ and $\hbar = 1$, so that masses, energies, and momenta are all expressed in electron volts (eV).

## 1.1 The Standard Theory of Particle Physics

All the knowledge we have about the fundamental components of matter and their interactions is encompassed in a theory commonly referred to as the "Standard Model." The first foundations of the SM were laid in the 1960s [18, 19, 20, 21, 22], but its current form has emerged gradually in almost a century of theoretical and experimental investigations. The SM is now regarded as a fully established and predictive model that describes a wide range of phenomena and has been tested experimentally with remarkable accuracy. The SM is a quantum field theory whose Lagrangian is constructed by imposing a set of gauge invariance groups, from which force mediator bosons arise, and including the presence of several fermions and a scalar. Although all of the elementary SM particles have been directly observed with the predicted properties, there exist strong motivations to believe that the SM is not sufficient to describe our nature at all energy scales. For example, there is overwhelming experimental evidence for phenomena, such as neutrino oscillations, the presence of dark matter, and the acceleration of the universe, which are not explained within the SM. Also, profound theoretical questions, such as the hierarchy problem, the lack of an observed CP violation in strong interactions, and the matter-antimatter imbalance in the observed universe, can not be satisfactorily addressed within the current theory. The hope for our generation is that a paradigm shift or the discovery of new physics might help answer at least some of these open questions.

The aim of the following sections is to describe the SM accurately enough to serve as a theoretical background for this thesis. A more complete overview of this topic can be found in [23, 24]. For the sake of completeness, a summary of the description of the gravitational interaction, whose effect is typically negligible in the setup of particle physics experiments, can be found in [25].

### Symmetries and Fields

The Standard Model is a renormalizable non-abelian gauge quantum field theory, coherent with special relativity. In the SM there are three types of fields: fermions of spin ½ which represent the constituents of matter, bosons of spin 1 that represent the interactions, and a scalar boson that allows particles to be massive.

A gauge field theory is characterized by a certain group of symmetries $\mathcal{G}$, with associated generators $T_a$ satisfying the algebra

$$[T_a, T_b] = i f_{abc} T_c \ ,$$

and the representation $r_\psi$ and $r_\phi$ under $\mathcal{G}$ of fermionic fields $\psi$ and scalar fields

$\phi$. The symmetries in $\mathcal{G}$ must be internal, meaning that they leave the space-time degrees of freedom of a particle untouched, and local, meaning that the parameters of this transformation are functions of the space-time position. The interactions are fixed once the gauge symmetry groups are assigned. Indeed, when these symmetry groups refer to global (a global symmetry is also a local symmetry) transformations, satisfying Noether's theorem, they bring forth conserved currents to which the gauge fields are coupled to. These transformations are not physical and they express only the redundancy of the description of the massless gauge fields with Lorentz 4-vectors. The Lagrangian of the theory can then be built to satisfy these gauge symmetries by using covariant derivatives of the form

$$D_\mu = \partial_\mu - igT^a A^a_\mu,$$

where $A^a_\mu$ are the gauge vector fields and $g$ is the charge of the field under the interaction induced by the symmetry. Considering all the Lorentz-invariant, gauge-invariant, and renormalizable operators, the Lagrangian $\mathcal{L}_\mathcal{G}$ for a generic group $\mathcal{G}$ can be written as

$$\mathcal{L}_\mathcal{G} = \mathcal{L}_{min} + \left(\frac{1}{2}\psi^T M\psi + \phi\psi^T\Gamma_1\psi + \phi^\dagger\psi^T\Gamma_2\psi + \text{h.c.}\right) - V(\phi), \qquad (1.1)$$

where $\psi$ and $\phi$ are the fermionic and scalar fields, $M$ and $\Gamma_{1,2}$ are in general matrices (sometimes necessarily vanishing to guarantee the invariance under $\mathcal{G}$), $V(\phi)$ is the potential for the scalar field, and

$$\begin{aligned} \mathcal{L}_{min} &= -\tfrac{1}{4}F^a_{\mu\nu}F^a_{\mu\nu} + i\bar{\psi}\not{D}\psi + |D_\mu\phi|^2 \\ F^a_{\mu\nu} &= \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + g f_{abc}A^b_\mu A^c_\nu \end{aligned}$$

represent the dynamic evolution of the fields and the fields' strength tensor. In dimension 4, vector bosons can only appear as mediators since they are renormalizable if and only if they are associated with a gauge symmetry [26]. It has to be noted that no mass term for the gauge boson of the form $-m^2 A_\mu A^\mu$ is present since it would violate gauge invariance.

The gauge symmetry group of the SM is the direct product

$$\mathcal{G}_{SM} = SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$$

and the associated conserved quantities are the color, the weak isospin, and the hypercharge. It is interesting to note that for every simple non-abelian factor of the symmetry group there is only a single arbitrary constant (i.e. the coupling is the

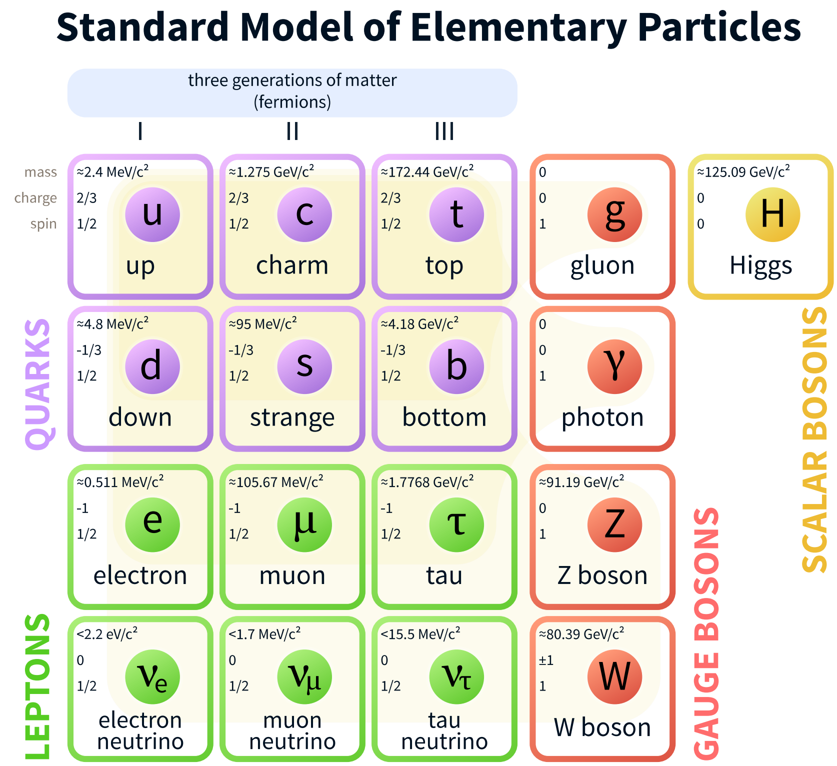


Figure 1.1: Fundamental particles present in the Standard Model. The 3 matter generations of fermions are shown in the first 3 columns, while gauge bosons and scalar boson are shown in the $4^{th}$ and $5^{th}$ column, respectively. Image from [27].

same for all the fields), whereas the charges under U(1) can assume independent arbitrary values for each field[1]. $SU(3)_C$, where C stands for color, is the group of Quantum Chromodynamics (QCD) and describes the nuclear strong interactions. This non-commutative group has 8 generators and thus generates 8 self-interacting vector bosons called gluons ($g$). $SU(2)_L \otimes U(1)_Y$ instead describes the electroweak interactions (EW), where the electric charge ($Q$) can be expressed in terms of the weak isospin ($T_3$) and of the hypercharge ($Y$) using the formula: $Q = \frac{1}{2}(Y + T_3)$. This subgroup produces 3 self-interacting vector bosons $\overrightarrow{W}$ and a neutral vector boson $B$. Thanks to the electroweak symmetry breaking (EWSB) discussed in the next section, these fields will manifest as the photon ($\gamma$) that mediates the electromagnetic interaction, and the $W^{\pm}$ and $Z$, mediators of nuclear weak interactions. The SM has only one complex scalar field $\phi$, whose remaining dynamic degree of freedom after the EWSB will represent the Higgs boson ($H$). The scalar field $\phi$ does not carry a color charge (i.e. is a single under $SU(3)_C$), is a double under $SU(2)_L$, and has unitary hypercharge. In the SM there are finally 12 spin ½ fields which

[1]This is actually only true up to the requirement of vanishing gauge anomalies [28].

differs for the representation of $\mathcal{G}_{SM}$ they fill and their Yukawa interaction with the scalar field (i.e. their masses). These fermionic fields can be organized into three

Table 1.1: Fields content of the SM from a gauge point of view, the three generations of fermions have not been distinguished since they have exactly the same gauge interactions.

| Particle | $SU(3)_C$ | $SU(2)_L$ | $U(1)_Y$ |
|---|---|---|---|
| $L = \begin{pmatrix} \nu \\ e_L \end{pmatrix}$ | 1 | 2 | -1/2 |
| $e_R$ | 1 | 1 | -1 |
| $Q = \begin{pmatrix} u_L \\ d_L \end{pmatrix}$ | 3 | 2 | +1/6 |
| $u_R$ | 3 | 1 | +2/3 |
| $d_R$ | 3 | 1 | -1/3 |
| $H$ | 1 | 2 | 1 |
| $g$ | 8 | 1 | 0 |
| $\vec{W}$ | 1 | 3 | 0 |
| $B$ | 1 | 1 | 0 |

sets, usually referred to as three generations of matter, each containing the same four kinds of field. Across generations, each field of a given kind differs from the other fields of the same kind solely because of the mass. As a consequence, all fermions of the same kind are predicted to have the same interactions and charges. Half of the fermions do not carry a color charge and are referred to as leptons. The other half, the quarks, fill the simplest non-trivial representation of $SU(3)_C$ and hence interact via the strong force. All the fermions couple to $SU(2)_L$ but only with their left-handed degrees of freedom, as indicated by the subscript $L$ for *left*. This asymmetry in the coupling is often called chirality and forces the fermion's mass term $\frac{1}{2}\psi^T M\psi$ to vanish in order to respect the gauge symmetry. Practically, while all right-handed fermions are singlets under $SU(2)_L$, the left-handed quarks and leptons pairs in each generation fill the doublet representation. The two quarks in the weak isospin doublet are simply called the up-type and the down-type quark and end up having an electric charge of +2/3 and -1/3. The quark fields across the three generations (generations also referred to as flavors), are respectively called up, charm, and top quark (up-type) and down, strange and bottom quark (down-type). The two leptons in the weak isospin doublet are called neutrino and charged lepton from the fact that the first kind ends up with no electric charge while the second one ends up with a unitary electric charge. The three charged lepton flavors are called electron, muon, and tau while the neutrinos simply take the names from the weak isospin doublet companion. While right-handed degrees of freedom are present for the quarks and the charged lepton, no right-handed neutrinos are present in the minimal SM. Finally, all fermions interact with $U(1)_Y$ with a hypercharge value such that the left and right degrees of freedom end up with the same electric charge.

Once the gauge symmetries and the field representations are fixed, we can explicitly write down the Lagrangian of the SM:

$$\mathcal{L}_{SM} = -\frac{1}{4}B_{\mu\nu}B^{\mu\nu} - \frac{1}{4}\sum_{k=1}^{3} W^k_{\mu\nu}W^{\mu\nu,k} - \frac{1}{4}\sum_{\alpha=1}^{8} G^\alpha_{\mu\nu}G^{\mu\nu,\alpha} \tag{1.2}$$

$$+ \sum_{\psi=\{Q,u_R,d_R,L,e_R\}} \sum_{i=1}^{3} i\overline{\psi}^i \not{D}\psi^i + (D_\mu\phi)^\dagger D^\mu\phi \tag{1.3}$$

$$- \sum_{i,j=1}^{3} \Gamma^{ij}_d \overline{Q}^i \phi d^j_R + \Gamma^{ij}_u \overline{Q}^i i\sigma_2\phi^* u^j_R + \Gamma^{ij}_\ell \overline{L}^i \phi e^j_R + h.c. \tag{1.4}$$

$$+ \mu^2\phi^\dagger\phi - \lambda(\phi^\dagger\phi)^2, \tag{1.5}$$

where $B_{\mu\nu}$, $W^i_{\mu\nu}$, and $G^\alpha_{\mu\nu}$ are the field strengths related to the hypercharge, weak, and strong interactions, respectively, the index $k$ runs over the three bosons of the weak $SU(2)_L$ group, $\alpha$ labels the eight gluons, and $i$ and $j$ label the three matter generations in the fermionic fields $\psi_i$. In principle, an infinite number of non-renormalizable terms with the operatorial part of the terms with mass dimension ($k$) larger than four can be added. These terms are suppressed with $1/\Lambda^n$, where $\Lambda$ is some energy scale and $n = k - 4$. At energies much smaller than $\Lambda$ the effects of these terms are negligible compared to the others, as they are suppressed by $(E/\Lambda)^n$. As there is no experimental evidence of these terms, they are not included within $\mathcal{L}_{SM}$. Such terms would provide evidence of new physics phenomena at a scale $\Lambda$, and hence the presence of a UV-complete theory that extends the SM but manifests itself in an approximate way at lower energies. This is exactly what happens in the theories of weak decays at energies much smaller than the $W$ mass. In the first line of $\mathcal{L}_{SM}$, we have the kinetic terms for the gauge fields, whereas the second line contains the kinetic term for the fermions and the scalar from which the matter interactions arise. The covariant derivative takes the explicit form:

$$D_\mu = \partial_\mu - i\frac{g'}{2}YB_\mu - igT_jW^j_\mu - ig_sS_\alpha G^\alpha_\mu, \tag{1.6}$$

where $B_\mu$, $W^j_\mu$, and $G^\alpha_\mu$ are the gauge fields for the hypercharge, weak, and strong interactions, respectively, with relative coupling constants $g_1$, $g_2$, and $g_3$. $Y$, $T_j$ and

$S_\alpha$ are the infinitesimal generators for $U(1)_Y$, $SU(2)_L$ and $SU(3)_C$ transformations, respectively. The fourth line presents the explicit form of the scalar potential introducing two arbitrary constants, $\mu$ and $\lambda$. Although $\lambda$ has to be positive in order to avoid a potential unbounded from below, $\mu^2$ could in principle assume positive or negative values. The functional form is dictated by the fact that only terms with an even number of scalar fields are $SU(2)_L$ invariant and that terms with more than four scalar fields are not renormalizable. Again, no right-handed neutrino is included in this formulation. Despite the fact that neutrinos are massive is well-assessed from flavor oscillation experiments [29], terms to generate neutrino masses are not present in the minimal SM since no experimental observations unequivocally prefer one of the many viable theories.

The description given so far presents the SM from its most fundamental point of view, focusing on the gauge symmetries and the field content. When considering its phenomenology, however, it presents two apparent discrepancies with the experimental observations. First of all, despite the fact that no mass term for gauge bosons and fermions is allowed, it is well established that the $W$ boson, the $Z$ boson, and all the fermions are massive. Second, the SM has four conserved charges in the EW sector, namely the hypercharge and the three weak charges related to three $SU(2)_L$ generators, but only a combination of these four, the electromagnetic charge $Q$, is explicitly conserved and the physical states are organized in multiplets of the electromagnetic and color charges. In order to reconcile this fundamental view with what we observe in nature, it is crucial to discuss the explicit form of the potential $V(\phi)$ (Eq. 1.5) and how it addresses these problems by inducing a spontaneous symmetry breaking.

**The Brout-Englert–Higgs Mechanism and the Flavor Structure**

The mechanism that allows the SM to generate the phenomenology we observe in nature is based on the idea that the symmetry group $SU(2)_L \otimes U(1)_Y$ is spontaneously broken. Spontaneous symmetry breaking (SSB), or symmetry realized à la *Nambu-Goldstone*, is defined as the situation in which the ground state of a system does not reflect the symmetries of the Lagrangian. A classical example of this phenomenon is ferromagnetism. In this case, the rotational invariance of the system is broken below the Curie temperature, due to the emergence of a magnetization towards a random direction, despite the Lagrangian describing a symmetric interaction. Another example, more similar to what happens in the SM, is the rest position of a ball initially placed on top of a cone. The potential seen by the ball and the initial state are symmetric under rotation around the axis of the cone. However, since the ball initially sits in a maximum of the potential, the ball will fall in a random direction down the side of the cone as soon as the system is left free to evolve. The system spontaneously reaches a ground state with the ball lying at the base of the cone which does not manifest the rotational symmetry of the potential.

In the SM the SSB mechanism occurs when $\mu$ from Eq. 1.5 assumes real and positive values. In this case, the potential for the scalar field presents a local maximum for $\phi = 0$ and a set of degenerate minima defined by

$$\phi^\dagger \phi = v^2 \equiv \frac{\mu^2}{2\lambda}.$$

Also referred to as the *EW scale*, $v$ is the only energy scale present in the SM since all the other terms do not contain any factor with a non-zero mass dimension. To better understand the phenomenology, it is possible to rewrite the 4 degrees of freedom of the scalar field $\phi$ in a way in which the fluctuations around the minimum are explicit:

$$\phi(x) = \begin{pmatrix} \phi_1^+(x) + i\phi_2^+(x) \\ \phi_1^0(x) + i\phi_2^0(x) \end{pmatrix} = \frac{1}{\sqrt{2}} \exp\left[\frac{i\sigma_i}{v}\pi^i(x)\right] \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix},$$

where fields are identified by the explicit space dependence, $\sigma_i$ are the Pauli matrices, $\pi^i$ represents rotational degrees of freedom and are referred to as Goldstone bosons, and $H$ represents a radial degree of freedom which is identified as the Higgs boson. In this representation, it is clear how the potential, which only depends on the modulo $|\phi| = \sqrt{\phi^\dagger \phi}$, is flat along the excitations of the Goldstone bosons and it instead constrains the radial excitation in a quadratic well. The flat directions identify the set of degenerate ground states, which are all linked by an $SU(2)$ rotation. In order to remove this redundancy, it is possible to choose a specific gauge fixing, called *unitary gauge*, by operating a transformation of the form

$$U(x) = \exp\left[-\frac{i}{v}\left(T_1\pi^1(x) + T_2\pi^2(x) + (T_3 - Y)\pi^3(x)\right)\right].$$

After the transformation $\phi(x) \to U(x)\phi(x)$, the Goldstone bosons result rotated away from the scalar field which takes the form

$$\phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix}$$

and the vacuum expectation value (vev)

$$\langle \phi \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}$$

is clearly invariant only under the unbroken $U(1)_{em}$ part of the $U(1)_Y \otimes SU(2)_L$ group identified by electric charge generator[2] $Q = \frac{1}{2}(T_3 + Y)$. The scalar potential

$$V(\phi) = -\mu^2\phi^\dagger\phi + \lambda(\phi^\dagger\phi)^2 = \frac{1}{2}(2\lambda v^2)H^2 + \lambda v H^3 + \frac{\lambda}{4}H^4$$

presents an explicit term for the Higgs boson mass $m_H = \sqrt{2\lambda v^2}$ and self interactions terms.

Computing the effect of the gauge transformation $U(x)$ on the EW bosons, it is possible to show that terms proportional to $v^2$ appear in the kinematic term $D_\mu\phi^\dagger D^\mu\phi$. These terms can be rearranged to provide mass terms for a combination of the EW bosons, whose generators are broken while leaving the gauge field associated with the unbroken $U(1)_{em}$ massless. The mass eigenstates are then the photon field $A$ and the weak interaction mediators $W^\pm$ and $Z^0$:

$$\begin{aligned} W^\pm &= \frac{1}{\sqrt{2}}(W^1 \mp W^2) & \qquad m_W &= g\frac{v}{2} \\ Z^0 &= \frac{1}{\sqrt{g^2 + g'^2}}(gW^3 - g'B) & \qquad m_Z &= \frac{v}{2}\sqrt{g^2 + g'^2} \\ A &= \frac{1}{\sqrt{g^2 + g'^2}}(gW^3 + g'B) & \qquad m_A &= 0. \end{aligned}$$

In the unitary gauge, the three Goldstone bosons $\pi^i$ are reabsorbed inside the $W^+$, $W^-$, and $Z$ fields to account for the fact that the initial spin-1 massless bosons only have two degrees of freedom but the massive ones also require an additional third degree for the longitudinal polarization. Having acquired a mass of the order of the vev, the exchange of virtual weak bosons can be approximated with four fermions point-like interaction at low energy scales. Thanks to this, the value of $v \approx 246$ GeV was known with high precision several years before the Higgs boson discovery. The Fermi constant, which is the coupling constant of the weak decays at low energy, can be expressed as $G_F = v^2/\sqrt{2}$. The Higgs boson discovery allowed, on the other hand, to disentangle the values of $\mu^2$ and $\lambda$.

Since the scalar field $\phi$ carries no color charge, the group $SU(3)_C$ remains unbroken and the gluons remain massless. At low energies, color interactions among quarks (the only color charged fermions) and gluons are strong since the coupling $g_S$ is of order one, hence no perturbative treatment can be done. At this scale, all the observable states are composite colorless systems. The most simple systems,

[2] For the scalar field, which has $Y = 1$, this explicitly takes the form $\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$.

called mesons, are made of symmetrical $\bar{q}^a q^a$ quark pairs and have integers spins. Three quarks systems are also possible in the complete anti-symmetric $\epsilon^{abc} q^a q^b q^c$ state, they are called baryons and have semi-integer spins. Composite particles made of bound states with two or more quarks are referred to as *hadrons*. In the ultraviolet regime ($\Lambda \gg 200\text{MeV}$), the behavior changes dramatically because vacuum polarization for this theory is such that the renormalized charge decreases with increasing energy. This phenomenon, called asymptotic freedom, allows a perturbative treatment of strong interactions at high energies.

Dirac mass terms for the fermions and interactions between them and the Higgs boson $H$ also arise explicitly after the gauge fixing. The Yukawa sector in Eq. 1.4 takes the form

$$\mathcal{L}_{SM}^{(\text{Yukawa})} = -\left[M_d^{ij} \bar{d}_L^i d_R^j + M_u^{ij} \bar{u}_L^i u_R^j + M_e^{ij} \bar{e}_L^i e_R^j\right] \left(1 + \frac{H}{v}\right) + h.c.$$

where $M^{ij} = \Gamma^{ij} v/\sqrt{2}$ is a $3 \times 3$ complex matrix proportional to the vev and the indices $i$ and $j$ run on the matter generation. Since the $\Gamma^{ij}$ matrices are generally non-diagonal, the fermion fields used so far are a linear combination of the mass eigenstates. However, it is known that for every $M$ there exists two unitary matrices $U_L$ and $U_R$ such that $U_L^\dagger M U_R = D$, where $D$ is a diagonal matrix. Using this fact, it is possible to redefine the fermion fields in order to have them matching the phenomenology of the observed mass states

$$\begin{aligned} \psi_L^i &\to U_L^{(\psi)ij} \psi_L^j \\ \psi_R^i &\to U_R^{(\psi)ij} \psi_R^j, \end{aligned} \tag{1.7}$$

where $\psi = u, d, e$ represents the three fermions types. After this rotation, the Yukawa sector becomes

$$\begin{aligned} \mathcal{L}_{SM}^{(\text{Yukawa})} &= -\sum_{\psi,i,j} \bar{\psi}_L^i \left[U_L^{(\psi)\dagger} M_\psi U_R^{(\psi)}\right]^{ij} \psi_R^j \left(1 + \frac{H}{v}\right) + h.c. \\ &= -\sum_{\psi,i} m_\psi^i \left(\bar{\psi}_L^i \psi_R^i + \bar{\psi}_R^i \psi_L^i\right) \left(1 + \frac{H}{v}\right) \\ &= -\sum_{\psi,i} m_\psi^i \bar{\psi}^i \psi^i \left(1 + \frac{H}{v}\right), \end{aligned}$$

where the Yukawa coupling to the Higgs has been reabsorbed in an explicit mass parameter which makes clear the diagonal structure and the proportionality of the interaction term with the Higgs boson with the mass itself. The rotation presented above is transparent to all the kinematic and to the dynamic terms of $\mathcal{L}_{SM}$ that do not mix the two components of the $SU(2)_L$ doublet because the two rotation matrices always contract with their conjugate to give the identity. The kinematic term remains $i\bar{\psi}\partial\!\!\!/\psi$ and so do the gluon-quark interaction $-ig_S S_\alpha G^\alpha_\mu \bar{\psi}\gamma^\mu\psi$. Among the electroweak terms, the *neutral currents* mediated by a mix of the hypercharge boson and the third diagonal generator of $SU(2)_L$ are unchanged

$$\mathcal{L}_{SM}^{(NC)} = e\bar{\psi}\gamma^m u Q\psi A_\mu + \sqrt{g'^2 + g^2}\bar{\psi}\gamma^\mu \left(T_3 - \frac{g'^2}{g'^2 + g^2}Q\right)\psi Z_\mu,$$

where $e = gg'/\sqrt{g'^2 + g^2}$ is the electric charge and $\psi = e^i, \nu^i, u^i, d^i$ is the Dirac spinor for all the fermions in the 3 generations $i = 1, 2, 3$. The only terms partially impacted by the rotation in Eq. 1.7 are the *charge currents* (CC), which are mediated by the first two generators of $SU(2)_L$, mixed into the $W^\pm$ bosons. In the lepton sector

$$\mathcal{L}_{SM}^{\text{(CC, lep)}} = \sum_{i=1}^{3} \frac{g}{\sqrt{2}} \bar{\nu}^i \gamma^\mu e^i_L W^+_\mu + h.c.,$$

the rotation can be reabsorbed in the definition of the neutrino fields $\nu$, which have no Yukawa term with the scalar field because no $\nu_R$ is present. As a consequence, the CC interactions and the mass terms are both diagonal in the same leptons base. Since all the SM terms do not mix leptons of a different generation, an accidental symmetry called *lepton flavor number* arises from the conserved quantity

$$N^i_L = \#(e^i) + \#(\nu^i) - \#(\bar{e}^i) - \#(\bar{\nu}^i).$$

The only terms actually affected by the rotation are the CC for the quarks which end up being non diagonal in the fermions mass eigenstates.

$$\begin{aligned}\mathcal{L}_{SM}^{\text{(CC, quark)}} &= \frac{g}{\sqrt{2}} \bar{u}^i_L U_L^{u\dagger,ik} \gamma^\mu U_L^{d,kj} d^j_L W^+_\mu + h.c.\\ &= \frac{g}{\sqrt{2}} V^{ij}_{\text{CKM}} \bar{u}^i_L \gamma^\mu d^j_L W^+_\mu + h.c.,\end{aligned}$$

where the complex Cabibbo-Kobayashi-Maskawa matrix

$$V_{\text{CKM}} = U_L^{u\dagger} U_L^d = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix}$$

has been defined to represent the non-diagonal CC interaction across matter generation that arises from the rotation used to set a fermion base corresponding to the phenomenologically observed particles in their mass eigenstates. By construction, $V_{\text{CKM}}$ is also unitary and hence, only depends on $(N-1)^2$ parameters where $N = 3$ is the number of flavors (i.e. matter generations) in the SM. Out of those 4 parameters, one is a complex phase which is the only source of CP violation in $\mathcal{L}_{SM}$. Out of the many possible different parameterizations for $V_{\text{CKM}}$, the one proposed in [30] is particularly useful to understand the magnitude of the off-diagonal elements, which allow the mixing among different flavors

$$V_{\text{CKM}} = \begin{pmatrix} 1 - \frac{1}{2}\epsilon^2 & \epsilon & \alpha\epsilon^3(\rho - i\eta) \\ -\epsilon & 1 - \frac{1}{2}\epsilon^2 & \alpha\epsilon^2 \\ \alpha\epsilon^3(1 - \rho - i\eta) & -\alpha\epsilon^2 & 1 \end{pmatrix} + O(\epsilon^4).$$

All the parameters have been experimentally observed to be of order 1, including $\epsilon = 0.225 \pm 0.001$ which sets the scale for the off-diagonal elements.

## 1.2 The CMS Experiment at LHC

This section presents an overview of the experimental facilities which are at the center of most of the work discussed in the thesis. The first part introduces the Large Hadron Collider (LHC), the particle accelerator that, by colliding two high energy proton beams, provides the experimental environment where the data are collected. A summary of the characteristics of the collisions (*events*) produced in LHC interaction points is also discussed. The second part describes the Compact Muon Solenoid (CMS) detector, which collects the experimental data from the LHC collisions. The text includes not only a description of the hardware of the CMS apparatus, but also outlines the details of the software, summarizing the algorithms and techniques adopted to reconstruct the particles produced in the events. Finally, the framework used to simulate and reconstruct events is described in the last section. An extensive and complete review of the LHC and CMS can be found in [31] and [32].

### The Large Hadron Collider

The Large Hadron Collider is a circular particle accelerator located underground at the French-Swiss border near Geneva. With its design maximum collision energy of 14 TeV, the LHC is the world's most powerful particle collider ever built and takes its name from the particles—protons and ions—which are there accelerated and brought into collision. It was built between 1998 and 2008 by the European Organization for Nuclear Research (CERN) in fulfillment of its core mission to provide particle beams for physics research, in collaboration with hundreds of universities and laboratories all over the world. The task of the LHC is to produce particle collisions at the energy frontier with the purpose of exploring open questions in fundamental physics that concern the basic laws that govern the interactions among elementary particles. The LHC provides experimental conditions similar to those existing in the first instants after the Big Bang, and allows the exploration of phenomena beyond the EW scale up to the TeV scale. The LHC succeeded the Tevatron [33], which was a circular proton-antiproton ($p\bar{p}$) collider operating at Fermilab until 2011 and was able to reach a center of mass energy $\sqrt{s} \sim 2$ TeV. The LHC is located about 100 meters underground beneath the CERN site in a 27 km long tunnel, the same tunnel previously used by the Large Electron Positron (LEP) collider [34]. Its size makes the LHC the largest machine mankind has ever built.

The core of the LHC is made of two adjacent parallel vacuum beam pipes, where the two proton or ion beams circulate in opposite directions in an ultra-high vacuum of about $10^{-11}$ mbar. The beams are steered in their orbits inside the pipes by a magnetic field up to 8.33 T strong, which is provided by 1232 superconducting dipole magnets that bend the particles' trajectory, and 474 quadrupole magnets that focus and squeeze the beams. All the LHC magnets are superconducting magnets made of copper-clad niobium-titanium, and approximately 96 tons of superfluid helium-4 are needed to keep them at their operating temperature of 1.9 K. In order to accelerate the particle, 16 radio-frequency (RF) cavities are placed along the beam pipes and are used to transfer about 2 MV per cavity to the orbiting particles. Before entering the LHC, due to magnetic field constraints, protons need to be grouped in bunches of about $10^{11}$ particles and accelerated to a minimum energy of 450 GeV. This is achieved through a chain of 4 accelerators present at the CERN site (fig. 1.2): a linear accelerator (LINAC), a Booster ring (PSB), the Proton Synchrotron (PS), and the Super Proto Synchrotron (SPS) that directly injects particles in to the LHC. Once they enter the LHC, the particle beams need about 20 minutes to reach their

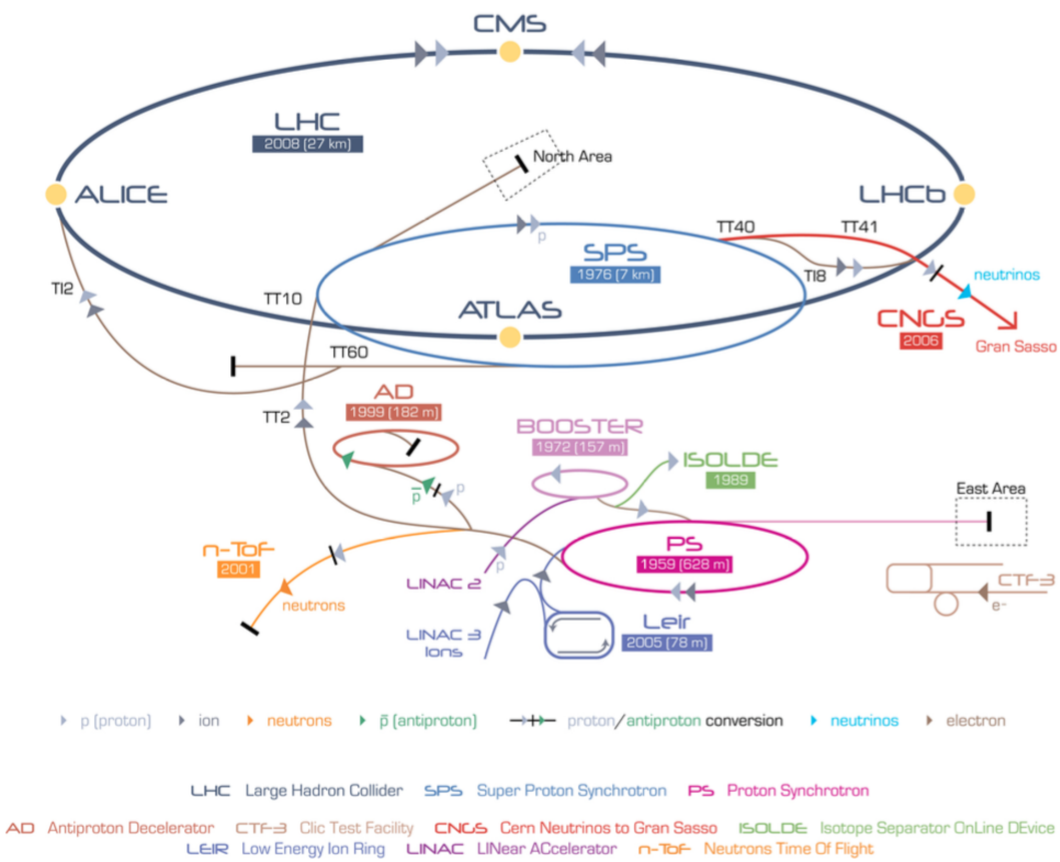


Figure 1.2: Scheme of the facilities of the CERN acceleration complex from [35].

maximum energy, with the bunches having passed through the RF cavities more than 10 million times. The synchrotron radiation of the protons is several orders of magnitudes smaller than the energy provided by the RF cavities, and thus the energy is capped by the ability to keep the particles in orbit. The radius of the LEP tunnel and the maximum magnetic field produced by the magnets are the limiting factors for the LHC top design energy of $\sqrt{s} = 14$ TeV. However, despite all of LHC's magnets were commissioned to a collision energy of over 14 TeV, no physics run has ever been run with an energy above $\sqrt{s} = 13$ TeV. This is because it has been observed that some dipole magnets have a lower memory than expected, demanding a larger number of quenches to reach nominal field. Retraining these magnets to 13 TeV require only a short period of time, whereas retraining to 14 TeV would take longer, taking time away from physics research. That's why it has been decided that, to speed the route to potential new physics, the optimal delivery of particle collisions for physics research is to run operation at 13 TeV, an energy anyways considerably higher than ever achieved before.

Since protons are composite particles, the effective energy available in the actual subatomic partons scattering process is smaller than 14 TeV. For example, a proton is a bound state $uud$ of two up and one down quark. These quarks are called *valence quarks* to distinguish them from the gluons and the quark anti-quark pair, which are part of the QCD sea. At the LHC energies, valence quarks carry in total about ½ of the energy, so $\frac{1}{6}\sqrt{s}$ is a rough estimate of the available energy. As a result, the effective physics reach for direct observations at LHC is at the energy scale of a few TeV. One of the main improvements of the LHC with respect to the Tevatron is the significant increase in luminosity. A higher luminosity means producing a larger number of collisions, which is crucial for exploring rare processes and reducing statistical uncertainties. This is mainly achieved through a tighter bunch spacing of about 25 ns, a higher number of protons per bunch, and advanced magnetic focusing optics. The luminosity $\mathcal{L}$ depends on the beams' parameters, and ultimately on the accelerator machine specifics, as

$$\mathcal{L} = \frac{fkn_p^2}{4\pi\sigma_x\sigma_y} = \frac{fkn_p^2}{4\beta^*\epsilon_n},$$

where $f$ is the revolution frequency, $k$ is the number of bunches ($\sim$ 3000 for LHC), $\sigma_x$ ($\sigma_y$) is the beam size on the $\hat{x}$ ($\hat{y}$) axis, $\beta^*$ is the amplitude function that expresses the squeezing at the interaction points, and $\epsilon_n$ is the normalized emittance that expresses the cross-sectional speeds in terms of a small angle regarding the direction of the beam. The LHC optics parameters at peak performance have a value of $\beta^* = 0.55$ m and $\epsilon_n = 3.75$ $\mu$m $\cdot$ rad, which translate into a peak instantaneous luminosity of $\mathcal{L} \approx 10^{34}$ cm$^{-2}$s$^{-1}$. The luminosity is directly related to the rate of events $dN/dt$ produced by a particular process following $N = \int \mathcal{L}(t)\sigma dt$, where $\sigma$ is the process cross section. The LHC luminosity decays with time from its peak value at injection time due to several effects. First of all, the luminosity is sizeable when compared to the total $pp$ cross section $\sigma_p p = 100$mb, which means that a large number of soft QCD $pp$ interactions (often called *minimum bias* from the name of the trigger used to register them) happen at every bunch crossing. These interactions slightly reduce the number of protons in each bunch and degrade the overall beam quality. Second, protons have additional interactions outside the bunch crossing region, mainly with the residual gas inside the beam pipe and intra-beam interactions among themselves inside a bunch. Third, the magnetic optics have limited efficiency for the orbit correction and focalization systems, so protons in the tails of the bunch are slowly shed away. As a result of these effects, the luminosity lifetime of the beam is approximately 15 hours, out of which only about 10 hours are typically used for physics runs. The amount of data available for the analyses is quoted by the integrated luminosity $L = \int \mathcal{L}dt$, and is measured in inverse picobarn (pb$^{-1}$).

Inside the LHC, particle collisions are induced by crossing the two beam paths in four points around the orbit. In each one of these regions, one of the four LHC experiments is hosted. The two general-purpose detectors, designed to explore the Higgs boson and the energy frontiers, ATLAS [36] and CMS [32], sit respectively below the CERN Meyrin site and at *point 5*, the opposite point on the LHC circumference. The other two experiments are LHCb [37], designed to study the flavor physics of bottom and charm quarks, and ALICE [38], dedicated to heavy-ion physics. As I am part of the CMS collaboration, my thesis focuses on proton-proton ($pp$) collisions recorded by CMS or simulated with a CMS-like detector. The LHC started the first physics data delivery in the spring of 2010, a period called Run 1, at a center of mass energy of 7 TeV. In these conditions CMS collected the first-ever 6 fb$^{-1}$ of integrated luminosity. In 2012, the center of mass energy was increased to 8 TeV, and CMS collected about 22 fb$^{-1}$. At the beginning of 2013, the LHC was shut down to prepare for Run 2, when the center of mass energy and luminosity were increased. From the LHC restart in early 2015 to the second shutdown at the end of 2018, CMS collected about 163.6 fb$^{-1}$ at a center of mass energy of 13 TeV. After a second shutdown, starting in 2019 and prolonged by the COVID-19 pandemic, LHC is now restarting operations and is expected to deliver the first beams in summer 2022 with the goal of doubling the luminosity integrated so far by the end of the run in 2025. A summary plot of the integrated luminosity by CMS is shown in Fig. 1.3. In January 2026, LHC will go down for its third long shutdown and will start a major

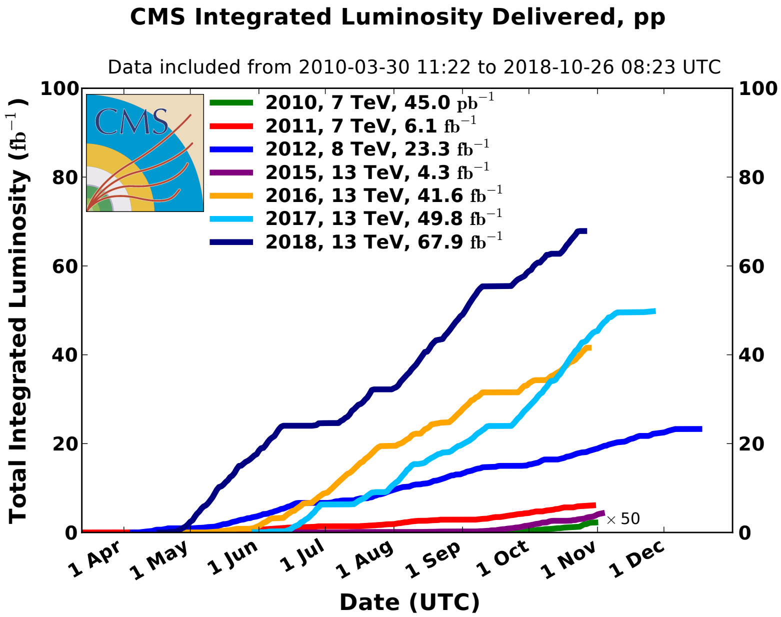


Figure 1.3: Cumulative luminosity versus day delivered to CMS during stable beams and for p-p collisions. Figure from [39].

upgrade, the High Luminosity LHC (HL-LHC [40]), which is expected to have a projected luminosity 10 times higher than LHC.

### Phenomenology of Proton-Proton Interactions

The proton-proton interaction at the intersection points of LHC is a complex phenomenon that involves the strong interaction of quarks and gluons composing the protons. At very low transferred momentum ($q^2$), protons behave like point-like particles, but for $q^2$ above $\sim$ 10 GeV a radically different picture appears. In this regime, thanks to the QCD asymptotic freedom, protons can be modeled as a bunch of loosely bound point-like particles called *partons*, each of whom carries a certain fraction $x$ of the whole proton energy.

Following this picture, the production cross section of a given final state from a $pp$ collision is the sum of the cross sections of all possible interactions among partons that can produce the final state, weighted by their probability:

$$\sigma(pp \to X|s) = \sum_{i,j} \int dx_i dx_j f_i(x_i) f_j(x_j) \hat{\sigma}(i + j \to X|\hat{s} = x_i x_j s)$$

where $i$ and $j$ run on the partons from each of the two protons, $X$ is the final state produced with cross section $\hat{\sigma}$, $f_i$ is the probability density function for the parton $i$ to have a certain fraction of the proton energy $x_i$, and $\hat{s}$ is the partonic center of mass energy. These probability functions are known as Parton Distribution Functions (PDFs) and have to be determined experimentally. The knowledge of these PDFs is one of the theoretical systematic uncertainties affecting LHC precision measurements. Even if the two proton beams have the same momentum in the laboratory frame, the actual energy fraction that the two partons taking part in the interaction have is in general different. The initial state is then not at rest in the laboratory frame along the $z$ direction while remaining approximately at rest in the transverse plane, besides next to leading order corrections due to proton scattering angle and QCD effects. For this reason, in $pp$ physics the transverse plane is often used to study the properties of the collisions and the transverse momentum is used as the energy scale of the interaction produced.

When two protons interact with each other there is usually one hard interaction, called *hard scattering*, and several other softer interactions that happen at the same time among the other partons inside the two protons. Given the nature of the strong interaction, initial and final state radiation of gluons is very common and the soft scattering of the remnants of the protons must be also considered. All these interactions together are called *underlying event* and can sometimes be detected in CMS. Due to color confinement, gluons and quarks produced in the collision of the protons do not propagate freely but create several quarks and anti-quarks pair from the QCD vacuum until recombination into *colorless* composite hadrons is achieved. This process, known as *hadronization*, involves non-perturbative QCD and can not, therefore, be easily predicted analytically. However, there exist phenomenological models, that using effective parameters tuned to reproduce the experimental results, can be used to produce Monte Carlo simulations of hadronic showers.

Whenever two proton bunches cross each other at the LHC, it is common that more than a pair of protons interact. Along with the main interaction that usually causes the detectors to trigger the data acquisition, the event is accompanied by several softer collisions happening between other protons in the bunches. This happens because the cross section for a soft QCD process in a proton-proton collision is several orders of magnitude larger than the process of interest. For example, the total $pp$ hard scattering cross-section, dominated by soft QCD interactions, is $\sim$ 100 mb while the production of $b$ quarks is around 0.5 mb and the electroweak physics lays below $\sim$ 100 nb. The number of these additional interactions is randomly distributed according to a distribution dictated by the LHC beam instantaneous parameters. In particular, at the CMS interaction point, the number of these additional non-interesting $pp$ collisions, called *pile-up* (PU), varies in the different runs and tends to decrease throughout an LHC fill. The average PU per bunch crossing is usually of the order of 10 and reaches about 30 in the highest luminosity runs. Pile-up produces a large number of soft particles which can be confused with the particles from the hard scattering and subsequent hadronization. Unfortunately, the presence of pile-up is the direct consequence of the trade-off between the requirements of high luminosity and the experimental ability to reconstruct interesting events.

### The CMS Detector

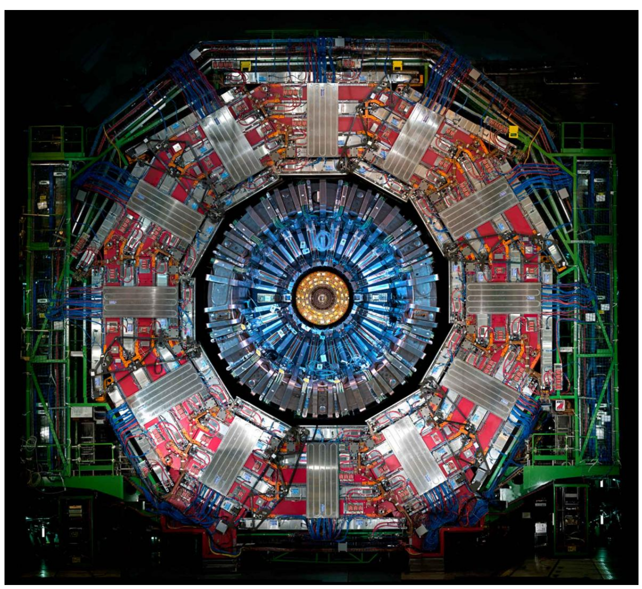

Figure 1.4: Picture of CMS in its location at point 5 of LHC. This section of the detector shows in full the 15 m of the outer diameter. (Image: Michael Hoch/Maximilien Brice)

The Compact Muon Solenoid (CMS) experiment is one of two large general-purpose detectors built at the LHC. The goal of the CMS experiment is to investigate a wide range of physics, including the measurement of the Higgs boson properties and of the other SM particles, search for new physics such as dark matter or Supersymmetry, and measure QCD properties in heavy ions collision. The CMS detector [32] has a cylindrical shape, centered at the nominal point where the LHC beams collide. It consists of a central cylindrical part called barrel and two external discs called endcaps, placed as the ends of the cylinder. CMS dimensions are impressive: it is about 20 m in length and about 14 m of total height. CMS is equipped with a 3.8 T superconducting solenoidal coil coaxial with the detector and the beams. A picture of the detector as well as an overview of the CMS subdetectors is present in Figures 1.4 and 1.5. The standard reference frame used in CMS is a right-handed

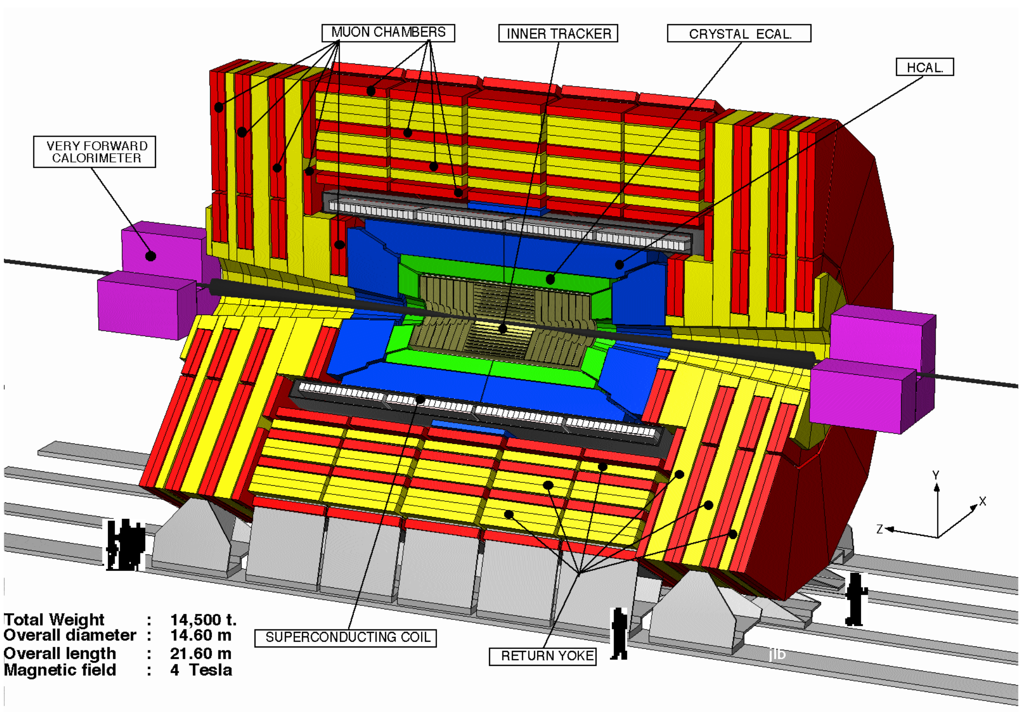


Figure 1.5: Longitudinal view of the CMS detector. Picture from `http://www-collider.physics.ucla.edu/cms/`.

Cartesian coordinate system with its origin in the geometrical center of the solenoid. The $\hat{x}$ axis points towards the center of the LHC ring, the $\hat{y}$ axis points upwards, and the $\hat{z}$ axis points along the beam line. A cylindrical coordinate system is more often used, described by the $r = \sqrt{x^2 + y^2}$ coordinate pointing from the axis of the cylinder outwards, and two angles defined by $\tan(\phi) = y/x$ and $\tan(\theta) = r/z$.

Instead of the angle $\theta$, the pseudorapidity,

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

is more commonly used, since it is additive under boosts along the $\hat{z}$ axis and it corresponds to the rapidity for mass-less particles. The rapidity is defined as

$$Y = \frac{1}{2}\ln\left(\frac{E + p_z}{E - p_z}\right)$$

where $E$ is the particle energy and $p_Z$ is the particle momentum along the beam axis. The plane identified by the $\hat{x}$ and $\hat{y}$ directions is referred to as *transverse plane*.

### The Layers Structure

The CMS detector is composed of modular subdetectors organized in a cylindrical nested layer structure (Fig. 1.6). Starting from the interaction point and moving outward along the cylindrical radius, CMS has inside the superconductive coil the following subdetectors: the silicon tracker, the electromagnetic calorimeter (ECAL), and the hadron calorimeter (HCAL). Outside the coil, a region reached only by muons and escaping neutrinos, CMS has only the muon detection system, which is embedded in the iron return yoke of the magnet. In the barrel, layers have cylindrical geometry coaxial with the beam whereas in the endcaps they have a disk geometry with their face perpendicular to the beam.

**Tracker** The tracker is the innermost subdetector and its goal is to accurately measure the trajectory of charged particles. The tracker's sensitive component is a silicon semiconductor waffle arranged in a $p-n$ or $n-n$ junction. When a charged particle crosses the material, it creates electron-hole pairs by knocking electrons off the silicon valence band. Thanks to the presence of a polarizing electric field, the electrons, and the holes are separated and collected by the electrodes, producing a signal proportional to the energy lost by the passing particle. This small pulse of current lasting a few nanoseconds is then amplified and collected to form what is called a *hit*. The hits are used as digital logic to reconstruct the particle's path. The CMS tracking system is divided into 2D tiles which present silicon pixels or strips with a pitch of the order of 100$\mu$m. The pixel modules are arranged into 3 concentric layers closed by two endcap disks on each side. Outside the pixel layers, the strip modules are arranged in 10 barrel layers and 3+9 endcap disks. The CMS tracker has a total of 66 million pixels and 10 million strips with the innermost pixel layer

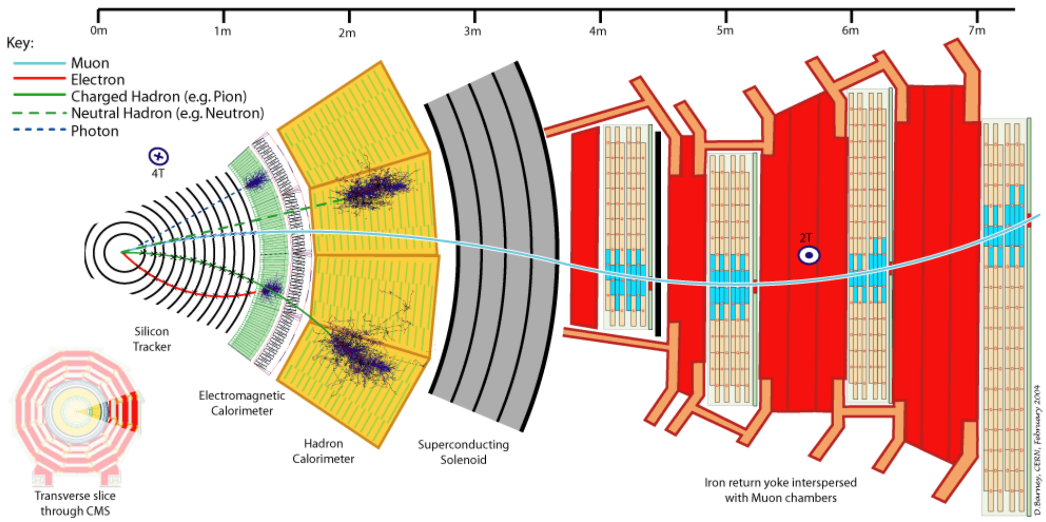


Figure 1.6: A transverse slice of the CMS detector, with the qualitative experimental signatures of the different particles. Picture from [41].

placed just 4.4 cm from the beam and the outermost strip layers placed about 1.2 m from the beam. The tracker's typical resolution of $10 - 50$ $\mu$m combined with the bending power of the coil allows CMS to measure the track transverse momentum with percent precision. The CMS tracker acceptance is limited to $|\eta| < 2.5$, where the more forward endcaps modules are placed.

**ECAL** The electromagnetic calorimeter is situated outside the tracker. The ECAL is a homogeneous calorimeter, segmented only in the $\eta - \phi$ plane, designed to absorb particles like electrons and photons. The ECAL is composed of scintillating PbWO4 crystals of about 3 cm $\times$ 3 cm $\times$ 20 cm, with an excellent resolution for energies ranging from 1 GeV to the TeV range. In the endcaps, a pre-shower detector (lead and silicon strips) is used to improve the $\pi^0$ rejection and to help identify the interaction vertex. The ECAL energy measurement resolution has a dependence on energy itself but typical values are of the order of 2% in the whole $|\eta| < 3$ acceptance.

**HCAL and HF** A second heavier calorimeter, called HCAL, meant to absorb hadrons is placed outside ECAL. The HCAL is a sampling calorimeter made of layers of brass and scintillator plates that extends up to $|\eta| < 3$. The HCAL thickness is limited to 1.2 m by the requirement to fit inside the coil. An additional calorimeter (HF) in quartz and scintillating fibers is installed to extend the acceptance up to $|\eta| < 5$. The energy resolution obtained using this detector is about 10-20% for high-energy particle jets.

**Muon system** The muon system is placed outside the coil and is composed of different types of gas detectors embedded in the return yoke of the magnet. In the barrel, up to $|\eta| < 1.3$, the muon system consists of four layers of drift tube chambers while in the endcaps, between $0.9 < |\eta| < 2.4$, cathode strip chambers are used due to the higher flux. Resistive plate chambers (RPCs) are also installed in front of most muon stations up to $|\eta| < 1.6$ in order to get a faster signal suitable for triggering. RPCs are parallel plate gaseous detectors that combine an adequate position resolution with a very fast response time. Muons with energies below a few hundred GeV lose energy in the matter only through ionization and can easily traverse all the CMS detector. Nevertheless, a muon in the barrel loses on average 3 GeV of transverse momentum before it reaches the first muon station, which is why muons with lower energy can not be identified. Since muons are the only kind of detectable particle surviving the whole detector, they represent the cleanest signals available at LHC and the muon system is crucial for their identification.

**Trigger** The rate of data produced from LHC events at CMS is much higher than the available bandwidth of the data processing chain and it would also be impossible to store every single produced event. It is enough to consider that CMS raw output is roughly 10 millions of TB a year (0.5 MB/event $\times$ 400 MHz $\times$ 6 months), which corresponds to more than 10 times the entire world's storage capacity filled up in a single year of data acquisition. As a consequence, it is necessary to deploy a trigger system that applies an online selection of the events to be kept before they can be definitely stored. The CMS triggering system is divided into two steps. The first one, called the Level-1 Trigger (L1), is implemented in hardware and, due to readout rate constraints, exploits only the information from calorimeters and muon chambers. This first step is completed in about 4 $\mu$s and it is tuned to reduce the event rate from the LHC bunch crossing rate of 40 MHz to approximately 100 kHz. An event accepted by the L1 is passed to the High-Level Trigger (HLT) which, through a simplified version of the software used for offline analyses, reconstructs the event using the full detector information and makes decisions based on high-level physical quantities. The HLT reduces the event rate further down to a few 100 Hz. Each full set of requirements applied to an event to survive the CMS trigger system is called *trigger path* which can be divided into an L1 seeding and an HLT path. The CMS trigger system also includes special paths designed to extend the CMS physics reach beyond the original design. In particular, *parking* paths are designed to save events with less stringent requirements in a separate slower pipeline that is activated when the CMS computing system is working below full capacity.

### Physics Object Reconstruction and the Particle Flow algorithm

Once the detector records an event, a great deal of offline work is still needed in order to extract the interpretable physics information from the registered raw electronic data. This crucial phase is called reconstruction and leverages the expected behavior of particles passing through the detector.

After being produced in the interaction region, the particles first enter the tracker, whose information is used to reconstruct charged-particle trajectories (tracks) and their origins (vertexes). Given the presence of the magnetic field in the tracker region, the charged particles propagate following a helicoidal trajectory which can be described by 5 parameters: the curvature $q/p_T$, $\phi$, $\eta$, the transverse impact parameters $d_{xy}$, and the longitudinal impact parameter $d_z$. The algorithm used in CMS for track reconstruction [42] is divided into three steps: track seeding, track finding, and track fitting. During the seeding, a loop on all pairs of hits compatible with some kinematic cuts selects segments of possible trajectories. The seeding starts from the innermost pixel detectors since the high resolution on the position of the hit reduces the number of options to consider. The track finding and fitting steps are based on a Kalman filter pattern recognition approach, which is initiated with the seed parameters. The track state given by the 5 parameters is propagated layer by layer with an inside-out approach and fitted to the hits using a minimum $\chi^2$ criterion. When all the layers are taken into account, a quality selection based on the $\chi^2$ and the number of missing hits is applied before producing the full parameters and covariance matrix output. In order to fully exploit the performances of the CMS tracker, a very precise alignment procedure is needed. This extremely complicated procedure parameterizes each tracker module with 3 spatial coordinates, 3 angular rotations, and 3 quadratic bending parameters. A mixed set of LHC collisions and cosmic rays is used to optimize the module parameters by running a secondary fit on the whole set of tracks. As a result, the position of each module is known with a resolution smaller than the typical strip or pixel size of the module. The track reconstruction step also provides the necessary input for the determination of the vertexes in which the $pp$ interactions took place, called primary vertices. The primary vertex reconstruction [43] is divided into two steps. First, a clustering algorithm is used to group tracks into compatible sets using the $z$ projection of the perigee point of the tracks with respect to the beam axis. Then, a linearized approach is used to fit the best 3-dimensional position of the vertex.

After passing through the tracker, particles moving outward encounter the ECAL. Here, electrons and photons are absorbed and the corresponding electromagnetic showers are detected as clusters of energy recorded in neighboring cells. The clusters are used to determine the energy and the direction of absorbed particles. Charged and neutral hadrons may initiate a hadronic shower in the ECAL as well, but they are mainly absorbed in the hadron calorimeter HCAL. The corresponding clusters are used to estimate their energies and directions. Finally, muons and neutrinos traverse the calorimeters with little or no interactions. Muons stand at the very center of the CMS physics focus and are a crucial part of many physics analyses. They are the only particles that can traverse the whole detectors and leave clear signatures of additional hits in the muon system. Moreover, they can not be produced directly in soft QCD interactions that dominate the physics production at LHC and constitute a signature for interesting events. For this reason, they are one of the main handles for triggering. In CMS the information left by the muons is at first reconstructed independently in the tracker (tracker muons) and in the muon system (standalone muons). Tracker muons have the advantage of traversing only through the very light and well-measured material of the CMS tracker but are limited by the short lever arm of 4 Tm bending Differently, standalone muons exploit the full lever arm of 12 Tm by including the beamspot constraint but are limited by multiple scattering. In a second iteration, the muon deposit in the two systems is combined with a Kalman filter in the two possible outside-in or inside-out ways. Global muons are formed by starting from a standalone muon and searching for a matching track. This gives major improvements in the measurement of the momentum for very high-energy muons, above 200 GeV. Soft muons, however, are formed by extrapolating the tracker track to the muon system taking into account all the detector effects and searching for at least a matching segment in the muon chambers. This allows the identification of low-energy muons that stop inside the muon chambers. Differently from all the other particles, neutrinos escape completely undetected. Using the fact that the initial state is at rest in the transverse plane, it is sometimes possible to deduce the momentum of the neutrino by summing all the other particles' momenta, using the quantity referred to as missing energy:

$$\vec{\not{E}}_T = 0 - \sum_i \vec{p}_T^{(i)}$$

where $\vec{p}_T = \vec{p} - (\vec{p} \cdot \hat{z})\hat{z}$ is the transverse momentum and the index $i$ runs over all the produced visible particles. Unfortunately, given the presence of PU, the impossibility of assigning neutral particles to a specific vertex, and the finite acceptance of the detector, the kinematics can be closed only up to a few tens of GeV, which makes this approach unsuitable for soft physics.

**Particle Flow algorithm** A significantly improved event description and an approximate identification of the different kinds of particles can be achieved by linking the information present in all the subdetectors. This approach is called particle-flow (PF) reconstruction [44] and heavily leverages on the following CMS detector features: a finely segmented tracker, a fine-grained electromagnetic calorimeter, a hermetic hadron calorimeter, a strong magnetic field, and an excellent muon spectrometer. The PF algorithm proceeds in four steps. The first step is clustering detector cells in sets called PF elements. Inside the calorimeters, (preshower, ECAL and HCAL) adjacent deposits compatible with a shower from a single particle or jet are clustered, while for the tracker the tracking procedure discussed above is performed. The second step is a linking algorithm that connects the information from different subdetectors to form PF blocks. After sorting in the $\phi - \eta$ plane, nearest-neighbor pairs of PF elements are considered and a distance, aimed at quantifying the quality of the link, is defined. The specific conditions required to link two elements depend on their nature. For example, a track and a calorimeter cluster are linked if the extrapolation of the track ends is within the cluster area. Different PF blocks may be associated with each other either by a direct link or by an indirect link through common elements. The third step, performed in each PF block, is the identification and reconstruction sequence which proceeds in the following order. First, muon candidates are identified, reconstructed, and the corresponding PF elements are removed from the PF block. Then, the electron identification and reconstruction, which tries to recollect also bremsstrahlung photons, is performed together with the treatment of energetic and isolated photons. Like for the muons, the corresponding tracks and ECAL or preshower clusters are excluded from further consideration.

The remaining elements in the block are then subjected to a cross-identification of charged hadrons, neutral hadrons, or photons, arising from parton fragmentation, hadronization, and in-flight decays. At the end of this procedure, all information is expressed in the form of PF candidates with one of the following labels: $e$, $\mu$, $\gamma$, $\pi^{\pm}$, and $K^0$. The $\pi^{\pm}$ id is used to identify a general charged hadron, while the $K^0$ id is used to identify neutral hadrons.

**Simulation**

Monte-Carlo simulation (MC) of physic processes and detector effects is one of the most used tools in modern physics analyses. Given the good level of description that simulation programs have reached, they provide a clean and fully controlled environment in which to develop new techniques and even understand the physics underlying real observation. On top of this, simulations are the only viable way to compare data and predictions in the harsh environment of hadron colliders. The CMS collaboration has its own dedicated simulation group which provides software [45] for event generation, detector simulation, and signal digitization with high fidelity to real performances. Simulation of physics processes proceeds through three steps: first, the physics of the hard scattering is simulated through a Monte Carlo program (e.g. MadGraph [46]) and saved in some standardized format; next, a second (sometimes the same) MC (i.e. Pythia [47]) takes care of the hadronic environment, color reconnection, and hadronization; then, mostly in the case of heavy flavored hadrons, an additional MC specialized for particle decays (i.e. EvtGen [48]) is called; finally, events are passed through a Geant4 [49, 50] detector simulation, which is also in charge of simulating long-lived particle decays happening outside a small region of a few microns. Interactions with the whole detector and readout electronics are also simulated in the last step. The final product of the simulation procedure is thus a simulated event with information and format equivalent to a real data event, and the addition of MC truth information such as undetected particles and true generated quantities. The MC produced in this way is then passed through the same trigger, reconstruction, and analysis chain as the real data.

## References


[1] J. J. Aubert et al. "Experimental Observation of a Heavy Particle $J$". In: *Phys. Rev. Lett.* 33 (23 1974), pp. 1404–1406. DOI: 10.1103/PhysRevLett.33.1404. URL: https://link.aps.org/doi/10.1103/PhysRevLett.33.1404.

[2] J. -E. Augustin et al. "Discovery of a Narrow Resonance in $e^+e^-$ Annihilation". In: *Phys. Rev. Lett.* 33 (23 1974), pp. 1406–1408. DOI: 10.1103/PhysRevLett.33.1406. URL: https://link.aps.org/doi/10.1103/PhysRevLett.33.1406.

[3] Thong Q. Nguyen et al. "Topology classification with deep learning to improve real-time event selection at the LHC." In: *Comput. Softw. Big Sci.* 3.1 (2019), p. 12. DOI: 10.1007/s41781-019-0028-1. arXiv: 1807.00083 [hep-ex].

[4] Olmo Cerri et al. "Variational autoencoders for new physics mining at the Large Hadron Collider." In: *J. High Energy Phys.* 2019.5 (2019), p. 36. DOI: 10.1007/JHEP05(2019)036.

[5] Oliver Knapp et al. "Adversarially Learned Anomaly Detection on CMS Open Data: Re-discovering the top quark." In: *Eur. Phys. J. Plus* 136.2 (2021), p. 236. DOI: 10.1140/epjp/s13360-021-01109-4. arXiv: 2005.01598 [hep-ex].

[6] Gaia Grosso. "Machine Learning techniques to search for New Physics." In: (2019). URL: http://cds.cern.ch/record/2653342.

[7] Arjona Martinez, J. et al. "Pileup mitigation at the Large Hadron Collider with graph neural networks." In: *Eur. Phys. J. Plus* 134.7 (2019), p. 333. DOI: 10.1140/epjp/i2019-12710-3. URL: https://doi.org/10.1140/epjp/i2019-12710-3.

[8] Eric A. Moreno et al. "JEDI-net: A jet identification algorithm based on interaction networks." In: *Eur. Phys. J. C* 80.1 (2020), p. 58. DOI: 10.1140/epjc/s10052-020-7608-4. arXiv: 1908.05318 [hep-ex].

[9] Eric A. Moreno et al. "Interaction networks for the identification of boosted $H \to b\bar{b}$ decays." In: *Phys. Rev. D* 102 (1 2020), p. 012010. DOI: 10.1103/PhysRevD.102.012010.

[10] Cheng Chen et al. "Analysis-specific fast simulation at the LHC with deep Learning." In: *Comput. Softw. Big Sci.* 5.1 (2021), p. 15. DOI: 10.1007/s41781-021-00060-4.

[11] Wozniak, Kinga Anna et al. "New physics agnostic selections for new physics searches." In: *EPJ Web Conf.* 245 (2020), p. 06039. DOI: 10.1051/epjconf/202024506039. URL: https://doi.org/10.1051/epjconf/202024506039.

[12] Olmo Cerri et al. "Identification of long-lived charged particles using time-of-flight systems at the upgraded LHC detectors." In: *J. High Energy Phys.* 2019.4 (2019), p. 37. DOI: 10.1007/JHEP04(2019)037.

[13] Adi Bornheim et al. "Geant4 simulation of the light yield and timing performance for BTL sensors." In: *CMS DN-2018/031* (2018).

[14] Adi Bornheim et al. "BTL sensor performance in testbeam campaigns at FNAL and lab source measurements." In: *CMS DN-2018/032* (2018).

[15] Si Xie et al. "Barrel MIP timing detector testbeam setup in 2018 and 2019." In: *CMS DN-2019/042* (2019).

[16] Adi Bornheim et al. "Thermal performance of the barrel timing layer for the CMS phase 2 upgrade." In: *CMS DN-2018/039* (2019).

[17] Olmo Cerri et al. "Thermal studies on BTL prototype modules." In: *CMS DN-2020/013* (2020).

[18] S. L. Glashow. "Partial Symmetries of Weak Interactions". In: *Nucl. Phys.* 22 (1961), pp. 579–588. DOI: 10.1016/0029-5582(61)90469-2.

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

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

[21] Steven Weinberg. "A Model of Leptons". In: *Phys. Rev. Lett.* 19 (1967), pp. 1264–1266. DOI: 10.1103/PhysRevLett.19.1264.

[22] Abdus Salam. "Weak and Electromagnetic Interactions". In: *Conf. Proc. C* 680519 (1968), pp. 367–377. DOI: 10.1142/9789812795915_0034.

[23] Michael Edward Peskin and Daniel V. Schroeder. *An Introduction to Quantum Field Theory*. Reading, USA: Addison-Wesley (1995) 842 p. Westview Press, 1995.

[24] R. Barbieri. *Lectures on the ElectroWeak Interactions*. Publications of the Scuola Normale Superiore. Scuola Normale Superiore, 2007. ISBN: 9788876423116. URL: https://books.google.com/books?id=FuHvAAAAMAAJ.

[25] Benjamin Schulz. "Review on the quantization of gravity". In: (2014). arXiv: 1409.7977 [gr-qc].

[26] M. Srednicki. *Quantum field theory*. Cambridge University Press, Jan. 2007.

[27] Wikipedia. *Standard Model — Wikipedia, The Free Encyclopedia*. http://en.wikipedia.org/w/index.php?title=Standard%20Model&oldid=1095725338. [Online; accessed 17-July-2022]. 2022.

[28] J. A. Minahan, P. Ramond, and R. C. Warner. “Comment on anomaly cancellation in the standard model”. In: *Phys. Rev. D* 41 (2 1990), pp. 715–716. DOI: 10.1103/PhysRevD.41.715. URL: https://link.aps.org/doi/10.1103/PhysRevD.41.715.

[29] Particle Data Group. “Review of Particle Physics”. In: *Progress of Theoretical and Experimental Physics* 2020.8 (2020). 083C01. ISSN: 2050-3911. DOI: 10.1093/ptep/ptaa104. URL: https://doi.org/10.1093/ptep/ptaa104.

[30] Lincoln Wolfenstein. “Parametrization of the Kobayashi-Maskawa Matrix”. In: *Phys. Rev. Lett.* 51 (21 1983), pp. 1945–1947. DOI: 10.1103/PhysRevLett.51.1945. URL: https://link.aps.org/doi/10.1103/PhysRevLett.51.1945.

[31] Lyndon Evans and Philip Bryant. “LHC Machine”. In: *Journal of Instrumentation* 3.08 (2008), S08001–S08001. DOI: 10.1088/1748-0221/3/08/s08001. URL: https://doi.org/10.1088/1748-0221/3/08/s08001.

[32] S. Chatrchyan et al. “The CMS Experiment at the CERN LHC”. In: *JINST* 3 (2008), S08004. DOI: 10.1088/1748-0221/3/08/S08004.

[33] Stephen Holmes, Ronald S Moore, and Vladimir Shiltsev. “Overview of the Tevatron collider complex: goals, operations and performance”. In: *Journal of Instrumentation* 6.08 (2011), T08001–T08001. DOI: 10.1088/1748-0221/6/08/t08001. URL: https://doi.org/10.1088/1748-0221/6/08/t08001.

[34] Ralph Assmann, Mike Lamont, and Steve Myers. “A brief history of the LEP collider”. In: *Nuclear Physics B - Proceedings Supplements* 109.2 (2002). Proceedings of the 7th Topical Seminar, pp. 17–31. ISSN: 0920-5632. DOI: https://doi.org/10.1016/S0920-5632(02)90005-8. URL: https://www.sciencedirect.com/science/article/pii/S0920563202900058.

[35] CERN. *TE-EPC-LPC in LHC*. http://te-epc-lpc.web.cern.ch/te-epc-lpc/machines/lhc/general.stm. 2017.

[36] The ATLAS Collaboration. “The ATLAS Experiment at the CERN Large Hadron Collider”. In: *Journal of Instrumentation* 3.08 (2008), S08003. URL: http://stacks.iop.org/1748-0221/3/i=08/a=S08003.

[37] The LHCb Collaboration. “The LHCb Detector at the LHC”. In: *Journal of Instrumentation* 3.08 (2008), S08005. URL: http://stacks.iop.org/1748-0221/3/i=08/a=S08005.

[38] The ALICE Collaboration. “The ALICE experiment at the CERN LHC”. In: *Journal of Instrumentation* 3.08 (2008), S08002. URL: http://stacks.iop.org/1748-0221/3/i=08/a=S08002.

[39] The CMS collaboration. *CMS Luminosity - Public results*. https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults. 2022.

[40] O. Aberle et al. *High-Luminosity Large Hadron Collider (HL-LHC): Technical design report*. CERN Yellow Reports: Monographs. Geneva: CERN, 2020. DOI: 10.23731/CYRM-2020-0010. URL: https://cds.cern.ch/record/2749422.

[41] Siona Ruth Davis. "Interactive Slice of the CMS detector". In: (2016). URL: https://cds.cern.ch/record/2205172.

[42] Wolfgang Adam et al. *Track Reconstruction in the CMS tracker*. Tech. rep. Geneva: CERN, 2006. URL: https://cds.cern.ch/record/934067.

[43] Serguei Chatrchyan et al. "Description and performance of track and primary-vertex reconstruction with the CMS tracker". In: *JINST* 9.10 (2014), P10009. DOI: 10.1088/1748-0221/9/10/P10009. arXiv: 1405.6569 [physics.ins-det].

[44] A. M. Sirunyan et al. "Particle-flow reconstruction and global event description with the CMS detector". In: *JINST* 12.10 (2017), P10003. DOI: 10.1088/1748-0221/12/10/P10003. arXiv: 1706.04965 [physics.ins-det].

[45] The CMS collaboration. *The CMS Software components*. http://cms-sw.github.io/. 2022.

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

[47] Torbjörn Sjöstrand et al. "An introduction to PYTHIA 8.2". In: *Comput. Phys. Commun.* 191 (2015), pp. 159–177. DOI: 10.1016/j.cpc.2015.01.024. arXiv: 1410.3012 [hep-ph].

[48] Anders Ryd et al. "EvtGen: A Monte Carlo Generator for B-Physics". In: (2005).

[49] S. Agostinelli and J. Allison et al. "Geant4—a simulation toolkit". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 506.3 (2003), pp. 250–303. ISSN: 0168-9002. DOI: https://doi.org/10.1016/S0168-9002(03)01368-8. URL: https://www.sciencedirect.com/science/article/pii/S0168900203013688.

[50] J. Allison and K. Amako et al. "Recent developments in Geant4". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 835 (2016), pp. 186–225. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2016.06.125.

*Chapter 2*

# DEVELOPMENT OF MACHINE LEARNING BASED PHYSICS TOOLS

> *“Essendo la Natura inesorabile ed immutabile, e mai non trascendente i termini delle leggi impostegli, come quella che nulla cura che le sue recondite ragioni e modi d’operare sieno o non sieno esposti alla capacità degli uomini."*
> Nature is relentless and unchangeable and it is indifferent as to whether its hidden reasons and actions are understandable to humans or not.
>
> — Galileo Galilei, *Il Saggiatore.*

The great technological development in computer hardware of the last decades, guided by the self-fulfilling Moore’s law [1], has drastically increased the computing power available to human endeavors. One of the fields that is benefiting greatly from this development is machine learning (ML) and in particular its application to complex and multi-variate problems. In particular, Deep Learning represents the cutting-edge ML technology which was built from the field of neural networks, initially developed since the 1960s [2], and the newly augmented computational capabilities driven for example by the use of GPUs. The success of deep learning in a wide variety of applications is permeating also in particle physics which is evolving its application far beyond the BDTs popular in the early 2000s. The challenge faced in developing ML techniques for particle physics research goes beyond the already complex task of training the model but includes also studies in the data pipeline, model development, and inference application.

Believing that a new breakthrough in fundamental physics will also be supported by the development of new emerging experimental techniques, I first approached the application of ML to experimental particle physics during my master thesis [3]. During my PhD, I first joined an ongoing effort to develop a topology classification with Deep learning to improve real-time event selection [4]. In that work, we show how an event topology classification based on deep learning could be used to improve the purity of data samples selected in real-time at the LHC. We tested different data representations, on which different kinds of multi-class classifiers are trained. Both raw data and high-level features are utilized. In the considered examples, a filter based on the classifier's score can be trained to retain 99% of the interesting events and reduce the false-positive rate by more than one order of magnitude. By operating such a filter as part of the online event selection infrastructure of the LHC experiments, one could benefit from a more flexible and inclusive selection strategy while reducing the amount of downstream resources wasted in processing false positives. The saved resources could translate into a reduction of the detector operation cost or into an effective increase of storage and processing capabilities, which could be reinvested to extend the physics reach of the LHC experiments. After this first work, I focused on a study to develop a model-independent tagger for unexpected events using artificial neural networks for physics anomaly detection [5] that is extensively discussed in Sec. 2.1. As a natural continuation, I collaborated on the first application of this proposed strategy on real data that showcased the feasibility of applying that method to real LHC data. The rediscovery of the top quark production at LHC [6] has been used as proof of the method's effectiveness and its independence from a reliable Monte Carlo simulation. In addition, I collaborated with or supervised younger students in several other ML research projects as summarized in Sec. 2.2.

## 2.1 Variational Autoencoders for New Physics Mining at the LHC

Using variational autoencoders [7] trained on known physics processes, we develop a one-sided threshold test to isolate previously unseen processes as outlier events. Since the autoencoder training does not depend on any specific new physics signature, the proposed procedure does not make specific assumptions about the nature of new physics. An event selection based on this algorithm would be complementary to classic LHC searches, typically based on model-dependent hypothesis testing. Such an algorithm would deliver a list of anomalous events, that the experimental collaborations could further scrutinize and even release as a catalog, similar to what is typically done in other scientific domains. Event topologies repeating in this dataset could inspire new-physics model building and new experimental searches.

Running in the trigger system of the LHC experiments, such an application could identify anomalous events that would be otherwise lost, extending the scientific reach of the LHC. The code developed for this project can be found in the repository in [8].

**Introduction**

One of the main motivations behind the construction of the CERN Large Hadron Collider (LHC) is the exploration of the high-energy frontier in search for new physics phenomena. New physics could answer some of the standing fundamental questions in particle physics, e.g., the nature of dark matter or the origin of electroweak symmetry breaking. In LHC experiments, searches for physics beyond the Standard Model (BSM) are typically carried on as fully-supervised data analyses: assuming a new physics scenario of some kind, a search is structured as a hypothesis test, based on a profiled-likelihood ratio [9]. These searches are said to be *model dependent*, since they depend on considering a specific new physics model.

Assuming that one is testing the *right* model, this approach is very effective in discovering a signal, as demonstrated by the discovery of the Standard Model (SM) Higgs boson [10, 11] at the LHC. On the other hand, given the (so far) negative outcome of many BSM searches at particle-physics experiments, it is possible that a future BSM model, if any, is not among those typically tested. The problem is more profound if analyzed in the context of the LHC big-data problem: at the LHC, 40 million proton-beam collisions are produced every second, but only $\sim$1000 collision events/sec can be stored by the ATLAS and CMS experiments, due to limited bandwidth, processing, and storage resources. It is possible to imagine BSM scenarios that would escape detection, simply because the corresponding new physics events would be rejected by a typical set of online selection algorithms. Establishing alternative search methodologies with reduced model dependence is an important aspect of future LHC runs. Traditionally, this issue was addressed with so-called model-independent searches, performed at the Tevatron [12, 13], at HERA [14], and at the LHC [15, 16], as discussed in the first subsection about related works at page 39.

In this work, we propose to address this need by deploying an unsupervised algorithm in the online selection system (trigger) of the LHC experiments.[1] This algorithm

[1] A description of the ATLAS and CMS trigger systems can be found in Ref. [17] and Ref. [18], respectively. In this study, we take the data-taking strategy of these two experiments as a reference. On the other hand, the proposed strategy could be adapted to other use cases.

would be trained on known SM processes and would be able to identify BSM events as anomalies. The selected events could be stored in a special stream, scrutinized by experts (e.g., to exclude the occurrence of detector malfunctions that could explain the anomalies), and even released outside the experimental collaborations, in the form of an open-access catalog. The final goal of this application is to identify anomalous event topologies and inspire future supervised searches on data collected afterwards.

As an example, we consider the case of a typical single-lepton data stream, selected by a hardware-based Level-1 (L1) trigger system. In normal conditions, the L1 trigger is the first of a two-stage selection. After a coarse (and often local) reconstruction and loose selection at L1, events are fully reconstructed in the High Lever Trigger (HLT), where a much tighter selection is applied. The selection is usually done having in mind specific signal topologies, eg., specific BSM models. In this study, we imagine to replace this model-dependent selection with a variational autoencoder (VAE) [19, 20] looking for anomalous events in the incoming single-lepton stream. The VAE is trained to compress the input event representation into a lower-dimension latent space and then decompress it, returning the shape parameters describing the probability density function (pdf) of each input quantity given a point in the compressed space. In addition, a VAE allows a stochastic modeling of the latent space, a feature which is missing in a simple AE architecture. The highlighted procedure is not specific of the considered single-lepton stream and could be easily extended to other data streams.

The distribution of the VAE's reconstruction loss on a validation sample is used to define a threshold, corresponding to a desired acceptance rate for SM events. All the events with loss larger than the threshold are considered as potential anomalies and could be stored in a low-rate anomalous-event data stream. In this work, we set the threshold such that $\sim$ 1000 SM events would be collected every month under typical LHC operation conditions. In particular, we took as a reference 8 months of data taking per year, with an integrated luminosity of $\sim 40\ \text{fb}^{-1}$. Assuming an LHC duty cycle of 2/3, this corresponds to an average instantaneous luminosity of $\sim 2.9 \times 10^{33}\ \text{cm}^{-2}\ \text{s}^{-1}$.

We then evaluate the BSM production cross section that would correspond to a signal excess of 100 BSM events selected per month, as well as the one that would give a signal yield $\sim 1/3$ of the SM yield.

For this, we consider a set of low-mass BSM resonances, decaying to one or more leptons and light enough to be challenging for the currently employed LHC trigger algorithms.

**Related Work**

Model-independent searches for new physics have been performed at the Tevatron [12, 13], HERA [14], and the LHC [15, 16]. These searches are based on the comparison of a large set of binned distributions to the prediction from Monte Carlo (MC) simulations, in search for bins exhibiting a deviation larger than some predefined threshold. While the effectiveness of this strategy in establishing a discovery has been a matter of discussion, a recent study by the ATLAS collaboration [16] rephrased this model-independent search strategy into a tool to identify interesting excesses, on which traditional analysis techniques could be performed on independent datasets (e.g., the data collected after running the model-independent analysis). This change of scope has the advantage of reducing the trial factor (i.e., the so-called *look-elsewhere* effect [21, 22]), which would otherwise wash out the significance of an observed excess.

Our strategy is similar to what is proposed in Ref. [16], with two substantial differences: (i) we aim to process also those events that could be discarded by the online selection, by running the algorithm as part of the trigger process; (ii) we do so by exploiting deep-learning-based anomaly detection techniques.

Applying deep learning at the trigger level has been proposed in Ref. [23]. Other works [24, 25, 26, 27] have investigated the use of machine-learning techniques to setup new strategies for BSM searches with minimal or no assumption on the specific new-physics scenario under investigation. In this work, we use VAEs [19, 20] based on high-level features as a baseline. Previously, autoencoders have been used in collider physics for detector monitoring [28, 29] and event generation [30]. Autoencoders have also been explored to define a jet tagger that would identify new physics events with anomalous jets [31, 32], with a strategy similar to what we apply to the full event in this work.

Anomaly detection has been a traditional use case for one-class machine learning methods, such as one-class Support Vector Machine [33] or Isolation Forest [34, 35]. A review of proposed methods can be found in Ref. [36]. Variational methods have been shown to be effective for novelty detection, as for instance is discussed in Ref. [37]. In particular, VAEs [19] have been proposed as an effective method for anomaly detection [20].

### Data Samples

The dataset used for this study is a refined version of the high-level-feature (HLF) dataset used in Ref. [23]. Proton-proton collisions are generated using the `PYTHIA8` event-generation library [38], fixing the center-of-mass energy to the LHC Run-II value (13 TeV) and the average number of overlapping collisions per beam crossing (pileup) to 20. These beam conditions loosely correspond to the LHC operating conditions in 2016.

Events generated by `PYTHIA8` are processed with the `DELPHES` library [39], to emulate detector efficiency and resolution effects. We take as a benchmark detector description the upgraded design of the CMS detector, foreseen for the High-Luminosity LHC phase [40]. In particular, we use the CMS HL-LHC detector card distributed with `DELPHES`. We run the `DELPHES` *particle-flow* (PF) algorithm, which combines the information from different detector components to derive a list of reconstructed particles, the so-called PF candidates. For each particle, the algorithm returns the measured energy and flight direction. Each particle is associated to one of three classes: charged particles, photons, and neutral hadrons. In addition, lists of reconstructed electrons and muons are given.

Many SM processes would contribute to the considered single-lepton dataset. For simplicity, we restrict the list of relevant SM processes to the four with the highest production cross sections, namely:

- Inclusive $W$ production, with $W \to \ell\nu$ ($\ell = e,\ \mu,\ \tau$).
- Inclusive $Z$ production, with $Z \to \ell\ell$ ($\ell = e,\ \mu,\ \tau$).
- $t\bar{t}$ production.
- QCD multijet production.[2]

These samples are mixed to provide a SM cocktail dataset, which is then used to train autoencoder models and to tune the threshold requirement that defines what we consider an anomaly. The cocktail is built scaling down the high-statistics samples

[2] To speed up the generation process for QCD events, we require $\sqrt{\hat{s}} > 10$ GeV, the fraction of QCD events with $\sqrt{\hat{s}} < 10$ GeV and producing a lepton within acceptance being negligible but computationally expensive.

($t\bar{t}$, $W$, and $Z$) to the lowest-statistics one (QCD, whose generation is the most computing-expensive), according to their production cross-section values (estimated at leading order with PYTHIA) and selection efficiencies, shown in Table 2.1.

Table 2.1: Acceptance and L1 trigger (i.e. $p_T^\ell$ and Iso requirement) efficiency for the four studied SM processes and corresponding values for the BSM benchmark models. For SM processes, we quote the total cross section before the trigger, the expected number of events per month and the fraction in the SM cocktail. For BSM models, we compute the production cross section corresponding to an average of 100 BSM events per month passing the acceptance and L1 trigger requirements. The monthly event yield is computed assuming an average luminosity per month of 5 fb$^{-1}$, corresponding to the running conditions discussed in Section 2.1.

| Standard Model processes | | | | | |
|---|---|---|---|---|---|
| Process | Acceptance | L1 trigger efficiency | Cross section [nb] | Event fraction | Events /month |
| $W$ | 55.6% | 68% | 58 | 59.2% | 110M |
| QCD | 0.08% | 9.6% | $1.6 \cdot 10^5$ | 33.8% | 63M |
| $Z$ | 16% | 77% | 20 | 6.7% | 12M |
| $t\bar{t}$ | 37% | 49% | 0.7 | 0.3% | 0.6M |

| BSM benchmark processes | | | | |
|---|---|---|---|---|
| Process | Acceptance | L1 trigger efficiency | Total efficiency | Cross-section 100 BSM events/month |
| $A \to 4\ell$ | 5% | 98% | 5% | 0.44 pb |
| $LQ \to b\tau$ | 19% | 62% | 12% | 0.17 pb |
| $h^0 \to \tau\tau$ | 9% | 70% | 6% | 0.34 pb |
| $h^\pm \to \tau\nu$ | 18% | 69% | 12% | 0.16 pb |

Events are filtered at generation requiring an electron, muon, or tau lepton with $p_T > 22$ GeV. Once detector effects are taken into account through the DELPHES simulation, events are further selected requiring the presence of one reconstructed lepton (electron or muon) with transverse momentum $p_T > 23$ GeV and a loose isolation requirement Iso $< 0.45$. If more than one reconstructed lepton is present, the highest $p_T$ one is considered. The isolation for the considered lepton $\ell$ is computed as:

$$\mathrm{Iso} = \frac{\sum_{p \neq \ell} p_T^p}{p_T^\ell}\,, \tag{2.1}$$

where the index $p$ runs over all the photons, charged particles, and neutral hadrons within a cone of size $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} < 0.3$ from $\ell$.[3]

[3] As common for collider physics, we use a Cartesian coordinate system with the $z$ axis oriented along the beam axis, the $x$ axis on the horizontal plane, and the $y$ axis oriented upward. The $x$ and $y$

The 21 considered HLF quantities are:

- The absolute value of the isolated-lepton transverse momentum $p_T^\ell$.
- The three isolation quantities (ChPFIso, NeuPFIso, GammaPFIso) for the isolated lepton, computed with respect to charged particles, neutral hadrons and photons, respectively.
- The lepton charge.
- A Boolean flag (isEle) set to 1 when the trigger lepton is an electron, 0 otherwise.
- $S_T$, i.e. the scalar sum of the $p_T$ of all the jets, leptons, and photons in the event with $p_T > 30$ GeV and $|\eta| < 2.6$. Jets are clustered from the reconstructed PF candidates, using the `FASTJET` [41] implementation of the anti-$k_T$ jet algorithm [42], with a jet-size parameter R=0.4.
- The number of jets entering the $S_T$ sum ($N_J$).
- The invariant mass of the set of jets entering the $S_T$ sum ($M_J$).
- The number of these jets being identified as originating from a $b$ quark ($N_b$).
- The missing transverse momentum, decomposed into its parallel ($p_{T,\parallel}^{\text{miss}}$) and orthogonal ($p_{T,\perp}^{\text{miss}}$) components with respect to the lepton $\ell$ direction. The missing transverse momentum is defined as the negative sum of the PF-candidate $p_T$ vectors:
$$\vec{p}_T^{\,\text{miss}} = -\sum_q \vec{p}_T^{\,q} \; . \tag{2.2}$$
- The transverse mass, $M_T$, of the isolated lepton $\ell$ and the $\vec{p}_T^{\,\text{miss}}$ system, defined as:
$$M_T = \sqrt{2 p_T^\ell E_T^{\text{miss}} (1 - \cos \Delta\phi)} \, , \tag{2.3}$$
with $\Delta\phi$ the azimuth separation between the $\vec{p}_T^{\,\ell}$ and $\vec{p}_T^{\,\text{miss}}$ vectors, and $E_T^{\text{miss}}$ the magnitude of $\vec{p}_T^{\,\text{miss}}$.
- The number of selected muons ($N_\mu$).

---

axes define the transverse plane, while the $z$ axis identifies the longitudinal direction. The azimuth angle $\phi$ is computed from the $x$ axis. The polar angle $\theta$ is used to compute the pseudorapidity $\eta = -\log(\tan(\theta/2))$. We fix units such that $c = \hbar = 1$.

- The invariant mass of this set of muons ($M_\mu$).
- The absolute value of the total transverse momentum of these muons ($p^\mu_{T,TOT}$).
- The number of selected electrons ($N_e$).
- The invariant mass of this set of electrons ($M_e$).
- The absolute value of the total transverse momentum of these electrons ($p^e_{T,TOT}$).
- The number of reconstructed charged hadrons.
- The number of reconstructed neutral hadrons.

This list of HLF quantities is not defined having in mind a specific BSM scenario. Instead, it is conceived to include relevant information to discriminate the various SM processes populating the single-lepton data stream. On the other hand, it is generic enough to allow (at least in principle) the identification of a large set of new physics scenarios.

In addition to the four SM processes listed above, we consider the following BSM models to benchmark anomaly-detection capabilities:

- A leptoquark $LQ$ with mass 80 GeV, decaying to a $b$ quark and a $\tau$ lepton.
- A neutral scalar boson with mass 50 GeV, decaying to two off-shell $Z$ bosons, each forced to decay to two leptons: $A \to 4\ell$.
- A scalar boson with mass 60 GeV, decaying to two tau leptons: $h^0 \to \tau\tau$.
- A charged scalar boson with mass 60 GeV, decaying to a tau lepton and a neutrino: $h^\pm \to \tau\nu$.

For each BSM scenario, we consider any direct production mechanism implemented in `PYTHIA8`, including associate jet production. We list in Table 2.1 the leading-order production cross section and selection efficiency for each model.

Figures 2.1 and 2.2 show the distribution of HLF quantities for the SM processes and the BSM benchmark models, respectively.

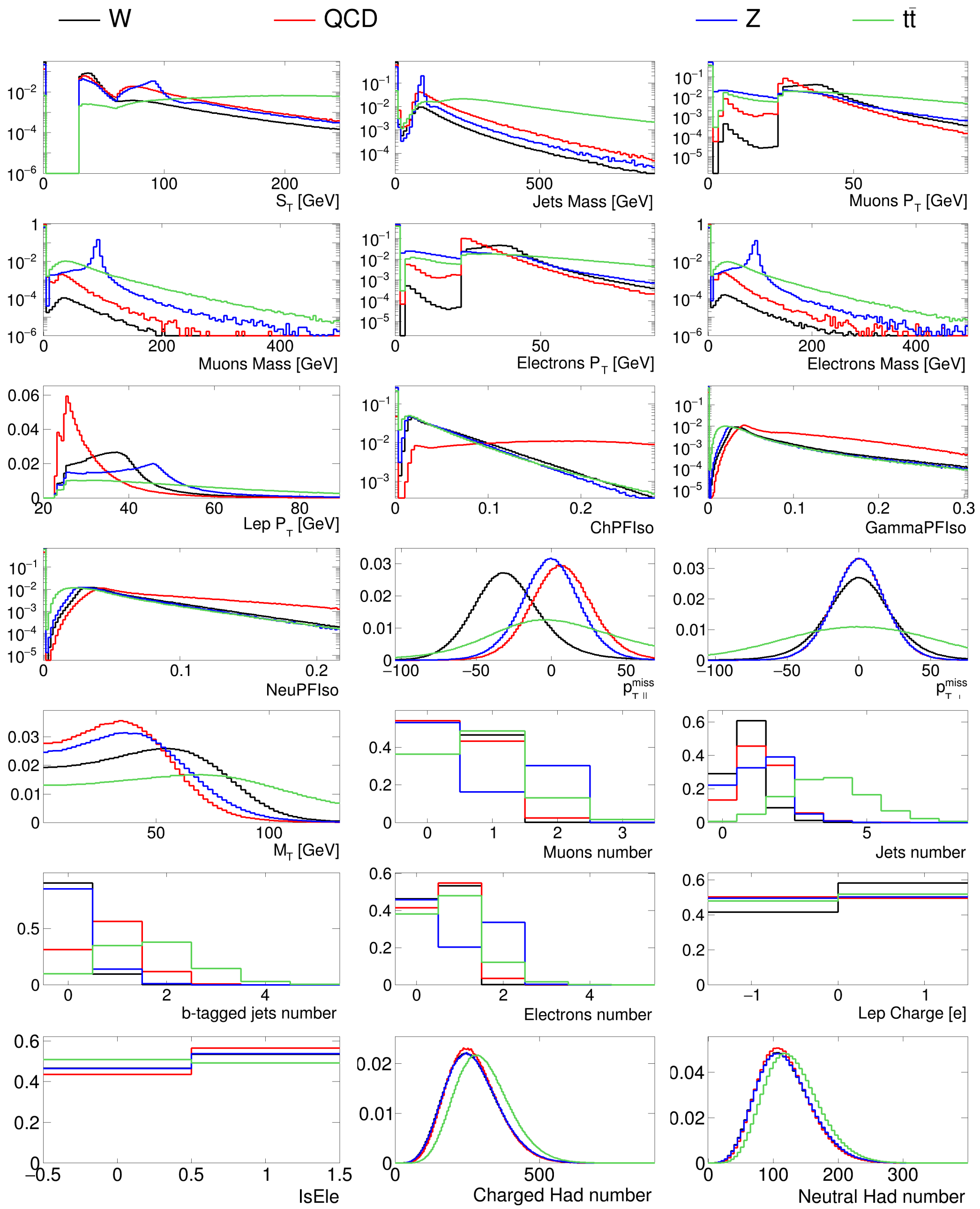


Figure 2.1: Distribution of the HLF quantities for the four considered SM processes.

### Model Description

We train VAEs on the SM cocktail sample described in Section 2.1, taking as input the 21 HLF quantities listed there. The use of HLF quantities to represent events limits the model independence of the anomaly detection procedure. While the list of features is chosen to represent the main physics aspects of the considered SM processes and is in no way tailored to specific BSM models, it is true that such a

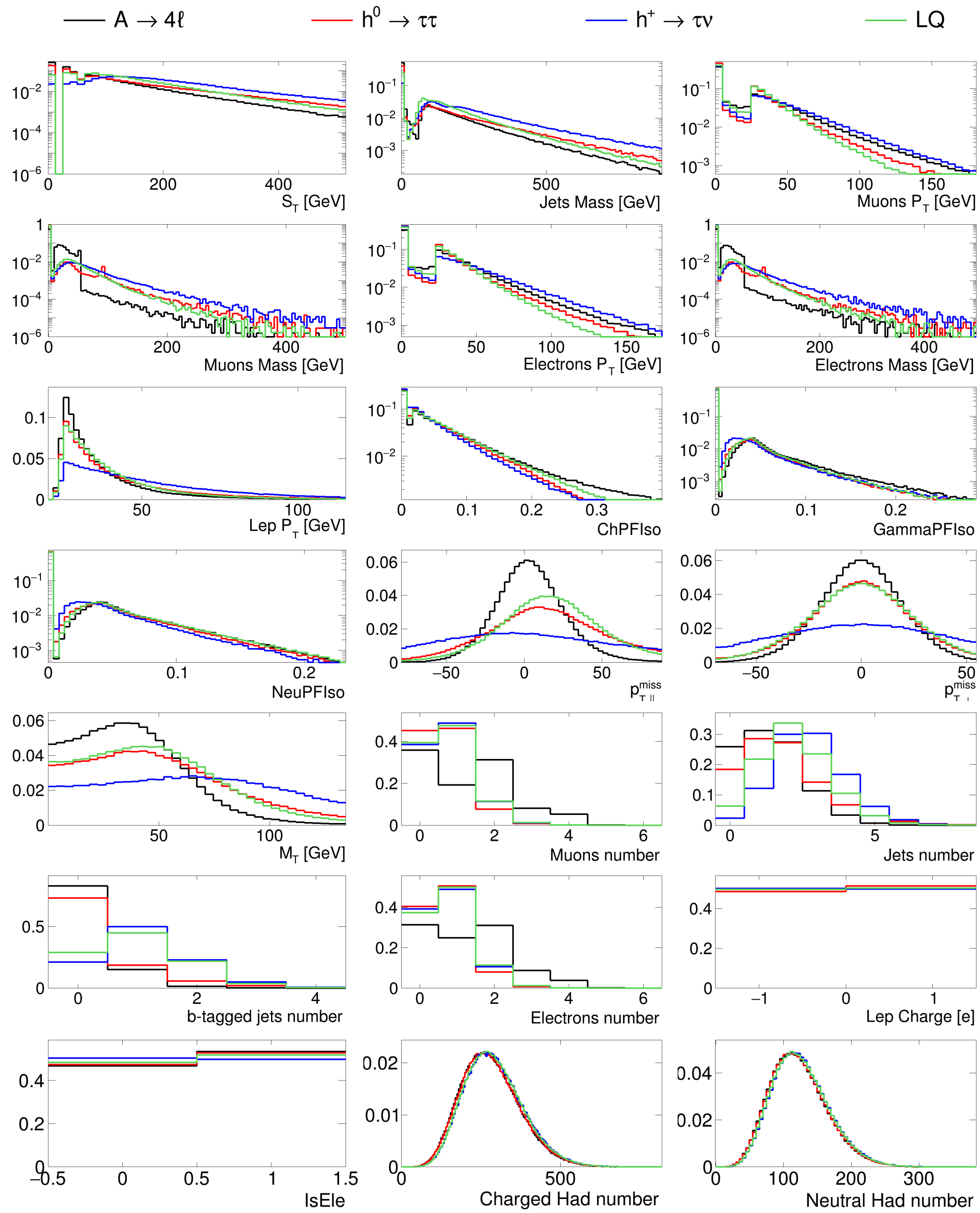


Figure 2.2: Distribution of the HLF quantities for the four considered BSM benchmark models.

list might be more suitable for certain models than for others. In this respect, one cannot guarantee that the anomaly-detection performance observed on a given BSM model would generalize to any BSM scenario. We will address in a future work a possible solution to reduce the residual model dependence implied by the input event representation.

In this section, we present both the best-performing autoencoder model, trained to encode and decode the SM training sample, and a set of four supervised classifiers, each trained to distinguish one of the four BSM benchmark models from SM events. We use the classification performance of these supervised algorithms as an estimate of the best performance that the VAE could get to.

**Autoencoders**

Autoencoders [7] are algorithms that compress a given set of inputs variables in a latent space (encoding) and then, starting from the latent space, reconstruct the HLF input values (decoding). The distribution of the loss function for a reference set of an AE is used in the context of anomaly detection to isolate potential anomalies. Since the compression capability learned on a given sample does not typically generalize to other samples, the tails of the loss distribution could be enriched by new kinds of events, different than those used to train the model. In the specific case considered in this study, the tail of the loss distribution for an AE trained on SM data might be enriched with BSM events.

In this work, we focus on VAEs [19]. For each event, a plain AE predicts an encoded point in the latent space and a decoded point in the original space. In other words, AEs are point-estimate algorithms. VAEs, instead, associate to each input event an estimated probability distribution in the latent space and in the original space. Doing so, VAEs provide both a best-point estimate and an estimate of the associated statistical noise. Besides this conceptual difference, VAEs have been shown to provide competitive performances for novelty [37] and anomaly [20] detection.

We consider the VAE architecture shown in Fig. 2.3, characterized by a four-dimensional latent space. Each latent dimension is associated to a Gaussian pdf and its two degrees of freedom (mean $\mu_z$ and RMS $\sigma_z$). The input layer consists of 21 nodes, corresponding to the 21 HLF quantities described in Section 2.1. This layer is connected to the latent space through a stack of two fully connected layers, each consisting of 50 nodes with ReLU activation functions. Two four-node layers are fully connected to the second 50-node layer. Linear activation functions are used for the first of these four-node layers, interpreted as the set of four $\mu_z$ of the four-dimension Gaussian pdf $p(z)$. The nodes of the second layer are activated by the functions:

$$\text{p-ISRLu}(x) = 1 + 5 \cdot 10^{-3} + \Theta(x)x + \Theta(-x)\frac{x}{\sqrt{1+x^2}} \,. \tag{2.4}$$

This activation allows to improve the training stability, being strictly positive defined,

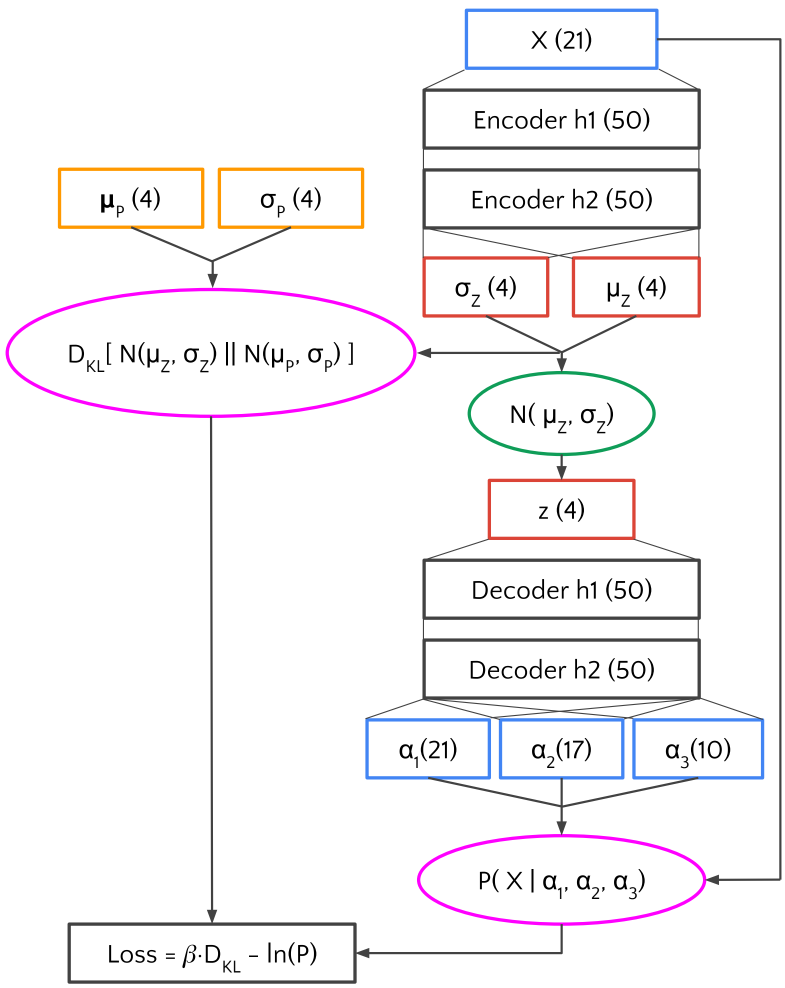


Figure 2.3: Schematic representation of the VAE architecture presented in the text. The size of each layer is indicated by the value within brackets. The blue rectangle $X$ represents the input layer, which is connected to a stack of two consecutive fully connected layers (black boxes). The last of the two black box is connected to two layers with four nodes each (red boxes), representing the $\mu_z$ and $\sigma_z$ parameters of the encoder pdf $p(z|x)$. The green oval represents the sampling operator, which returns a set of values for the 4-dimensional latent variables $z$. These values are fed into the decoder, consisting of two consecutive hidden layers of 50 nodes each (black boxes). The last of the decoder hidden layer is connected to the three output layers, whose nodes correspond to the parameters of the predicted distribution in the initial 21-dimension space. The pink ovals represent the computation of the two parts of the loss function: the KL loss and the reconstruction loss (see text). The computation of the KL requires 8 additional learnable parameters ($\mu_p$ and $\sigma_p$, represented by the orange boxes on the top-left part of the figure), corresponding to the means and RMS of the four-dimensional Gaussian prior $p(z)$. The total loss in computed as described by the formula in the bottom-left black box (see Eq. (2.6)).

non linear, and with no exponentially growing term (which might have created instabilities in the early epochs of the training). The four nodes of this layer are interpreted as the $\sigma_z$ parameters of $p(z)$. After several trials, the dimension of the latent space has been optimized to 4 in order to keep a good training stability without impacting the VAE performances. The decoding step originates from a point in the latent space, randomly sampled according to the predicted pdf (green oval in Fig. 2.3). The coordinates of this point in the latent space are fed into a sequence of two hidden dense layers, each consisting of 50 neurons with ReLU activation functions. The last of these layers is connected to three dense layers of 21, 17, and 10 neurons, activated by linear, p-ISRLu and clipped-tanh functions, respectively. The clipped-tanh function is written as:

$$C_{\tanh}(x) = \frac{1}{2}(1 + 0.999 \cdot \tanh x) \ . \tag{2.5}$$

Given the latent-space representation, the 48 output nodes represent the parameters of the pdfs describing the input HLF probability, i.e., the $\alpha$ parameters of Eq.(2.8).

The total VAE loss function $\mathrm{Loss}_{\mathrm{Tot}}$ is a weighted sum of two pieces [43]: a term related to the reconstruction likelihood ($\mathrm{Loss}_{\mathrm{reco}}$) and the Kullback-Leibler divergence ($D_{\mathrm{KL}}$) between the latent space pdf and the prior:

$$\mathrm{Loss}_{\mathrm{Tot}} = \mathrm{Loss}_{\mathrm{reco}} + \beta D_{\mathrm{KL}} \ , \tag{2.6}$$

where $\beta$ is a free parameter. We fix $\beta = 0.3$, for which we obtained good reconstruction performances.[4] The prior $p(z)$ chosen for the latent space is a four-dimension Gaussian with a diagonal covariance matrix. The means ($\mu_P$) and the diagonal terms of the covariance matrix ($\sigma_P$) are free parameters of the algorithm and are optimized during the back-propagation. The Kullback-Leibler divergence between two Gaussian distributions has an analytic form. Hence, for each batch, $D_{\mathrm{KL}}$ can be expressed as:

$$\begin{aligned} D_{\mathrm{KL}} &= \frac{1}{k}\sum_i D_{\mathrm{KL}}\left(N(\mu_z^i, \sigma_z^i) \ \ || \ \ N(\mu_P, \sigma_P)\right) \\ &= \frac{1}{2k}\sum_{i,j}\left(\sigma_P^j \sigma_z^{i,j}\right)^2 + \left(\frac{\mu_P^j - \mu_z^{i,j}}{\sigma_P^j}\right)^2 + \ln\frac{\sigma_P^j}{\sigma_z^{i,j}} - 1 \ , \end{aligned} \tag{2.7}$$

where $k$ is the batch size, $i$ runs over the samples and $j$ over the latent space dimensions. Similarly, $\mathrm{Loss}_{\mathrm{reco}}$ is the average negative-log-likelihood of the inputs

[4] Following Ref. [43], we tried to increase the value of $\beta$ up to 4 without observing a substantial difference in performance.

given the predicted $\alpha$ values:

$$\begin{aligned} \text{Loss}_{\text{reco}} &= -\frac{1}{k}\sum_{i} \ln\left[P(x \mid \alpha_1, \alpha_2, \alpha_3)\right] \\ &= -\frac{1}{k}\sum_{i,j} \ln\left[f_j(x_{i,j} \mid \alpha_1^{i,j}, \alpha_2^{i,j}, \alpha_3^{i,j})\right] . \end{aligned} \tag{2.8}$$

In the equation, $j$ runs over the input space dimensions, $f_j$ is the functional form chose to describe the pdf of the $j$-th input variable and $\alpha_m^{i,j}$ are the parameters of the function. Different functional forms have been chosen for $f_j$, to properly describe different classes of HLF distributions:

- **Clipped Log-normal + $\delta$ function**: used to describe $S_T$, $M_J$, $p_T^{\mu}$, $M_{\mu}$, $p_T^{e}$, $M_e$, $p_T^{\ell}$, ChPFIso, NeuPFIso and GammaPFIso:

$$P(x \mid \alpha_1, \alpha_2, \alpha_3) = \begin{cases} \alpha_3\delta(x) + \frac{1-\alpha_3}{x\alpha_2\sqrt{2\pi}} \exp\left(-\frac{(\ln x - \alpha_1)^2}{2\alpha_2^2}\right) \; for\ x \geq 10^{-4} \\ 0 \; for\ x < 10^{-4} \end{cases} . \tag{2.9}$$

- **Gaussian**: used for $p_{T,\parallel}^{\text{miss}}$ and $p_{T,\perp}^{\text{miss}}$:

$$P(x \mid \alpha_1, \alpha_2) = \frac{1}{\alpha_2\sqrt{2\pi}} \exp\left(-\frac{(x-\alpha_1)^2}{2\alpha_2^2}\right) . \tag{2.10}$$

- **Truncated Gaussian**: a Gaussian function truncated for negative values and normalized to unit area for $X > 0$. Used to model $M_T$:

$$P(x \mid \alpha_1, \alpha_2) = \Theta(x) \cdot \frac{1 + 0.5 \cdot (1 + \text{erf}\frac{-\alpha_1}{\alpha_2\sqrt{2}})}{\alpha_2\sqrt{2\pi}} \exp\left(-\frac{(x-\alpha_1)^2}{2\alpha_2^2}\right) . \tag{2.11}$$

- **Discrete truncated Gaussian**: like the truncated Gaussian, but normalized to be evaluated on integers (i.e. $\sum_{n=0}^{\infty} P(n) = 1$). This function is used to describe $N_{\mu}$, $N_e$, $N_b$ and $N_J$. It is written as:

$$P(n \mid \alpha_1, \alpha_2) = \Theta(x) \left[\text{erf}\left(\frac{n + 0.5 - \alpha_1}{\alpha_2\sqrt{2}}\right) - \text{erf}\left(\frac{n - 0.5 - \alpha_1}{\alpha_2\sqrt{2}}\right)\right] \mathcal{N} , \tag{2.12}$$

  where the normalization factor $\mathcal{N}$ is set to:

$$\mathcal{N} = 1 + \frac{1}{2}\left(1 + \text{erf}\left(\frac{-0.5 - \alpha_1}{\alpha_2\sqrt{2}}\right)\right) . \tag{2.13}$$

- **Binomial**: used for (isEle) and lepton charge:

$$P(n \mid p) = \delta_{n,m} p + \delta_{n,l}(1-p) \; , \tag{2.14}$$

where $m$ and $l$ are the two possible values of the variable (0 or 1 for (isEle) and -1 or 1 for lepton charge) and $p = C_{\tanh}(\alpha_1)$.

- **Poisson**: used for charged-particle and neutral-hadron multiplicities:

$$P(n \mid \mu) = \frac{\mu^n e^{-\mu}}{\Gamma(n+1)} \; , \tag{2.15}$$

where $\mu = \text{p-ISRLu}(\alpha_1)$.

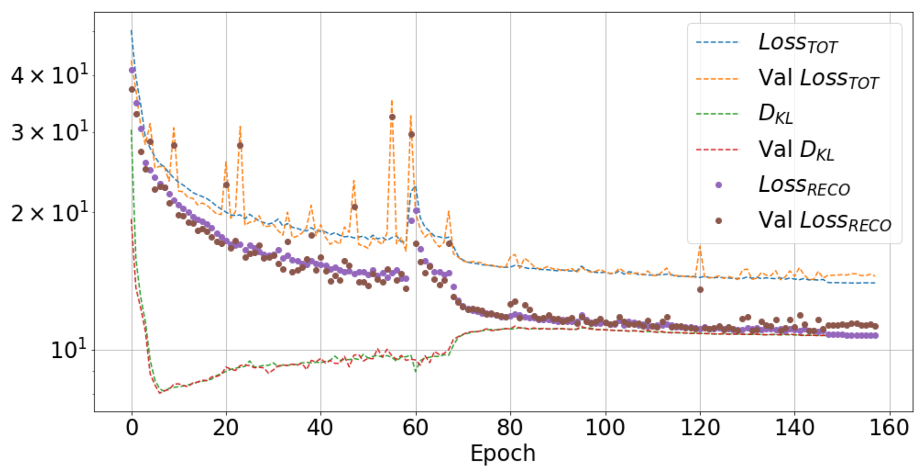


Figure 2.4: Training history for VAE. Total loss, reconstruction negative-log-likelihood ($\text{Loss}_{\text{reoc}}$) and KL divergence ($D_{KL}$) are shown separately for training and validation (indicated in thelegend by the Val prefix) sets though all the training epochs.

These custom functions provide an improved performance with respect to the standard choice of an MSE loss. When using the MSE loss, one is implicitly writing the likelihood of the input quantities as a product of Gaussian functions with equal variance. This choice is clearly a poor description of the input distributions at hand in this application and it results in a poor representation of the cores and the tails of the input distributions. Instead, the use of these tailored functions allows to correctly describe the distribution cores and to improve the description of the tails.

We point out that the final performance depends on the choice of the p(x|z) functional form (i.e., on the modeled dependence of the observed features on the latent variables) and the p(z) prior function. The former was tuned looking at the distributions for SM events. The latter is arbitrary. We explored techniques to optimize the choice of p(z), learning it from the data [44]. In this case, no practical advantage in terms of anomaly detection was observed. An improved choice of p(x|z) and the possibility of learning p(z) during the train could potentially further boost the performances of this algorithm and will be the subject of future studies with real LHC collision data.

The model shown in Fig. 2.3 is implemented in `KERAS+TENSORFLOW` [45, 46], trained with the Adam optimizer [47] on a SM dataset of 3.45M events, equivalent to an integrated luminosity of $\sim$ 100 $\mathrm{pb}^{-1}$. The SM validation dataset is made of 3.45M of statistically independent examples. Such a sample would be collected in about ten hours of continuous running, under the assumptions made in this study (see page 37). In training, we fix the batch size to 1000. We use early stopping with patience set to 20 and $\delta_{\min} = 0.005$, and we progressively reduce the learning rate on plateau, with patience set to 8 and $\delta_{\min} = 0.01$.

The model's training history is shown in Fig. 2.4. Figure 2.5 shows the comparison of the input and output distributions for the 21 HLF quantities in the validation dataset. A general good agreement is observed on the bulk of the distributions, even if some of the distributions are not well described in the tails. These discrepancies do not have a sizable impact on the anomaly-detection strategy, as shown at page 54. Nevertheless, alternative architectures were tested, in order to reduce these discrepancies. For instance, we increased or decreased the dimensionality of the latent space, we changed the value of $\beta$ in Eq.(2.6), we changed the number of neurons in the hidden layers, tried the RMSprop optimizer, and used plain Gaussian functions to describe the 21 input features. Some of these choices improved the encoding-decoding capability of the VAE, with up to a 10% decrease of the loss function at the end of the training. On the other hand, none of these alternative models provided a sizable improvement in the anomaly-detection performance. For simplicity, we decided to limit our study to the architecture in Fig. 2.3 and dropped these alternative models.

**Supervised Classifiers**

For each of the four BSM benchmark models, we train a fully-supervised classifier, based on a Boosted Decision Tree (BDT). Each BDT receives as input the same

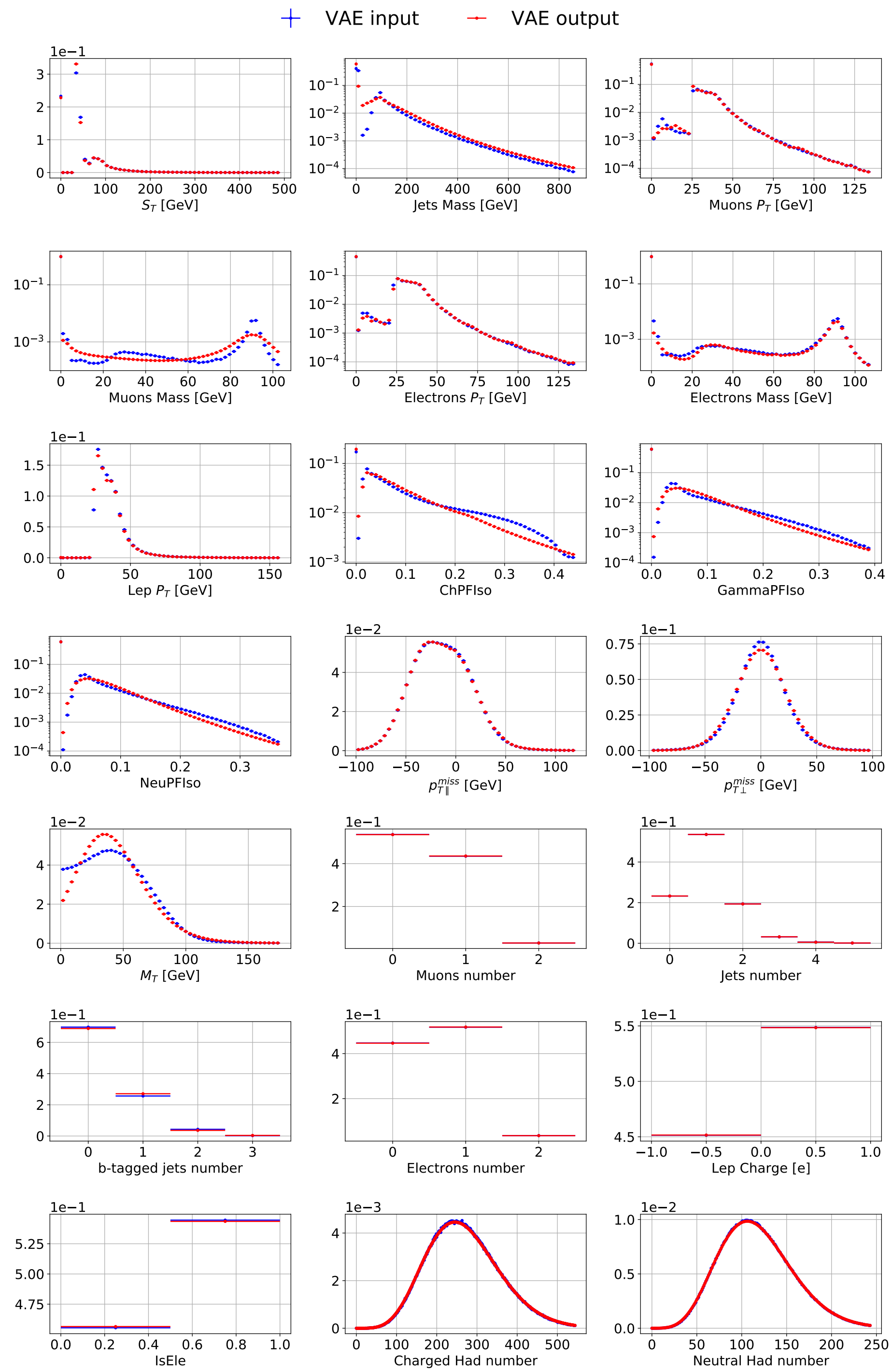


Figure 2.5: Comparison of input (blue) and output (red) probability distributions for the HLF quantities in the validation sample. The input distributions are normalized to unity. The output distributions are obtained summing over the predicted pdf of each event, normalized to the inverse of the total number of events (so that the total sum is normalized to unity).

21 features used by the VAE and is trained on a labeled dataset consisting of the SM cocktail (the background) and one of the four BSM benchmark models (the signal). The implementation is done through the Gradient Boosted Classifier of the scikit-learn library [48]. The algorithm was tuned with up to 150 estimators, minimum samples per leaf and maximum depth equal to 3, a learning rate of 0.1, and a tolerance of $10^{-4}$ on the validation loss function (choose to be the default deviance). Each BDT, tailored to a specific BSM model, is trained on 3.45M SM events and about 0.5M BSM events, consistently up-weighted in order to match the size of the SM sample during the training.

Table 2.2: Classification performance of the four BDT classifiers described in the text, each trained on one of the four BSM benchmark models. The two set of values correspond to the area under ROC curve (AUC), and to the true positive rate (TPR) for a SM false positive rate $\epsilon_{SM} = 5.4 \cdot 10^{-6}$, i.e., to $\sim$ 1000 SM events accepted every month.

| Process | AUC | TPR [%] |
|---|---|---|
| $A \to 4\ell$ | 0.98 | 5.4 |
| $LQ \to b\tau$ | 0.94 | 0.2 |
| $h^0 \to \tau\tau$ | 0.90 | 0.1 |
| $h^\pm \to \tau\nu$ | 0.97 | 0.3 |

We show in Table 2.2 and in Figure 2.6 the classification performance of the four supervised BDTs, which set a qualitative upper limit for the VAE's results. Overall, the four models can be discriminated with good accuracy, with some loss of performance for those models sharing similarities with specific SM processes (e.g., $h^0 \to \tau\tau$ exhibiting single- and double-lepton topology with missing transverse energy, typical of $t\bar{t}$ events). In the table, we also quote the true-positive rate (TPR) for each BSM model corresponding to a working point of SM false positive rate $\epsilon_{SM} = 5.4 \cdot 10^{-6}$, corresponding to an average of $\sim$ 1000 SM events accepted every month.

In addition to BDTs, we experimented with fully-connected deep neural networks (DNNs) with two hidden layers. Despite trying different architectures, we did not find a configuration in which the DNN classifiers could outperform the BDTs. This is due to the fact that, given the limited complexity of the problem at hand, a simple BDT can extract the maximum discrimination power from the 21 inputs. The limiting factor preventing larger auc values is not to be found in the model complexity but in the discriminating power of the 21 input features. Not being

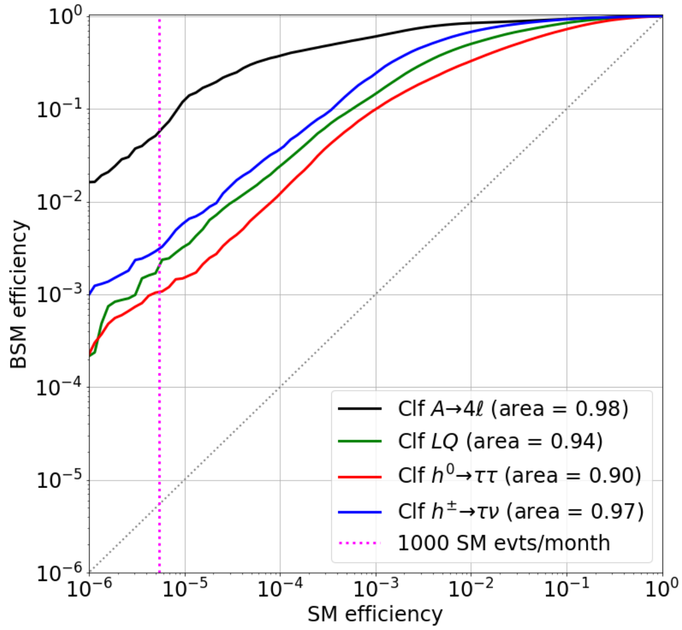


Figure 2.6: ROC curves for the fully-supervised BDT classifiers, optimized to separate each of the four BSM benchmark models from the SM cocktail dataset.

tailored on the benchmark BSM scenarios, these features do not carry all the needed information for an optimal signal-to-background separation. While one could obtain a better performance with more tailored classifiers, the purpose of this exercise was to provide a fair comparison with the VAE. In view of these considerations, we decided to use the BDTs as reference supervised classifiers.

**Results with VAE**

An event is classified as anomalous whenever the associated loss, computed from the VAE output, is above a given threshold. Since no BSM signal has been observed by LHC experiments so far, it is reasonable to expect that a new-physics signal, if any, would be characterized by a low production cross section and/or features very similar to those of a SM process. In view of this, we decided to use a tight threshold value, in order to reduce as much as possible any SM contribution.

Figure 2.7 shows the distribution of the $\text{Loss}_{\text{reco}}$ and $D_{\text{KL}}$ loss components for the validation dataset. In both plots, the vertical line represents a lower threshold such that a fraction $\epsilon_{SM} = 5.4 \cdot 10^{-6}$ of the SM events would be retained. This threshold value would result in $\sim$ 1000 SM events to be selected every month, i.e., a daily rate

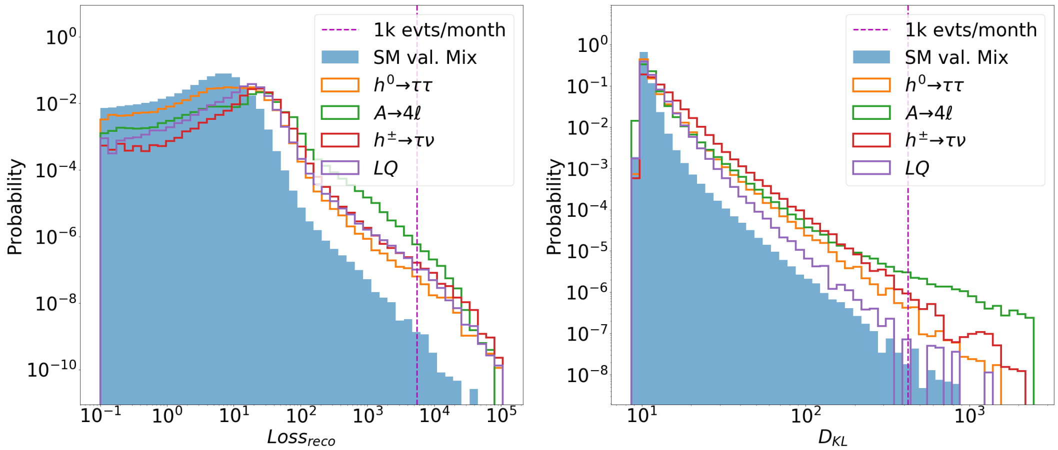


Figure 2.7: Distribution of the VAE's loss components, $\text{Loss}_{\text{reco}}$ (left) and $D_{\text{KL}}$ (right), for the validation dataset. For comparison, the corresponding distribution for the four benchmark BSM models are shown. The vertical line represents a lower threshold such that $5.4 \cdot 10^{-6}$ of the SM events would be retained, equivalent to $\sim 1000$ expected SM events per month.

of $\sim$ 33 SM events, as illustrated in Table 2.3. The acceptance rate is calculated assuming the LHC running conditions listed in Section 2.1. Table 2.3 also reports the by-process VAE selection efficiency and the relative background composition of the selected sample.

Figure 2.7 also shows the $\text{Loss}_{\text{reco}}$ and $D_{\text{KL}}$ distributions for the four benchmark BSM models. We observe that the discrimination power, loosely quantified by the integral of these distributions above threshold, is better for $\text{Loss}_{\text{reco}}$ than $D_{\text{KL}}$ and that the impact of the $D_{\text{KL}}$ term on $\text{Loss}_{\text{Tot}}$ is negligible. Anomalies are then defined as events laying on the right tail of the expected $\text{Loss}_{\text{reco}}$ distribution. Due to limited statistics in the training sample, the p-value corresponding to the chosen threshold value could be uncalibrated. This could result in a deviation of the observed rate from the expected value, an issue that one can address tuning the threshold. On the other hand, an uncalibrated p-value would also impact the number of collected BSM events, and the time needed to collect an appreciable amount of these events.

Once the $\text{Loss}_{\text{reco}}$ selection is applied, the anomalous events do not cluster on the tails of the distributions of the input features. Instead, they tend to cover the full feature-definition range. This is an indication of the fact that the VAE does more than a simple selection of feature outliers, which is what is done by traditional single-lepton trigger or by dedicated cross triggers (e.g., triggers that select events

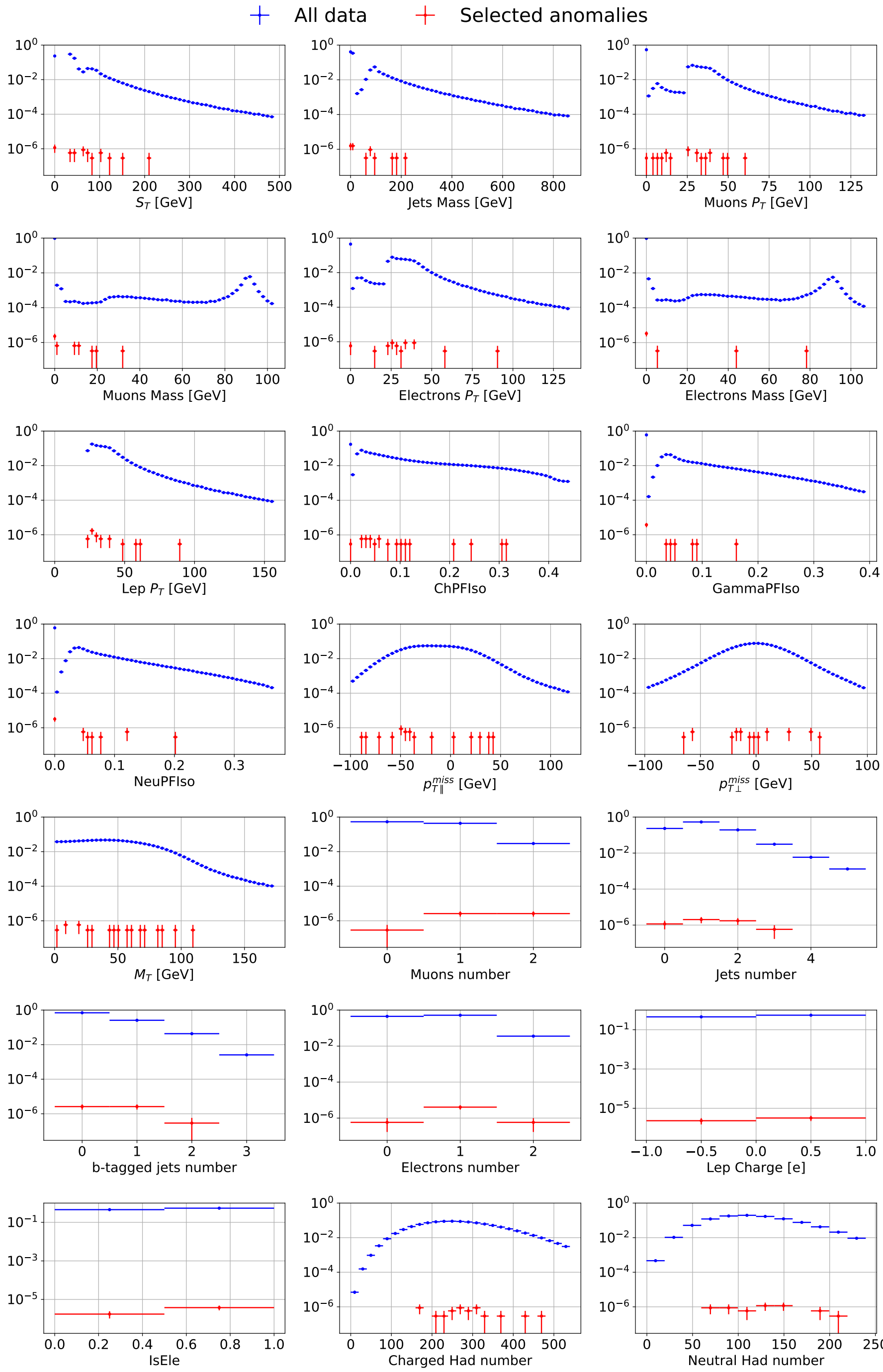


Figure 2.8: Comparison between the input distribution for the 21 HLF of the validation dataset (blue histograms) and the distribution of the SM outlier events selected from the same sample by applying the $\text{Loss}_{\text{reco}}$ threshold (red dots). The outlier events cover a large portion of the HLF definition range and do not cluster on the tails.

Table 2.3: By-process acceptance rate for the anomaly detection algorithm described in the text, computed applying the threshold on $\text{Loss}_{\text{reco}}$ shown in Fig. 2.7. The threshold is tuned such that a fraction of about $\epsilon_{SM} = 5.4 \cdot 10^{-6}$ of SM events would be accepted, corresponding to $\sim$ 1000 SM events/month, assuming the LHC running conditions listed in Section 2.1. The sample composition refers to the subset of SM events accepted by the anomaly detection algorithm. All quoted uncertainties refer to 95% CL regions.

| Standard Model processes | | | |
|---|---|---|---|
| Process | VAE selection | Sample composition | Events/month |
| $W$ | $3.6 \pm 0.7 \cdot 10^{-6}$ | 32% | $379 \pm 74$ |
| QCD | $6.0 \pm 2.3 \cdot 10^{-6}$ | 29% | $357 \pm 143$ |
| $Z$ | $21 \pm 3.5 \cdot 10^{-6}$ | 21% | $256 \pm 43$ |
| $t\bar{t}$ | $400 \pm 9 \cdot 10^{-6}$ | 18% | $212 \pm 5$ |
| Tot | | | $1204 \pm 167$ |

with soft leptons and large missing transverse energy, $S_T$, etc.). This is shown in Fig. 2.8 for SM events. A similar conclusion can be obtained from Fig. 2.9, showing the distribution of the 21 input HLF quantities for the $A \rightarrow 4\ell$ benchmark model, before and after applying the threshold requirement on the VAE loss.

The left plot in Fig. 2.10 shows the ROC curves obtained from the $\text{Loss}_{\text{reco}}$ distribution of the four BSM benchmark models and the SM cocktail, compared to the corresponding BDT curves of Section 2.1. As expected, the results obtained with the supervised BDTs outperform the VAE. On the other hand, the VAE can probe at the same time the four scenarios with comparable performances. This is a consequence of the trade off between precision and model independence and an illustration of the complementarity between the approach presented in this work and traditional supervised techniques. The right plot in Fig. 2.10 shows the one-sided p-value computed from the cocktail SM distribution, both for the SM events themselves (flat by construction) and for the four BSM processes. As the plot shows, BSM processes tend to concentrate at small p-values, which allows their identification as anomalies.

Table 2.4 summarizes the VAE's performance on the four BSM benchmark models. Together with the selection efficiency corresponding to $\epsilon_{SM} = 5.4 \cdot 10^{-6}$, the table reports the effective cross section (cross section after applying the trigger requirements) that would correspond to 100 BSM events selected in a month (assuming an integrated luminosity of 5 fb$^{-1}$). Similarly, we quote the cross section that would result in a signal-to-background ratio of 1/3 on the sample of events selected by the VAE. The VAE can probe the four models down to small cross section values,

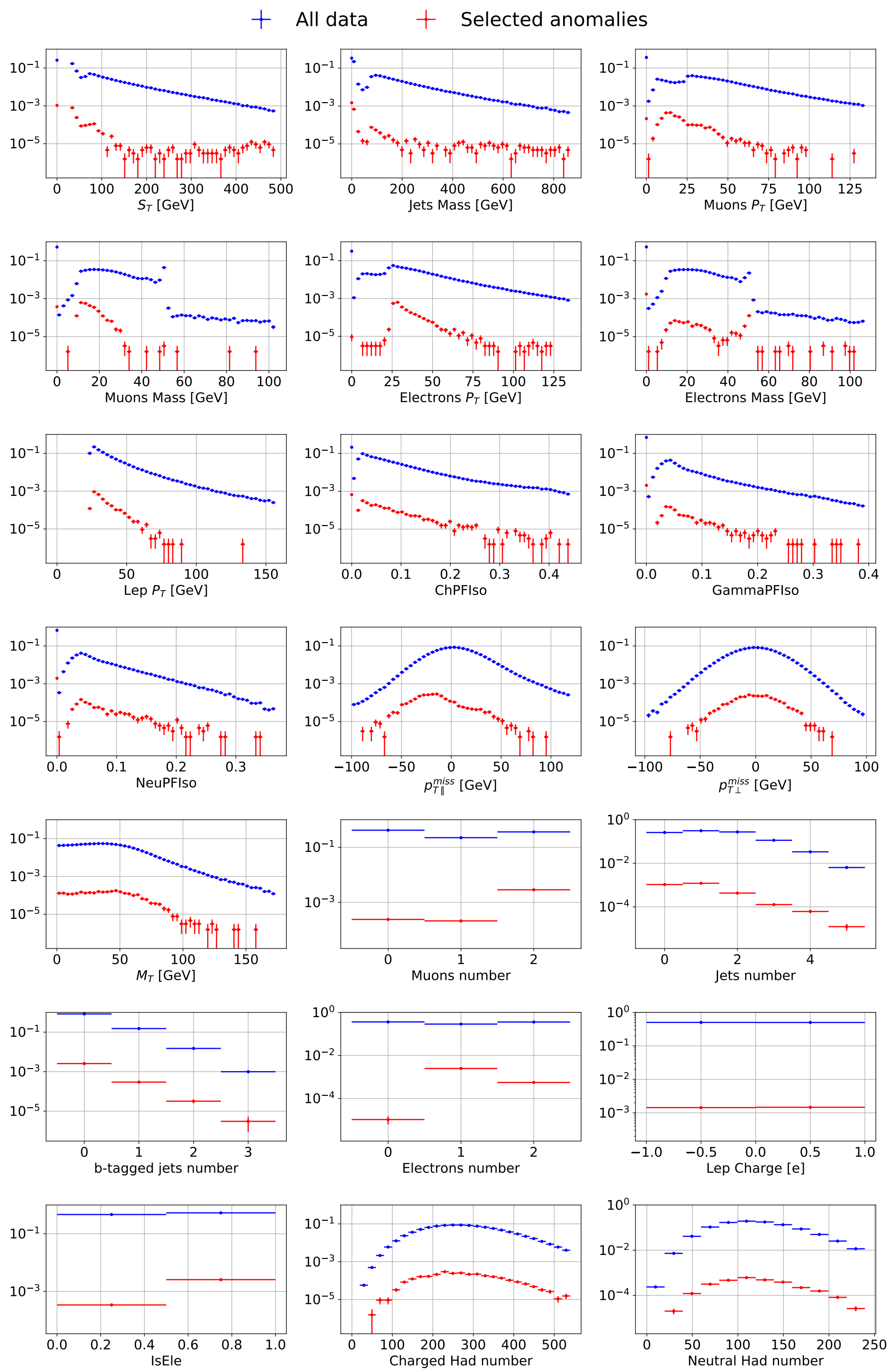


Figure 2.9: Comparison between the distribution of the 21 HLF distribution for $A \rightarrow 4\ell$ full dataset (blue) and $A \rightarrow 4\ell$ events selected by applying the $\text{Loss}_{\text{reco}}$ threshold (red). The selected events are not trivially sampled from the tail.

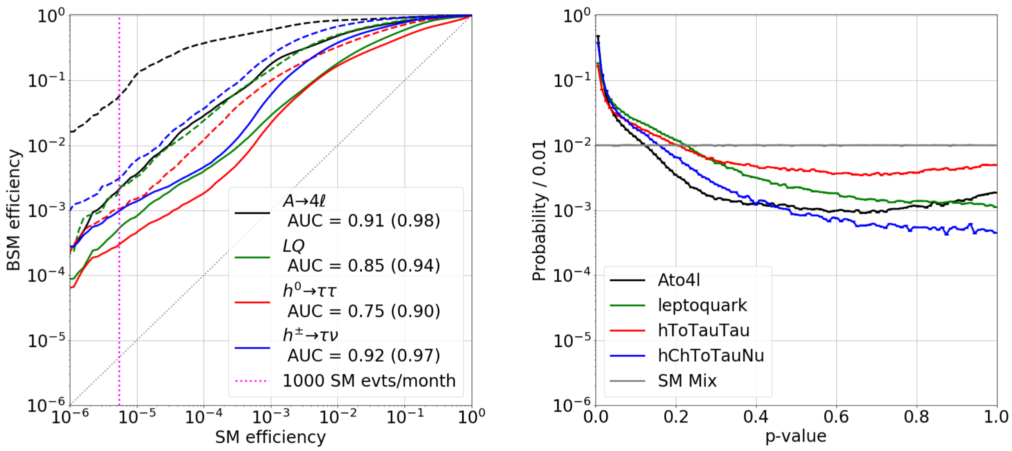


Figure 2.10: Left: ROC curves for the VAE trained only on SM events (solid), compared to the corresponding curves for the four supervised BDT models (dashed) described in Section 2.1. Right: Normalized p-value distribution distribution for the SM cocktail events and the four BSM benchmark processes.

comparable to the existing exclusion bounds for these mass ranges. As an example, Ref. [49] excludes a $LQ \rightarrow \tau b$ with a mass of 150 GeV and production cross section larger than $\sim$ 10 pb, using 4.8 fb$^{-1}$ at a center-of-mass energy of 7 TeV, while most recent searches [50] cannot cover such a low mass value, due to trigger limitations.

Table 2.4: Breakdown of BSM processes efficiency, and cross section values corresponding to 100 selected events in a month and to a signal-over-background ratio of 1/3 (i.e., an absolute yield of $\sim$ 400 events/month). The monthly event yield is computed assuming an average luminosity per month of 5 fb$^{-1}$, computing by taking the LHC 2016 data delivery ($\sim$ 40 fb$^{-1}$ collected in 8 months). All quoted efficiencies are computed fixing the VAE loss threshold $\epsilon_{SM} = 5.4 \cdot 10^{-6}$.

| BSM benchmark processes | | | |
|---|---|---|---|
| Process | VAE selection efficiency | Cross-section 100 events/month [pb] | Cross-section S/B = 1/3 [pb] |
| $A \rightarrow 4\ell$ | $2.8 \cdot 10^{-3}$ | 7.1 | 27 |
| $LQ \rightarrow b\tau$ | $6.7 \cdot 10^{-4}$ | 30 | 110 |
| $h^0 \rightarrow \tau\tau$ | $3.6 \cdot 10^{-4}$ | 55 | 210 |
| $h^{\pm} \rightarrow \tau\nu$ | $1.2 \cdot 10^{-3}$ | 17 | 65 |

Unlike a traditional trigger strategy, a VAE-based selection is mainly intended to select a high-purity sample of interesting event, at the cost of a typically small selection efficiency. To demonstrate this point, we consider a sample selected with the VAE and one selected using a typical inclusive single lepton trigger (SLT), consisting on a tighter selection than the one described in section 2.1. In particular, we require $p_T^\ell > 27$ GeV and ISO $< 0.25$. We consider the signal-over-background ratio (SBR) for the VAE's threshold selection and the SLT. While these quantities depend on the production cross section of the considered BSM model, their ratio

$$\frac{\mathrm{SBR}_{\mathrm{VAE}}}{\mathrm{SBR}_{\mathrm{SLT}}} = \left(\frac{\epsilon_{\mathrm{SLT}}}{\epsilon_{\mathrm{VAE}}}\right)_{SM} \cdot \left(\frac{\epsilon_{\mathrm{VAE}}}{\epsilon_{\mathrm{SLT}}}\right)_{BSM} \tag{2.16}$$

is only a function of the selection efficiency for the SLT ($\epsilon_{\mathrm{SLT}}$) and the for the VAE $\epsilon_{\mathrm{VAE}}$ for SM and BSM events. Table 2.5 shows how the SBR reached by the VAE is about two order of magnitude larger than what a traditional inclusive SLT could reach.

Table 2.5: Selection efficiencies for a typical single lepton trigger (SLT) and the proposed VAE selection, shown for the four benchmark BSM models and for the SM cocktail. The last row quotes the corresponding BSM-to-SM ratio of signal-over-background ratios (SBRs), quantifying the purity of the selected sample.

| | SM | $A \to 4\ell$ | $LQ \to b\tau$ | $h^0 \to \tau\tau$ | $h^\pm \to \tau\nu$ |
|---|---|---|---|---|---|
| $\epsilon_{\mathrm{VAE}}$ | $5.3 \cdot 10^{-6}$ | $2.8 \cdot 10^{-3}$ | $6.7 \cdot 10^{-4}$ | $3.6 \cdot 10^{-4}$ | $1.2 \cdot 10^{-3}$ |
| $\epsilon_{\mathrm{SLT}}$ | 0.6 | 0.5 | 0.6 | 0.7 | 0.6 |
| $\epsilon_{SLT}/\epsilon_{\mathrm{VAE}}$ | $1.1 \cdot 10^5$ | $1.8 \cdot 10^2$ | $9.0 \cdot 10^2$ | $1.7 \cdot 10^3$ | $5.8 \cdot 10^2$ |
| $\mathrm{SBR}_{\mathrm{VAE}}/\mathrm{SBR}_{\mathrm{SLT}}$ | - | 625 | 125 | 70 | 191 |

**Comparison with Auto-Encoder**

For sake of completeness, we repeated the strategy presented in this work on a simple AE. The architecture was fixed to be as close as possible to that of the VAE introduced in Sec. 2.1. The change from VAE to AE imply these two changes: the output layer has the same dimensionality of the input layer; the latent layer includes four neurons (as opposed to 8), corresponding to the four latent variables $z$ (and not to the $\mu$ and $\sigma$ parameters of the $z$ distribution). An MSE loss function is used. The optimizer and callbacks used to trained the VAE are are used in this case. Figure 2.11 shows the loss function distribution and a comparison between the ROC curves of the VAE and AE. These distributions directly compare to the left plots of Figs. 2.7 and 2.10, since in that case only the reconstruction part of the loss was used. For convenience, the VAE ROC curves are also shown here, represented by the dashed lines. When considering the four BSM benchmark models presented in this work, the AE provides competitive performances, for some choice of the SM accepted-event rate. On the other hand, the VAE usually outperforms a plain AE

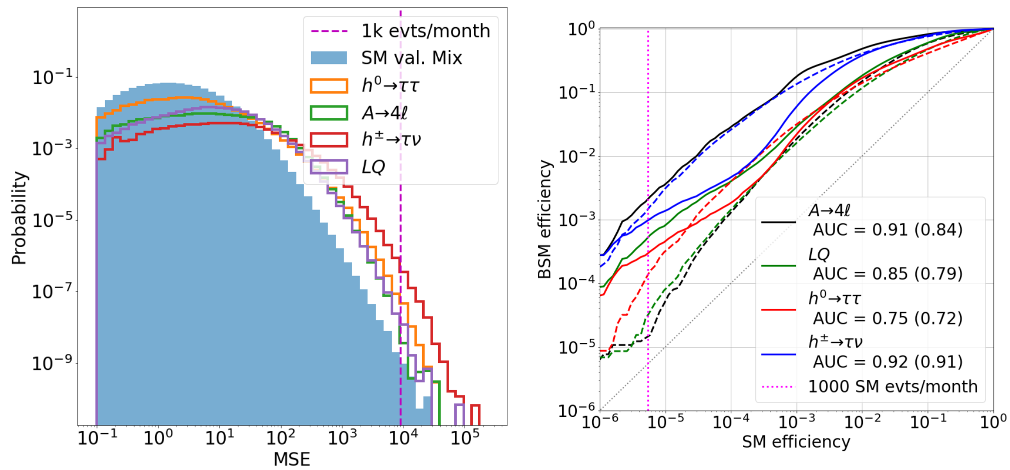


Figure 2.11: Left: Distribution of the AE loss (MSE) for the validation dataset. The distribution for the SM processes and the four benchmark BSM models are shown. Right: ROC curves for the AE (dashed lines) trained only on SM mix, compared to the corresponding VAE curves from Fig. 2.10 (solid). The vertical dotted line represents the $\epsilon_{SM} = 5.4 \cdot 10^{-6}$ threshold considered in this study.

for the rate considered in this study ($\epsilon_{SM} = 5.4 \cdot 10^{-6}$). With the exception of the $h^{\pm} \to \tau\nu$ model (for which the AE provides a 30% larger efficiency than the VAE), the VAE provides larger efficiency on the BSM models, with improvements as large as two orders of magnitude (for the $A \to 4\ell$ model).

**How to Deploy a VAE for BSM Detection**

The work presented in here suggests the possibility of deploying a VAE as a trigger algorithm associated to dedicated data streams. This trigger would isolate anomalous events, similarly to what was done by the CMS experiment at the beginning of the first LHC run. With an early new physics signal being a possibility at the LHC start, the CMS experiment deployed online a set of algorithms (collectively called *hot line*) to select potentially interesting new-physics candidates. At that time, anomalies were characterized as events with high-$p_T$ particles or high particle multiplicities, in line with the kind of early-discovery new physics scenarios considered at that time. The events populating the hot-line stream were immediately processed at the CERN computing center (as opposed to traditional physics streams, that are processed after 48 hours). The hot-line algorithms were tuned to collect $O(10)$ events per day, which were then visually inspected by experts.

While the focus of the work presented is not an early discovery, the spirit of the application we propose would be similar: a set of VAEs deployed online would select a limited number of events every day. These events would be collected in a dedicated dataset and further analyzed. The analysis technique could go from visual inspection of the collisions to detailed studies of reconstructed objects, up to some kind of model-independent analysis of the collected dataset, e.g. a deep-learning implementation of a model-independent hypothesis testing [24] directly on the loss distribution (provided a reliable sample of background-only data).

While a pure SM sample to train VAEs could only be obtained from a MC simulation, the presence of outlier contamination in the training sample has typically a tiny impact on performance. One could then imagine to train the VAE models on so-far collected data and use them on the events entering the HLT system. Such a training could happen offline on a dedicated dataset, e.g., deploying triggers randomly selecting events entering the last stage of the trigger system. The training could even happen online, assuming the availability of sufficient computing resources. As it happens with normal triggers, at the very beginning one would use some MC sample or some control sample from previously collected data to estimate the threshold corresponding to the target SM rate. Then, as it happens normally during HLT operations, the threshold will have to be monitored on real data and adjusted if needed.

To demonstrate the feasibility of a train-on-data strategy, we enrich the dataset used in Section 2.1 with a signal contamination of $A \to 4\ell$ events. As a starting point, the amount of injected signal is tuned to a luminosity of 100 pb$^{-1}$ and a cross section of 7.1 pb, corresponding to the value at which the VAE in Section 2.1 would select 100 $A \to 4\ell$ events in one month. This results in about 700 $A \to 4\ell$ events added to the training sample. The VAE is trained following the procedure outlined in Section 2.1 and its performance is compared to that obtained on a signal-free dataset of the same size. The comparison of the ROC curves for the two models is shown in Fig. 2.12. In the same figure, we show similar results, derived injecting a $\times 10$ and $\times 100$ signal contamination. A performance degradation is observed once the signal cross section is set to 710 pb (i.e., 100 times larger than the sensitivity value found in Section 2.1). At that point, the contamination is so large that the signal becomes as abundant as $t\bar{t}$ events and would have easily detectable consequences. For comparison, at a production cross section of 27 pb a third of the events selected by the VAE in Section 2.1 would come from $A \to 4\ell$ production (see Table 2.4). Such a large yield would still have negligible consequences on the training quality.

This test shows that a robust anomaly-detecting VAE could be trained directly on data, even in presence of previously undetected (e.g., at Tevatron, 7 TeV and 8-TeV LHC) BSM signals.

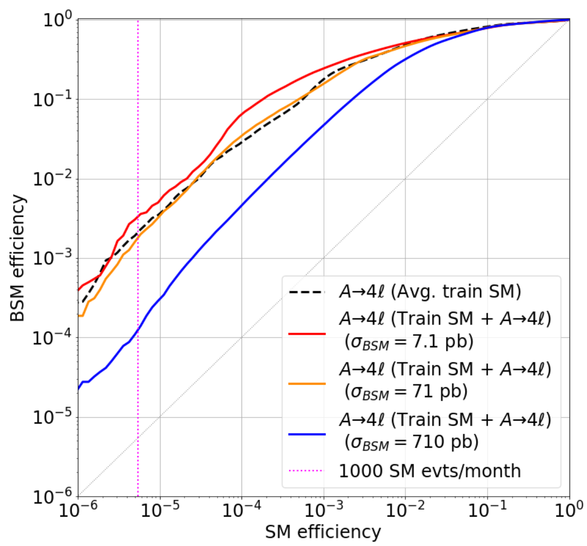


Figure 2.12: ROC curves for the VAE trained on SM contaminated with and without $A \to 4\mu$ contamination. Different levels of contamination are reported corresponding to 0.02% ($\sigma$ = 7.15 pb - equal to the estimated one to have 100 events per month), 0.19% ($\sigma$ = 71.5 pb) and 1.89% ($\sigma$ = 715 pb) of the training sample.

The possibility of training the VAE on data would substantially simplify the implementation of the strategy proposed in this work, since any possible systematic bias in the data would be automatically taken into account during the training process. In addition, it would make the procedure robust against other systematic effects (e.g., energy scale, efficiency, etc.) that would affect a MC-based training.

### Conclusions and Outlook

We present a strategy to isolate potential BSM events produced by the LHC, using variational autoencoders trained on a reference SM sample. Such an algorithm could be used in the trigger system of general-purpose LHC experiments to identify recurrent anomalies, which might otherwise escape observation (e.g., being filtered out by a typical trigger selection). Taking as an example a single-lepton data stream, we show how such an algorithm could select datasets enriched with events originating from challenging BSM models. We also discuss how the algorithm could be trained directly on data, with no sizable performance loss, more robustness against systematic uncertainties, and a big simplification of the training and deployment procedure.

The main purpose of such an application is not to enhance the signal selection efficiency for BSM models. Indeed, this application is tuned to provide a high-purity sample of potentially interesting events. We showed that events produced by not-yet-excluded BSM models with cross sections in the range of $O(10)$ to $O(100)$ pb could be isolated in a $\sim 30\%$ pure sample of $\sim 43$ events selected per day. The price to pay to reach such a purity is a relatively small signal efficiency and a strong bias in the dataset definition, which makes these events marginal and difficult to use in a traditional data-driven and supervised search for new physics.

The final outcome of this application would be a list of anomalous events, that the experimental collaborations could further scrutinize and even release as a catalog, similar to what is typically done in other scientific domains. Repeated patterns in these events could motivate new scenarios for beyond-the-standard-model physics and inspire new searches, to be performed on future data with traditional supervised approaches.

We stress the fact that the power of the proposed approach is in its generality and not in its sensitivity to a particular BSM scenario. We show that a simple BDT could give a better discrimination capability for a given BSM hypothesis. On the other hand, such a supervised algorithm would not generalize to other BSM scenarios. The VAE, instead, comes with little model dependence and therefore generalizes to unforeseen BSM models. On the other hand, the VAE cannot guarantee optimal performance in any scenario. As typical of autoencoders used for anomaly detection, our VAE model is trained to learn the SM background at best, but there is no guarantee that the best SM-learning model will be the best anomaly detection algorithm. By definition, the anomaly detection capability of the algorithm does not enter the loss function, as well as, by construction, no signal event enters the training sample. This is the price to pay when trading discrimination power for model independence.

We believe that such an application could help extend the physics reach of the current and next stages of the CERN LHC. The proposed strategy is demonstrated for a single-lepton data stream coming from a typical L1 selection. On the other hand, this approach could be generalized to any other data stream coming from any L1 selection, so that the full $\sim 100$ Hz rate entering the HLT system of ATLAS or CMS could be scrutinized. While the L1 selection still represents a potentially dangerous bias, an algorithm running in the HLT could access 100 times more events than the $\sim$ 1 kHz stream typically available for offline studies. Moreover, thanks to progress in the deployment of deep neural networks on FPGA boards [51], it is conceivable that VAEs for anomaly detection could be also deployed in the L1 trigger systems in a near future. In this way, the VAE would access the full L1 input data stream.

## 2.2 Other Contributions

As mentioned in the introduction, I contributed to field of development of machine-learing tools not only though the lead of the work presented in the previous section but also by direct collaboration in other projects and supervision of younger students on projects initiated from some of my previous work. In this section, a brief summary of those contribution is presented while a full description of the results is left to the references present in each paragraph.

### Re-discovering the Top Quark with Anomaly Detection Algorithms

As a natural continuation of the work in [5], we developed in [6] a real data test of the strategy there proposed with minor algorithm developments. In this work, we apply an Adversarially Learned Anomaly Detection (ALAD) algorithm to the problem of detecting new physics processes in proton-proton collisions at the LHC. Anomaly detection based on ALAD matches performances reached by Variational Autoencoders, with a substantial improvement in some cases. Training the ALAD algorithm on 4.4 $fb^{-1}$ of 8 TeV CMS Open Data, we show how a data-driven anomaly detection and characterization would work in real life, re-discovering the top quark by identifying the main features of the $t\bar{t}$ experimental signature at the LHC.

### Pileup Mitigation with Graph Neural Networks

As discussed in the introduction chapter, at the LHC, the high transverse-momentum events studied by experimental collaborations occur in coincidence with parasitic low transverse-momentum collisions, usually referred to as pileup. Pileup mitigation is a key ingredient of the online and offline event reconstruction as pileup affects the reconstruction accuracy of many physics observables. In [52] we present a classifier based on Graph Neural Networks, trained to retain particles coming from high transverse-momentum collisions while rejecting those coming from pileup collisions. This model is designed as a refinement of the PUPPI algorithm [53], employed in many LHC data analyses since 2015. Thanks to an extended basis of input information and the learning capabilities of the considered network architecture, we show an improvement in pileup-rejection performances with respect to state-of-the-art solutions.

**Jet Taggers Based on Interaction Networks**

In [54] we investigate the performance of a jet identification algorithm based on interaction networks (JEDI-net) to identify all-hadronic decays of high-momentum heavy particles produced at the LHC and distinguish them from ordinary jets originating from the hadronization of quarks and gluons. The jet dynamics are described as a set of one-to-one interactions between the jet constituents. Based on a representation learned from these interactions, the jet is associated to one of the considered categories. Unlike other architectures, the JEDI-net models achieve their performance without special handling of the sparse input jet representation, extensive pre-processing, particle ordering, or specific assumptions regarding the underlying detector geometry. The presented models give better results with less model parameters, offering interesting prospects for LHC applications.

In [55] we develop an algorithm based on an interaction network to identify high-transverse-momentum Higgs bosons decaying to bottom quark-antiquark pairs and distinguish them from ordinary jets that reflect the configurations of quarks and gluons at short distances. The algorithm's inputs are features of the reconstructed charged particles in a jet and the secondary vertices associated with them. Describing the jet shower as a combination of particle-to-particle and particle-to-vertex interactions, the model is trained to learn a jet representation on which the classification problem is optimized. The algorithm is trained on simulated samples of realistic LHC collisions, released by the CMS Collaboration on the CERN Open Data Portal. The interaction network achieves a drastic improvement in the identification performance with respect to state-of-the-art algorithms.

**Analysis-Specific Fast Simulation**

In [56] we present a fast-simulation application based on a deep neural network designed to create large analysis-specific datasets. Taking as an example the generation of W + jet events produced in $\sqrt{s}$ = 13 13 TeV proton–proton collisions, we train a neural network to model detector resolution effects as a transfer function acting on an analysis-specific set of relevant features computed at generation level, i.e. in absence of detector effects. Based on this model, we propose a novel fast-simulation workflow that starts from a large amount of generator-level events to deliver large analysis-specific samples. The adoption of this approach would result in about an order-of-magnitude reduction in computing and storage requirements for the collision simulation workflow. This strategy could help the high energy physics community to face the computing challenges of the future High-Luminosity LHC.

**Agnostic Selections For New Physics Searches**

In [57] we discuss a model-independent strategy for boosting new physics searches based on jets with the help of an unsupervised anomaly detection algorithm. Prior to a search, each input event is preprocessed by the algorithm - a variational autoencoder (VAE). Based on the loss assigned to each event, input data can be split into a background control sample and a signal enriched sample. Following this strategy, one can enhance the sensitivity to new physics with no assumption on the underlying new physics signature. Our results show that a typical BSM search on the signal enriched group is more sensitive than an equivalent search on the original dataset.

**References**


[1] Wikipedia. *Moore's law — Wikipedia, The Free Encyclopedia*. `http://en.wikipedia.org/w/index.php?title=Moore's%20law&oldid=1097387000`. [Online; accessed 14-July-2022]. 2022.

[2] Wikipedia. *Artificial neural network — Wikipedia, The Free Encyclopedia*. `http://en.wikipedia.org/w/index.php?title=Artificial%20neural%20network&oldid=1097455717`. [Online; accessed 14-July-2022]. 2022.

[3] Olmo Cerri. "Hadronic recoil in the W boson production at LHC for a W mass measurement with the CMS experiment". Presented 20 Sep 2017. 2017. URL: `https://cds.cern.ch/record/2285935`.

[4] Thong Q. Nguyen et al. "Topology classification with deep learning to improve real-time event selection at the LHC." In: *Comput. Softw. Big Sci.* 3.1 (2019), p. 12. DOI: `10.1007/s41781-019-0028-1`. arXiv: `1807.00083 [hep-ex]`.

[5] Olmo Cerri et al. "Variational autoencoders for new physics mining at the Large Hadron Collider." In: *J. High Energy Phys.* 2019.5 (2019), p. 36. DOI: `10.1007/JHEP05(2019)036`.

[6] Oliver Knapp et al. "Adversarially Learned Anomaly Detection on CMS Open Data: Re-discovering the top quark." In: *Eur. Phys. J. Plus* 136.2 (2021), p. 236. DOI: `10.1140/epjp/s13360-021-01109-4`. arXiv: `2005.01598 [hep-ex]`.

[7] Wikipedia. *Variational autoencoder — Wikipedia, The Free Encyclopedia*. `http://en.wikipedia.org/w/index.php?title=Variational%20autoencoder&oldid=1096166874`. [Online; accessed 17-July-2022]. 2022.

[8] Olmo Cerri. *Anomaly Detection VAE*. `https://github.com/ocerri/AnomalyDetection`. 2020.

[9] "Procedure for the LHC Higgs boson search combination in summer 2011". In: (2011).

[10] Georges Aad et al. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". In: *Phys. Lett.* B716 (2012), pp. 1–29. DOI: `10.1016/j.physletb.2012.08.020`. arXiv: `1207.7214 [hep-ex]`.

[11] Serguei Chatrchyan et al. "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC". In: *Phys. Lett.* B716 (2012), pp. 30–61. DOI: `10.1016/j.physletb.2012.08.021`. arXiv: `1207.7235 [hep-ex]`.

[12] T. Aaltonen et al. "Global Search for New Physics with 2.0 $\mathrm{fb}^{-1}$ at CDF". In: *Phys. Rev.* D79 (2009), p. 011101. DOI: 10.1103/PhysRevD.79.011101. arXiv: 0809.3781 [hep-ex].

[13] Victor Mukhamedovich Abazov et al. "Model independent search for new phenomena in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV". In: *Phys. Rev.* D85 (2012), p. 092015. DOI: 10.1103/PhysRevD.85.092015. arXiv: 1108.5362 [hep-ex].

[14] F. D. Aaron et al. "A General Search for New Phenomena at HERA". In: *Phys. Lett.* B674 (2009), pp. 257–268. DOI: 10.1016/j.physletb.2009.03.034. arXiv: 0901.0507 [hep-ex].

[15] *MUSiC, a Model Unspecific Search for New Physics, in pp Collisions at* $\sqrt{s} = 8$ TeV. Tech. rep. CMS-PAS-EXO-14-016. Geneva: CERN, 2017. URL: https://cds.cern.ch/record/2256653.

[16] Morad Aaboud et al. "A strategy for a general search for new phenomena using data-derived signal regions and its application within the ATLAS experiment". In: *Eur. Phys. J.* C79.2 (2019), p. 120. DOI: 10.1140/epjc/s10052-019-6540-y. arXiv: 1807.07447 [hep-ex].

[17] Morad Aaboud et al. "Performance of the ATLAS Trigger System in 2015". In: *Eur. Phys. J.* C77.5 (2017), p. 317. DOI: 10.1140/epjc/s10052-017-4852-3. arXiv: 1611.09661 [hep-ex].

[18] Vardan Khachatryan et al. "The CMS trigger system". In: *JINST* 12.01 (2017), P01020. DOI: 10.1088/1748-0221/12/01/P01020. arXiv: 1609.02366 [physics.ins-det].

[19] D. P Kingma and M. Welling. "Auto-Encoding Variational Bayes". In: *ArXiv e-prints* (Dec. 2013). arXiv: 1312.6114 [stat.ML].

[20] Jinwon An and Sungzoon Cho. "Variational autoencoder based anomaly detection using reconstruction probability". In: *Special Lecture on IE* 2 (2015), pp. 1–18.

[21] L. Lyons. "Open statistical issues in particle physics". In: *ArXiv e-prints* (Nov. 2008). arXiv: 0811.1663 [stat.AP].

[22] Eilam Gross and Ofer Vitells. "Trial factors for the look elsewhere effect in high energy physics". In: *Eur. Phys. J.* C70 (2010), pp. 525–530. DOI: 10.1140/epjc/s10052-010-1470-8. arXiv: 1005.1891 [physics.data-an].

[23] Thong Q. Nguyen et al. "Topology classification with deep learning to improve real-time event selection at the LHC". In: (2018). arXiv: 1807.00083 [hep-ex].

[24] Raffaele Tito D'Agnolo and Andrea Wulzer. "Learning New Physics from a Machine". In: *Phys. Rev.* D99.1 (2019), p. 015014. DOI: 10.1103/PhysRevD.99.015014. arXiv: 1806.02350 [hep-ph].

[25] Jack H. Collins, Kiel Howe, and Benjamin Nachman. “Anomaly Detection for Resonant New Physics with Machine Learning”. In: *Phys. Rev. Lett.* 121.24 (2018), p. 241803. DOI: 10.1103/PhysRevLett.121.241803. arXiv: 1805.02664 [hep-ph].

[26] Andrea De Simone and Thomas Jacques. “Guiding New Physics Searches with Unsupervised Learning”. In: (2018). arXiv: 1807.06038 [hep-ph].

[27] Jan Hajer et al. “Novelty Detection Meets Collider Physics”. In: (2018). arXiv: 1807.10261 [hep-ph].

[28] Adrian Alan Pol et al. “Detector monitoring with artificial neural networks at the CMS experiment at the CERN Large Hadron Collider”. In: (2018). arXiv: 1808.00911 [physics.data-an].

[29] *Anomaly detection using Deep Autoencoders for the assessment of the quality of the data acquired by the CMS experiment*. Tech. rep. Geneva: CERN, 2018. URL: http://cds.cern.ch/record/2630433.

[30] *Deep generative models for fast shower simulation in ATLAS*. Tech. rep. ATL-SOFT-PUB-2018-001. Geneva: CERN, 2018. URL: http://cds.cern.ch/record/2630433.

[31] Theo Heimel et al. “QCD or What?” In: (2018). arXiv: 1808.08979 [hep-ph].

[32] Marco Farina, Yuichiro Nakai, and David Shih. “Searching for New Physics with Deep Autoencoders”. In: (2018). arXiv: 1808.08992 [hep-ph].

[33] Bernhard Schölkopf et al. “Estimating the support of a high-dimensional distribution”. In: *Neural computation* 13.7 (2001), pp. 1443–1471.

[34] Fei Tony Liu, Kai Ming Ting, and Zhi-Hua Zhou. “Isolation forest”. In: *Data Mining, 2008. ICDM’08. Eighth IEEE International Conference on*. IEEE. 2008, pp. 413–422.

[35] Fei Tony Liu, Kai Ming Ting, and Zhi-Hua Zhou. “Isolation-based anomaly detection”. In: *ACM Transactions on Knowledge Discovery from Data (TKDD)* 6.1 (2012), p. 3.

[36] Charu C Aggarwal. “Outlier analysis”. In: *Data mining*. Springer. 2015, pp. 237–263.

[37] Mevlana Gemici et al. “Generative Temporal Models with Memory”. In: *CoRR* abs/1702.04649 (2017). arXiv: 1702.04649. URL: http://arxiv.org/abs/1702.04649.

[38] Torbjörn Sjöstrand et al. “An Introduction to PYTHIA 8.2”. In: *Comput. Phys. Commun.* 191 (2015), pp. 159–177. DOI: 10.1016/j.cpc.2015.01.024. arXiv: 1410.3012 [hep-ph].

[39] J. de Favereau et al. “DELPHES 3, A modular framework for fast simulation of a generic collider experiment”. In: *JHEP* 02 (2014), p. 057. DOI: 10.1007/JHEP02(2014)057. arXiv: 1307.6346 [hep-ex].

[40] D Contardo et al. *Technical Proposal for the Phase-II Upgrade of the CMS Detector*. Tech. rep. CERN-LHCC-2015-010. LHCC-P-008. CMS-TDR-15-02. Geneva: CERN, 2015. URL: `https://cds.cern.ch/record/2020886`.

[41] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. "FastJet User Manual". In: *Eur. Phys. J.* C72 (2012), p. 1896. DOI: `10.1140/epjc/s10052-012-1896-2`. arXiv: `1111.6097 [hep-ph]`.

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

[43] Irina Higgins et al. "beta-vae: Learning basic visual concepts with a constrained variational framework". In: (2016).

[44] Jakub M. Tomczak and Max Welling. "VAE with a VampPrior". In: *CoRR* abs/1705.07120 (2017). arXiv: `1705.07120`. URL: `http://arxiv.org/abs/1705.07120`.

[45] François Chollet et al. *Keras*. `https://github.com/fchollet/keras`. 2015.

[46] Martin Abadi et al. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Software available from tensorflow.org. 2015. URL: `https://www.tensorflow.org/`.

[47] D. P. Kingma and J. Ba. "Adam: A Method for Stochastic Optimization". In: *ArXiv e-prints* (Dec. 2014). arXiv: `1412.6980 [cs.LG]`.

[48] F. Pedregosa et al. "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.

[49] Serguei Chatrchyan et al. "Search for pair production of third-generation leptoquarks and top squarks in $pp$ collisions at $\sqrt{s} = 7$ TeV". In: *Phys. Rev. Lett.* 110.8 (2013), p. 081801. DOI: `10.1103/PhysRevLett.110.081801`. arXiv: `1210.5629 [hep-ex]`.

[50] Albert M Sirunyan et al. "Search for third-generation scalar leptoquarks and heavy right-handed neutrinos in final states with two tau leptons and two jets in proton-proton collisions at $\sqrt{s} = 13$ TeV". In: *JHEP* 07 (2017), p. 121. DOI: `10.1007/JHEP07(2017)121`. arXiv: `1703.03995 [hep-ex]`.

[51] Javier Duarte et al. "Fast inference of deep neural networks in FPGAs for particle physics". In: *JINST* 13.07 (2018), P07027. DOI: `10.1088/1748-0221/13/07/P07027`. arXiv: `1804.06913 [physics.ins-det]`.

[52] Arjona Martinez, J. et al. "Pileup mitigation at the Large Hadron Collider with graph neural networks." In: *Eur. Phys. J. Plus* 134.7 (2019), p. 333. DOI: `10.1140/epjp/i2019-12710-3`. URL: `https://doi.org/10.1140/epjp/i2019-12710-3`.

[53] Daniele Bertolini et al. “Pileup Per Particle Identification”. In: *JHEP* 10 (2014), p. 059. DOI: 10.1007/JHEP10(2014)059. arXiv: 1407.6013 [hep-ph].

[54] Eric A. Moreno et al. “JEDI-net: A jet identification algorithm based on interaction networks.” In: *Eur. Phys. J. C* 80.1 (2020), p. 58. DOI: 10.1140/epjc/s10052-020-7608-4. arXiv: 1908.05318 [hep-ex].

[55] Eric A. Moreno et al. “Interaction networks for the identification of boosted $H \rightarrow b\overline{b}$ decays.” In: *Phys. Rev. D* 102 (1 2020), p. 012010. DOI: 10.1103/PhysRevD.102.012010.

[56] Cheng Chen et al. “Analysis-specific fast simulation at the LHC with deep Learning.” In: *Comput. Softw. Big Sci.* 5.1 (2021), p. 15. DOI: 10.1007/s41781-021-00060-4.

[57] Wozniak, Kinga Anna et al. “New physics agnostic selections for new physics searches.” In: *EPJ Web Conf.* 245 (2020), p. 06039. DOI: 10.1051/epjconf/202024506039. URL: https://doi.org/10.1051/epjconf/202024506039.

*Chapter 3*

# A NEW ERA OF TIMING DETECTORS

*“Mens et Manus."*
Mind and Hand.

— MIT motto

As the end of the LHC’s run 2 is approaching, CMS and all the other experiments at the LHC are preparing for the major upgrade often referred to as phase-II. This upgrade is going to be crucial not only for the experiment’s current research direction but also for setting the new standards of achievable physics goals. With approximately three years of planned installation and commissioning time without any LHC physics runs, the phase-II upgrade is the second ever big opportunity to implement major technological and systems updates since CMS was originally designed in the early 2000s. Furthermore, this upgrade will crucially prepare the experiment to face the new and more challenging conditions that the High-Luminosity (HL) LHC is expected to create at its interaction points. The HL-LHC is expected to operate at a stable luminosity of $5.0 \times 10^{34}$ cm$^{-2}$s$^{-1}$, yielding 140 pileup collisions by continuously tuning the beam focus and the crossing profile during a fill. However, an even more extreme scenario, with $7.5 \times 10^{34}$ cm$^{-2}$s$^{-1}$ luminosity and 200 pileup collisions per beam crossing, is also being considered for its appeal of delivering 40% more accumulated data. At 140 or 200 pileup collisions, the spatial overlap of tracks and energy deposits from these collisions can degrade the identification and the reconstruction of the produced particles, and can increase the rate of false triggers. These expected conditions of rate and pileup will far exceed the capabilities of the existing CMS detector, which will consequently require significant upgrades to continue to function efficiently. The base and primary goal the phase-II upgrade is then to maintain the excellent performance of the CMS detector in efficiency, resolution, and background rejection for all final state particles and physical quantities used in data analyses.

As one of the major parts of the phase-II upgrade, the CMS collaboration is planning [1, 2] to include the MIP Timing Detector (MTD), a new subdetector with the purpose of measuring the precise time of arrival for charged particles. The MTD will consist of a single-layer device between the CMS tracker and ECAL. It will comprise a barrel and an endcap region, with different technologies based on different performance, radiation, mechanics, and schedule requirements and constraints. The barrel timing layer will cover the pseudorapidity region up to $|\eta| = 1.6$ and will be based on LYSO crystal scintillators read out with silicon photomultipliers (SiPMs)[3, 4]. The endcap region will take over from $|\eta| = 1.6$ to $|\eta| = 3$ to create a hermetic coverage and will be based on Ultra-Fast Silicon Detectors, a planar silicon device based on the Low-Gain Avalanche Detector (LGAD) technology [5, 6]. The MTD will measure timing information for minimum ionizing particles (MIPs) with a 30 ps resolution at the beginning of HL-LHC operation in 2026 and will degrade slowly to 60 ps by the end of HL-LHC operations as a result of radiation damage. The added timing information will help to recover the current CMS performances in the HL-LHC environment and to open the possibility for new research directions. For example, the timing assigned to each track will enable the use of 4D-vertexing, which is expected to render an effective 5-fold pileup reduction by helping assign charged tracks to the correct interaction vertices. Precision timing will also enable new time-based lepton isolation and improved b-tagging algorithms [7].

## 3.1 Identification of Long-lived Charged Particles using Time-Of-Flight

In [8] we study the impact of precision timing detection systems on the LHC experiments' long-lived particle search program during the HL-LHC era. We develop algorithms that allow us to reconstruct the mass of such charged particles and perform particle identification using the time-of-flight measurement. We investigate the reach for benchmark scenarios as a function of the timing resolution and find sensitivity improvement of up to a factor of ten over searches that use ionization energy loss information, depending on the particle's mass.

### Introduction

The CMS and ATLAS collaborations have been exploring precision timing detector concepts intended to enable the time-of-arrival measurement of charged particles with a resolution of a few tens of picoseconds [2, 9]. Such a detector promises to significantly mitigate the impact of the large number of simultaneous collisions within the same bunch crossing (pileup) expected for the High-Luminosity LHC.

The proposed HL-LHC beamspot extends about 10 cm along the beam-axis and about 200 ps in time. For the phase-II CMS tracker, the rate of spurious merging of vertices begins to become significant for vertices separated by less than one mm. For an average number of collisions per bunch-crossing of 140–200, it has been shown that the fraction of spurious tracks from pileup collisions associated with the reconstructed primary vertex is between 20 and 30%. With the addition of time measurement for tracks, this fraction of spurious track-to-vertex association is reduced to about 5% [2]. As a result, significant improvements in the efficiency of particle identification, including isolated leptons and photons, forward jet identification, as well as missing transverse energy resolution recovery, are expected.

Previous studies have demonstrated that significant reach enhancement of the HL-LHC physics program can be realized by using a combination of the time-of-arrival measurement with secondary vertices to reconstruct the mass of long-lived exotic neutral particles [2] produced by proton-proton collisions at the LHC. In this work, we complement those studies by enabling the reconstruction of the masses of long-lived exotic charged particles or heavy stable charged particles (HSCP) by using the position of the production vertex, the time-of-flight, and the track momentum. The resonance mass reconstruction yields a uniquely enhanced capability to identify new particles and to suppress backgrounds for searches of long-lived exotic particles. We discuss methods for reconstructing the time of the primary collision vertex and strategies for using the resulting time-of-arrival measurement to infer the mass of the HSCP. We demonstrate the effectiveness of these methods by evaluating the vertex time resolution and the mass resolution, and we show that an improvement in search sensitivity of a factor of 5–10 can be achieved for HSCP masses above 300 GeV.

**Time-of-Flight Particle Identification**

The time-of-flight $t_{\mathrm{OF}}$ refers to the time needed for a particle to travel between two spatial points. When the length of the particle trajectory ($L$) is known, one can compute the velocity of the particle as $\beta = L/ct_{\mathrm{OF}}$. Combining the latter with the momentum ($\vec{p}$) measurement of the particle, typically obtained by precision tracking detectors at colliders, it is possible to extract the particle mass via the relation

$$\beta = \frac{p}{\sqrt{p^2 + m^2}}.$$

Using TOF measured to a resolution of a few tens of picoseconds and detectors separated by about one meter from the collision point, one can obtain significant discrimination power between different mass hypotheses. Therefore, TOF measurements are a powerful tool for particle identification (PID). We illustrate its use for PID with a simplified example using the geometry of the CMS detector. The CMS detector is cylindrical and we define the $\hat{z}$ axis to be the line going parallel through the center of the cylinder. The total length of the cylindrical detector is equal to $H_d = 6$ m. A time-of-arrival measurement is performed in a cylindrical detector layer located at a radius of $R_d = 1.29$ m and the TOF is extracted using the time of the particle production vertex. The axial magnetic field is $B_z = 3.8$ T. A particle is characterized by its transverse momentum $p_T$, its mass $m$ and its pseudo-rapidity $\eta$ = - $\ln\tan\frac{\theta}{2}$, where $\theta$ is the polar angle measured from the $\hat{z}$ axis. The TOF measurement is assumed to have Gaussian uncertainty with an instrumental resolution of $\sigma_{\text{TOF}}$ and the transverse momentum is measured with a resolution of a few percent.

To estimate the separation power between different mass hypotheses, we compute the mass for which we can achieve separation significance higher than $3\sigma$ ( p-val $< 0.03$ ) from the pion mass ($m_\pi$) hypothesis. The TOF can be expressed, as a function of the particle kinematic variables as:

$$t_{\text{OF}} = \frac{L(p_T, \eta)}{cp}\sqrt{p^2 + m^2}. \tag{3.1}$$

Under the correct mass hypothesis $m = m^*$,

$$t_{\text{OF}}^{(\text{meas})} - t_{\text{OF}}|_{m=m^*} \sim N(0, \sigma_{\text{TOF}})$$

where $t_{\text{OF}}^{(\text{meas})}$ is the measured value, $t_{\text{OF}}|_{m=m^*}$ is the expected value for a mass equal to $m^*$ and $N(\mu, \sigma)$ is a Gaussian distribution with mean $\mu$ and standard deviation $\sigma$. Therefore, all the particles satisfying the relation

$$\frac{L}{c\sigma_{\text{TOF}}}\left|\sqrt{1 + \frac{m^2}{p^2}} - \sqrt{1 + \frac{m_\pi^2}{p^2}}\right| > 3$$

have a $3 - \sigma$ significance separation from the pion hypothesis. In Figure 3.1, we show as a function of the particle $p_T$, the minimum mass that a particle must have to be incompatible with the charged pion (left) and kaon (right) hypothesis, as well as its evolution with $\eta$ and the TOF instrumental resolution $\sigma_{\text{TOF}}$. Three TOF resolution scenarios are considered: 300 ps, corresponding to the best time resolution currently available in CMS for calorimeter deposits; 30 ps, being an ambitious but reasonable

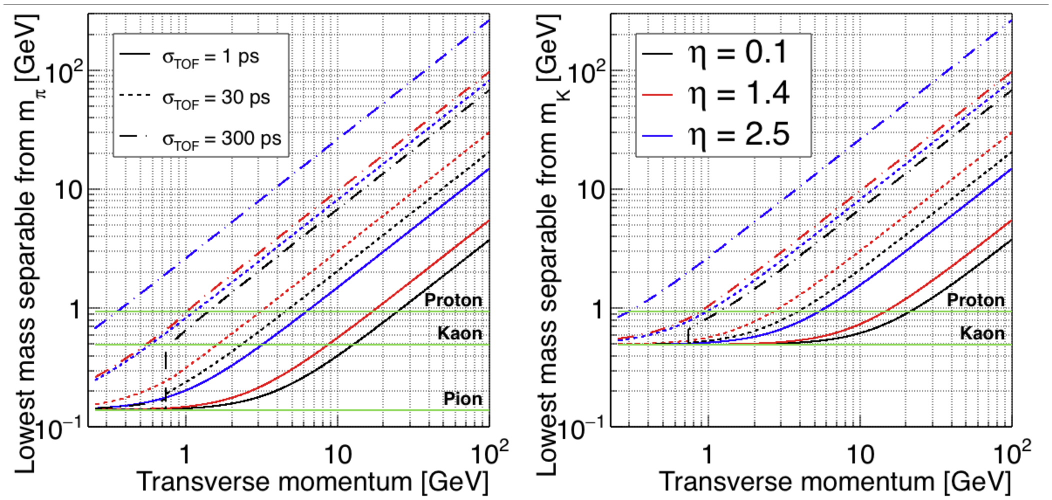


Figure 3.1: Mass of the lighter particle that has at least a $3\sigma$ discrepancy in $t_{\mathrm{OF}}$ from the hypothesis $m = m_\pi$ (left) and $m = m_K$ (right) as a function of the particle transverse momentum. Different colors represents different pseudo-rapidity regions and different styles represent different time resolution scenarios as indicated in the legend.

target for the performance during HL-LHC operation; and 1 ps, with ambitious future technology. The considered pseudo-rapidity values correspond to tracks ending in the central barrel (0.1), near the intersection between barrel and endcap (1.4), and close to the forward limit of the detector acceptance (2.5). In the HL-LHC scenario, with the expected CMS detector performance, separation between pions and kaons is achievable up to about 5-7 GeV in transverse momentum.

Similarly, we can use the TOF measurement to reconstruct the mass resonance of an exotic massive particle. For pair-produced particles at the LHC, its $p_T$ is typically of the same order as its mass. With this assumption, the mass of the particle estimated through Eqn 3.1 using the measured TOF will exhibit a peaked shape reflecting the shape of the particle resonance convoluted with detector resolution effects. This peak structure is a powerful discriminator between a new particle and the background; as the peak becomes narrower, the discrimination power increases. The resolution of the mass resonance can be expressed as:

$$(\Delta m)^2 = m^2 \left[ \left( \frac{\Delta p_T}{p_T} \right)^2 + \left( \frac{1}{1-\beta^2} \right)^2 \left( \frac{\sigma_{\mathrm{TOF}}}{t_{\mathrm{OF}}} \right)^2 \right]. \tag{3.2}$$

In Figure 3.2, we show the expected relative mass resolution as a function of the particle mass, as well as its evolution with $\eta$ and the TOF instrumental resolution

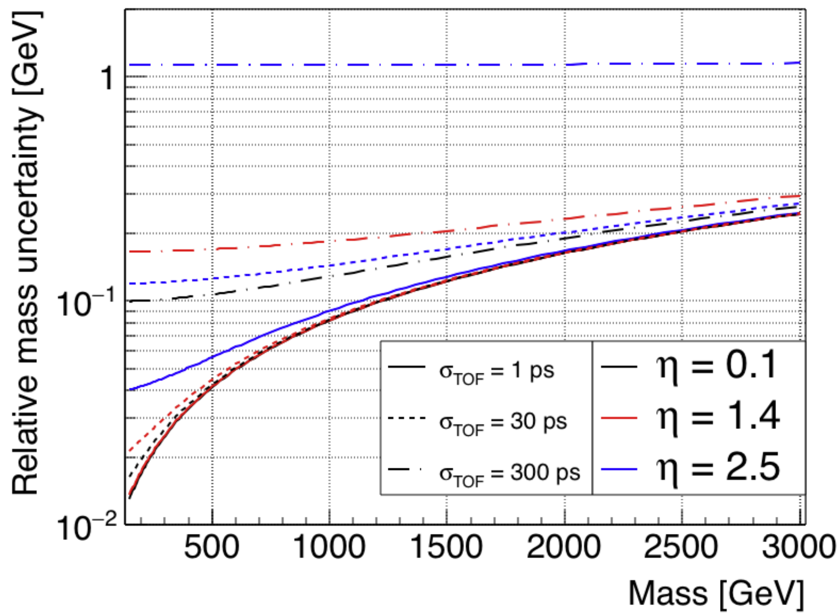


Figure 3.2: Expected mass resolution as a function of mass reconstructed based on the TOF measurement as indicated in the legend, for stable charged particles with a transverse momentum equal to their mass.

$\sigma_{\text{TOF}}$. In the HL-LHC scenario, with the expected CMS detector performance, the estimated resolution on the mass of a 1 TeV particle is about 7% .

**Signal Model and Monte Carlo Simulation**

Many extensions of the standard model (SM) include heavy, long-lived, charged particles [10, 11, 12] that might be produced at the LHC with a speed significantly less than the speed of light. Those with lifetimes greater than a few nanoseconds can travel distances larger than the typical LHC detector and appear stable. These particles are generically referred to as heavy stable charged particles (HSCPs). Precise time measurements can improve the detection of such particles. We consider signals consisting of HSCPs that interact via the strong force and hadronize with SM quarks to form R-hadrons. Similar to reference [13], we consider the pair production of top squarks $\tilde{t}_1$, with masses in the range 100-2500 GeV, generated under the Split SUSY scenario. For the purpose of this study all other masses of the SUSY spectrum are sequestered and considered to be larger than O(10) TeV.

As there are no SM processes that produce a pair of HSCPs, the dominant background at the LHC is QCD multijet production, which can mimic the signature of the HSCPs due to instrumental resolution.

The instrumental resolution is independent of the specific production process and QCD multijet production is dominant because of its large cross section.

For all the Monte Carlo simulation samples, Pythia v8.230 [14], with the 4C tune [15], is used to generate proton proton collisions at 14 TeV. Modeling of multiple interactions and initial and final state radiation are turned on. The signal simulation sample is generated, accounting for all tree-level squark pair-production processes, including gluon-gluon, q-q̄ and q-q initial states. Hadronization to form R-hadrons is activated and all possible R-hadrons are allowed based on the initialized squark masses and the constituent masses of the other partons in the hadron. QCD multijet background samples are generated in the following bins of $\hat{p}_t$, defined as the $p_T$ of the first two partons: 30-50, 50-100, 100-150, 150-185, 185-300, 300-600 and 600-infinity (values expressed in GeV). The pile-up collisions are generated using soft QCD processes which are then mixed with background or signal interaction.

Detector effects are simulated using the Delphes 3.4.1 [16] parametric simulation with a dedicated configuration card used to emulate the performance of the CMS detector after the HL-LHC upgrade. Events are simulated for a scenario with an average pile-up of 140 and the beam-spot is assumed to be Gaussian with $\sigma$ of 160 ps and 5.3 cm for the time and z coordinates, respectively. A negligible beam-spot size in the $x - y$ plane has been assumed, following the operational parameters of the HL-LHC [17]. The tracker performance is parametrized for the geometry of the CMS HL-LHC upgrade design [18]. The timing detector is simulated with Gaussian time-of-arrival resolution with $\sigma$ of 30 ps for all the tracks reaching $R = 1.29$ m with $|\eta| < 3$. This parametrization is consistent with testbeam results [19, 20] and the CMS MIP timing detector technical proposal [2]. We have developed new Delphes modules to implement the mass reconstruction and PID based on the TOF measurement of tracks. A new module to reconstruct vertices simultaneously using both the space and time measurements of tracks was developed, as well as a module to implement the TOF reconstruction. Details of these reconstruction modules are described in the following Section (Section 3.1).

### Space-time Vertex and TOF Reconstruction

To reconstruct the velocity of a particle, it is necessary to measure the time difference and length between two points along its trajectory. The tracker allows a measurement of the track length with a typical relative resolution of order $10^{-3}$. However, only the absolute time-of-arrival of the particle at the point where the particle trajectory intersects the timing detector layer is measured. In order to measure the TOF between the particle production vertex (inner point) and the impact point on the timing detector layer (outer point), the time of the collision that produced the charged particle must be obtained. The naive approach of using the bunch-crossing time as the reference time will yield a resolution not better than 160 ps, due to the spread in time of the collision beam-spot. Instead, we developed a method that precisely reconstructs the space-time coordinate of all vertices. This allows a velocity measurement to be made for all tracks which can be associated to a vertex. Once associated to a vertex, the vertex time is considered as the production time of the particle and it is used to compute the TOF.

**Space-time Vertex Reconstruction** Each track reconstructed in the collision event is clustered together using a deterministic annealing (DA) algorithm [21] to reconstruct collision vertexes. The DA is a clustering algorithm inspired by the observation of annealing processes in physical chemistry that use similar concepts to avoid local minima during the optimization. Certain systems can be driven to their low-energy states by annealing, which is a gradual reduction of temperature, going adiabatically through a series of phase transitions. In the corresponding clustering problem, a Gibbs distribution is defined over the set of all possible configurations which assigns higher probability to configurations of lower energy. The energy is defined as an appropriate function that describes the consistency between cluster centers and the point to be clustered. This distribution is parameterized by a parameter referred to as the "temperature." As the temperature is lowered it becomes more discriminating, concentrating most of the probability in a smaller subset of low-energy configurations. At the limit of low temperature it assigns nonzero probability only to global minimum configurations, corresponding to the output clusters. DA algorithms have been shown to perform well for vertex finding in CMS using the three dimensional spatial coordinates [22]. We have developed a new Delphes module that implements the DA algorithm, extending it to include the time-coordinate. Beside the specific interest for the TOF measurement, the use of the time coordinate in the vertex reconstruction procedure to distinguish tracks from collisions that are very closely separated in space will be crucial at the HL-LHC to maintain the current level of detector performance [2], due to the large amount of pile-up expected.

As shown in [23], for clustering it is convenient to substitute tracks with representative points. We substituted each track with the time ($t_{\text{CA}}$) and the position ($z_{\text{CA}}$) of closest approach to the beam axis. The position of closest approach $z_{\text{CA}}$ is extracted as one of the parameters in the track fit, while the time of closest approach is computed, under a given mass hypothesis $m$, with the following relation:

$$t_{\mathrm{CA}} = t_{\mathrm{TD}} - \int_T \frac{\hat{\beta}}{c\beta} d\vec{x} = t_{\mathrm{TD}} - \frac{L}{c}\frac{\sqrt{p^2+m^2}}{p}$$

where the line integral is computed along the track trajectory ($T$) from the point of closest approach to the location of the Timing Detector (TD), and $t_{\mathrm{TD}}$ is the measured time-of-arrival at the TD. To associate each track with a representative point, it is assumed that all the tracks are from charged pions with mass $m_\pi$. This is a good first order approximation, as the majority of charged particles produced at the LHC are pions or have masses close to the pion mass. For tracks from particles that have significantly different mass, $t_{\mathrm{CA}}$ will be shifted with a magnitude that depends on the particle's momentum and $\eta$. This shift will either result in the track being unsuccessfully clustered, or result in the track being clustered to the wrong vertex. The signal model we are considering in this work produces prompt heavy long-lived particles, so the choice of using the point of closest approach to the beam axis has a negligible effect. For cases where secondary vertexes are important, further development is needed to properly deal with the secondary vertex clustering and time reconstruction.

The deterministic annealing includes a large class of algorithms with many tunable parameters. Only a few parameters are of interest for the application presented in this work, that we discuss in more detail in subsequent sections. Parameters resulting in the choice of the energy function and the covariance matrix, the method chosen to assign tracks to clusters at the end of the cooling, and the choice of the temperatures at which we stop the annealing process are the most crucial ones.

To simplify the notation, the subscript CA will be dropped in the following discussion. The energy between the track $i$ and the vertex prototype $k$ is defined as

$$E_{ik} = \frac{(t_i - t_{\mathrm{V}}^k)^2}{\sigma_{t,i}^2} + \frac{(z_i - z_{\mathrm{V}}^k)^2}{\sigma_{z,i}^2} \tag{3.3}$$

where $\{t_{\mathrm{V}}, z_{\mathrm{V}}\}$ is the prototype position and $\sigma_{z,i}$ ($\sigma_{t,i}$) is the uncertainty on the position (time) of closest approach for track $i$. Using this choice, the typical temperature ($T = 1/\beta$) scale of the vertexes is set to be of $O(1)$. The track partition function is then defined as

$$Z_i = \rho e^{-\beta\mu^2} + \sum_k e^{-\beta E_{ik}}$$

where $\rho$ and $\mu$ are free parameters used to gauge the outlier rejection. The parameter $\mu$ can be approximately interpreted as the number of standard deviations after which a track is called an outlier for a given vertex prototype. In this study, we fixed $\mu$ to 4. The $\rho$ parameter is initially set to 0 and increased to 1 in small incremental steps at the end of the annealing loop. It is crucial to increment $\rho$ slowly in order to activate the outlier rejection quasi-adiabatically.

To penalize particles with high impact parameter, each track is weighted according to:

$$w = \frac{1}{1 + e^{\left(\frac{d_0}{\sigma_{d_0}}\right)^2 - S_0}}$$

where $d_0$ is the reconstructed track impact parameter and $S_0$ is a parameter which determines when the impact parameter weighting becomes important. For our study, we set $S_0$ to be at 1 standard deviation. With the definition of $p(k,i) = p_k e^{-\beta E_{ik}} Z_i^{-1}$ and $p_k = \sum_i w_i p(k,i) / \sum_j w_j$, the vertex prototype time position is computed as

$$t_{\mathrm{v}}^k = \left( \sum_i \frac{p(k,i) w_i}{\sigma_{t,i}^2} t_i \right) \Big/ \left( \sum_i \frac{p(k,i) w_i}{\sigma_{t,i}^2} \right),$$

and $z_{\mathrm{v}}^k$ is computed similarly.

Further defining

$$p(i,k) = \frac{w_i p(k,i)}{p_k}$$
$$w_{xy}^{i,k} = \frac{p(i,k)}{\sigma_{x,i}\sigma_{y,i}}$$

where both $x$ and $y$ can be $t$ or $z$, the vertex covariance matrix used has the form

$$C_{xy}^k = \frac{\sum_i w_{xy}^{i,k} (x_i - x_{\mathrm{v}}^k)(y_i - y_{\mathrm{v}}^k)}{\sum_i w_{xy}^{i,k}}.$$

The annealing cycle starts at $\beta = +\infty$, where the critical temperature for the only prototype ($\beta_0$) is computed. The system is immediately cooled down slightly above $\beta_0$. The annealing loop, set as follows, is run until $\beta_s = 0.2$ is reached:

1. Prototype positions are updated until

$$\max_k \left[ \left( \frac{\Delta t_{\mathrm{v}}^k}{\sigma_{\mathrm{v}}^t} \right)^2 + \left( \frac{\Delta z_{\mathrm{v}}^k}{\sigma_{\mathrm{v}}^z} \right)^2 \right] \leq 0.5$$

where $\Delta$ expresses the variation in the update cycle, and $\sigma_{\mathrm{v}}^{t} = 10\,\mathrm{ps}$ and $\sigma_{\mathrm{v}}^{t} = 0.1$ mm are normalization factors chosen to represent the expected vertex resolution.

2. If two vertices are less than 2 $\sigma$ apart, then they are consdiered not resolvable. Therefore, prototypes with a normalized distance smaller than 2 are merged and the cycle is updated.

3. The temperature is reduced by the cooling factor $f_C = 0.8$. We have observed that this parameter has small impact provided it is of order 1, which keeps the cooling process quasi-adiabatic.

4. Vertices below critical point are split into two along the maximum eigenvalue direction.

At this point, a purging loop is run to remove prototypes with low probability or less than 2 tracks, for which that vertex is the closest. The procedure is repeated, cooling down the system until $\beta_p = 1$. Finally, a ultimate cooling is performed until $\beta_M = 1.5$ to sharpen the cluster border and tracks are assigned to the closer prototype. Figure 3.3 shows a typical space-time configuration at the end of the DA clustering.

The cluster position gives a satisfactory estimate for the vertex position and in this study no further vertex fitting is performed. Tracks that are not assigned to a cluster are then potential candidates for a heavy charged particle.

**Particle Identification** A second Delphes module has been developed to identify potential HSCP candidates and, in general, cluster particles which have not been classified because of the inconsistency of the $m_\pi$ hypothesis assumed at the beginning of the DA. In practice, for each unmatched track with $|d_0|/\sigma_{d_0} < 3$, a two step procedure is followed. Standard Model particles are considered in the following order: pion, kaon, proton, electron and muon. For each mass hypothesis the $t_{\mathrm{CA}}$ is recomputed and the compatibility with the vertices obtained from the DA is tested. The vertices are tested in the order of decreasing $\sum p_T^2$. The track is assigned to the first vertex compatible to within $2\sigma$ and the mass is fixed to that given hypothesis. If no match is found for all of the above particle hypotheses, the track is passed to the second step. The vertex with the highest $\sum p_T^2$ that satisfies spatial compatibility is considered, and the mass which minimizes the closest approach distance from that vertex is estimated as the mass of that charged particle.

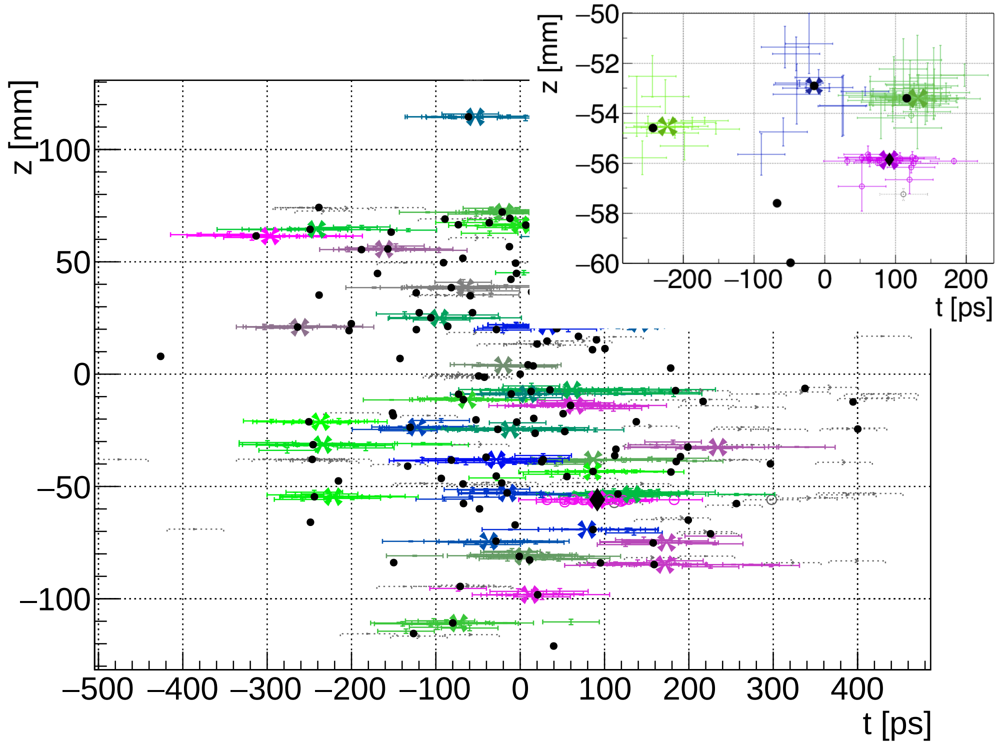


Figure 3.3: The $t - z$ plane at the end of DA clustering for an example $gg \to \tilde{t}_1\bar{\tilde{t}}_1$ event with 140 pileup collisions. Representative points of each track are shown as error bars; the positions of clustered vertices are shown as crosses, and the true position of generated vertices are shown as black circles for pileup collisions, and black diamonds for the signal vertices. Tracks assigned to each vertex are shown using the same color. Tracks assigned to no vertex are shown in light gray.
Top right corner: zoom of the region near the reconstructed signal vertex (purple).

The method developed has been verified to obtain reasonable performance in identifying heavy charged particles. Further details of the performance are discussed in Section 3.1. Future improvements, including better criteria for the mass choice hypothesis or allowing the mass to be a free parameter in the DA clustering, may yield further improved results.

**Vertex Reconstruction and Particle Identification Performance** We evaluate the performance of the vertex reconstruction described above using a sample of simulated signal events with an injected $\tilde{t}_1$ mass of 500 GeV.

In Figure 3.4, we show the difference between the true position simulated in the Monte Carlo generator and reconstructed production point $\hat{z}$ coordinate for each track. Tracks from the collision that produced the top squarks and R-Hadrons are labeled as PV and are shown separately from tracks resulting from pileup interactions. The performance for PV tracks is better compared to the pileup ones

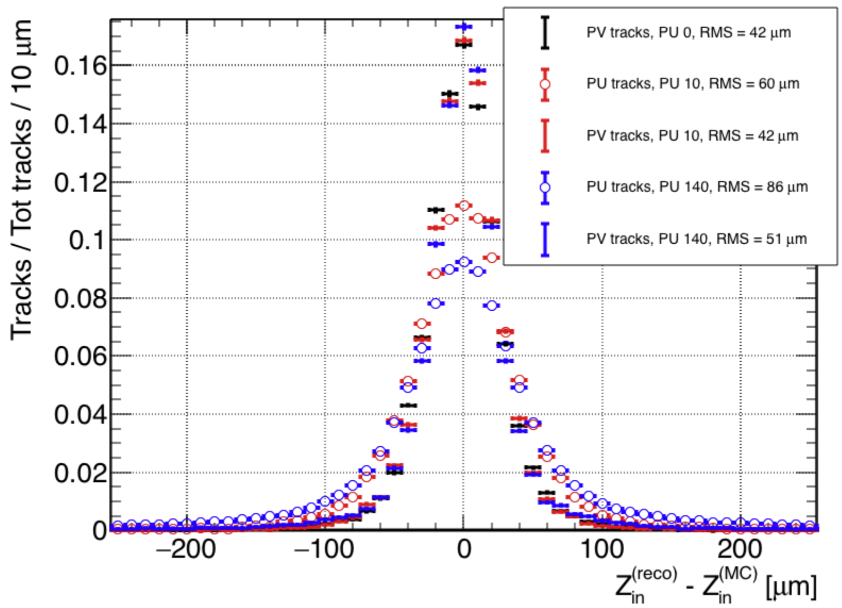


Figure 3.4: The resolution on the track production point (vertex) $\hat{z}$ coordinate obtained from the vertex reconstruction procedure described in Section 3.1 from a sample of signal events is shown. Tracks from $p-p$ interaction which contains the generated signal process (PV) and from pile up interaction (PU) are show separately. Different colors correspond to the resolution for scenarios with different number of PU interaction per bunch crossing.

due to the higher number of tracks and larger transverse momentum. In Figure 3.5, we show the analogous plot for the resolution on the TOF. The TOF resolution has a very small dependence on the amount of pileup and remains around 30 ps even for a scenario with 140 pileup collisions per bunch crossing.

To evaluate the particle identification power of the above resolution performance, we show a two dimensional histogram of the velocity versus the momentum (Figure 3.6) for all the tracks associated to a vertex in the same signal sample used in Figures 3.4 and 3.5. We can observe separation between protons, kaons, and pions for momenta up to a few GeV, in agreement with the discussion from Section 3.1. The reconstructed R-Hadrons are very well separated from the SM particles and lie far outside the boundaries of the displayed plot. Finally, Figure 3.7 shows the mass spectrum reconstructed using the TOF measurement as described in Section 3.1.

Peaks corresponding to the different particles can be clearly seen in the plot: electrons, muons, pions, kaons, protons and R-Hadrons, in order of increasing mass. For a 500 GeV R-hadron we achieve a mass resolution of about 10%.

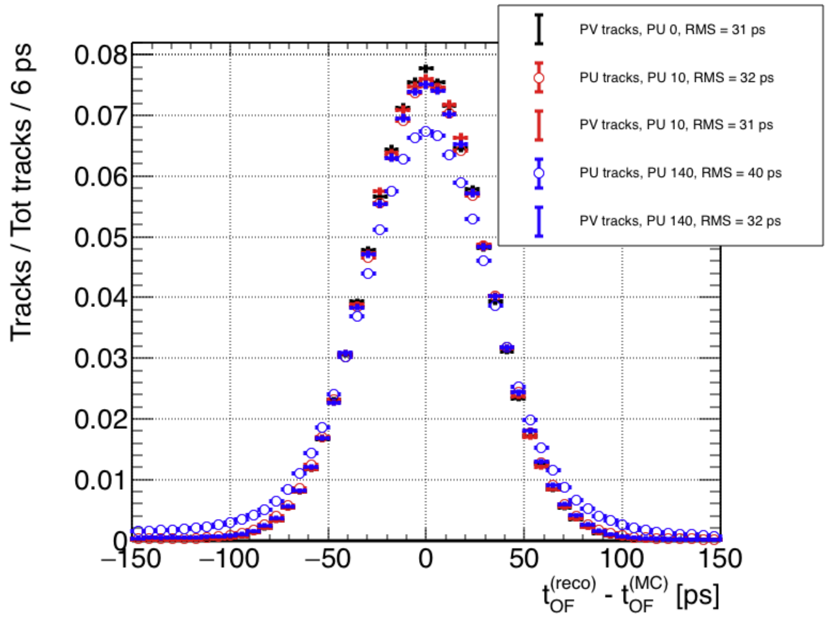


Figure 3.5: The resolution on the TOF from a sample of signal events is shown. Tracks from $p - p$ interaction which contains the generated signal process (PV) and from pile up interaction (PU) are show separately. Different colors correspond to the resolution for scenarios with different number of PU interaction per bunch crossing.

**Benchmark Search for Heavy Stable Charged Particles**

We perform a simple search analysis for heavy R-hadrons using the TOF measurement to illustrate the notable impact that a TOF detector would have at a proton-proton collider such as the HL-LHC. We compare the cross section limits for heavy stable charged particles (HSCPs) obtained with this search with existing limits from CMS and show that significant gains in sensitivity are possible with a new TOF detector.

**Trigger** We consider two benchmark trigger scenarios for the R-hadron search. In the baseline scenario, we employ the proposed CMS L1 track trigger [24] to require large scalar sum of the transverse momentum ($H_T$) of all tracks associated with a particular collision vertex. Based on CMS studies [24], the best estimate for a L1 track $H_T$ trigger with reasonable trigger rates, requires a threshold of 350 GeV. In Figure 3.8, we plot the track $H_T$ spectra for the background on the top panel and top squark signal on the bottom panel for a few different top squark mass points, along with the trigger threshold at 350 GeV. For top squark masses above 500 GeV, the track HT trigger will still retain more than 50% of the signal. However for smaller top

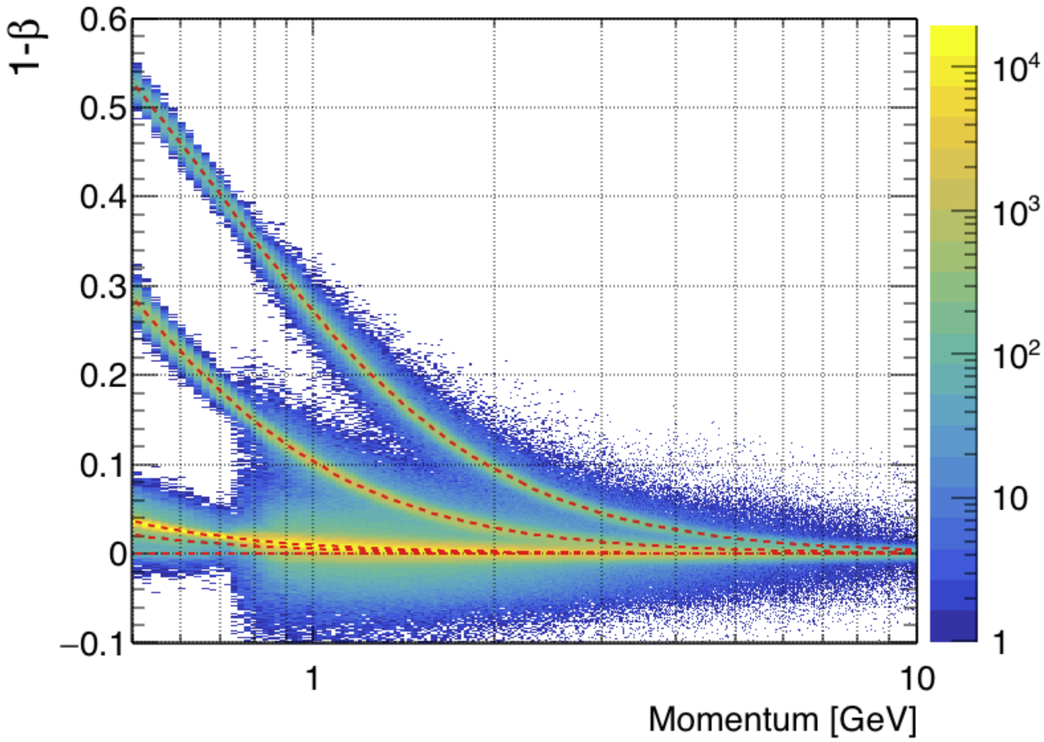


Figure 3.6: The reconstructed velocity ($\beta$) versus the reconstructed momentum for tracks from both PV and PU interactions is shown for a scenario with 140 PU and 30 ps time-of-arrival resolution. Dashed red lines show the analytical relation for different masses: (from top to bottom) proton, kaon, pion, muon and electron mass.

squark masses the signal efficiency decreases significantly. Therefore, we consider a second scenario with the added assumption that the TOF information becomes available in the L1 trigger. Seeded by tracks with $p_T > 10$ GeV, and using a similar procedure as described in Section 3.1, we require that the largest reconstructed track mass based on the TOF measurement is above 10 GeV. Combining this track mass requirement with a less stringent track $H_T$ requirement of $H_T > 150$ GeV, allows us to reduce the background rate to a level similar to the rate of the more stringent $H_T > 350$ trigger (below 150 Hz), while significantly improving the signal efficiency, for top squarks with mass below 200 GeV, from about 20% to above 80%. This specific scenario, as well as more generalized analyses of long-lived particle production [25] demonstrate potential of a TOF-based trigger in the upgraded CMS trigger system.

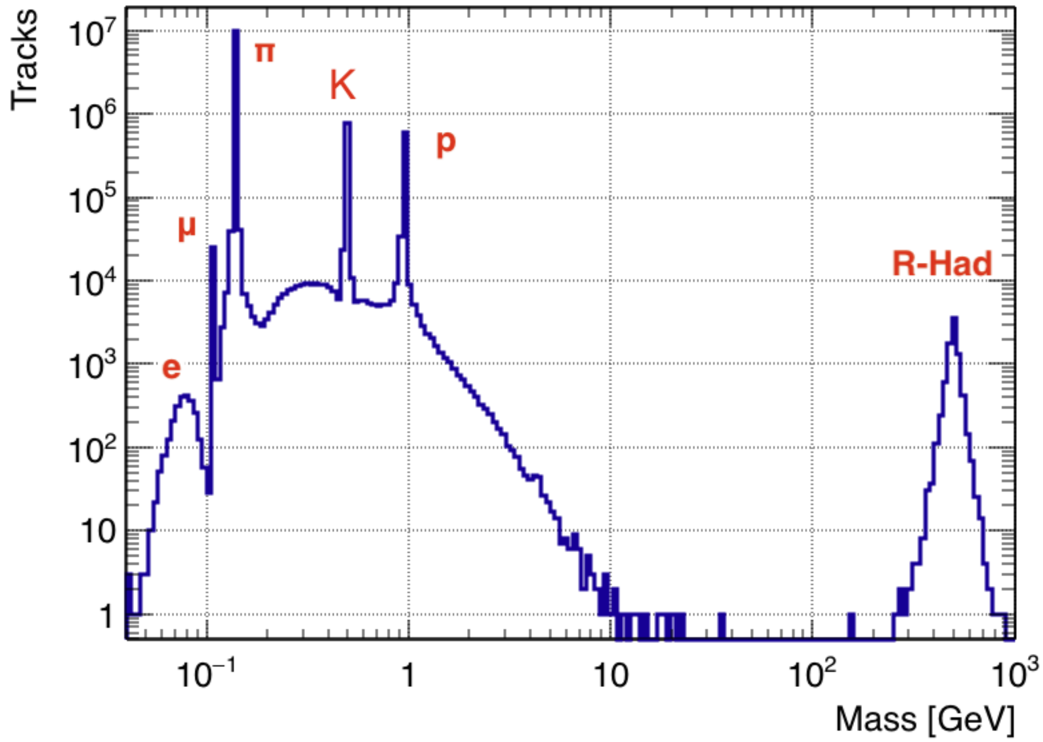


Figure 3.7: Reconstructed mass spectrum from a sample of signal events for a scenario with 140 PU and 30 ps time-of-arrival resolution. Tracks coming from the main collision and also those from pileup collisions are included in the plot.

**Search strategy** We consider a scenario where long-lived top squark pairs are produced and hadronize into stable R-hadrons in the detector volume. Events are split into two categories, one where R-hadrons from both top squarks are detected (the two R-hadron category), and one where an R-hadron from only one of the top squarks is detected (the single R-hadron category). Events are classified into the two R-hadron category if two R-hadron tracks are reconstructed with $p_T$ larger than 50 GeV, and the relative difference in track mass is less than 10%. Otherwise, events are classified into the single R-hadron category if there is one R-hadron track with $p_T > 100$ GeV.

In Figure 3.9 we show the acceptance times selection efficiency for signal events as a function of the generated top squark mass. This efficiency is dominated by the limited detector acceptance, particularly in the forward region. By comparing the efficiency with the baseline scenario that uses the $H_T$ trigger with the scenario where TOF is used in the trigger, we observe clearly that incorporating TOF measurements

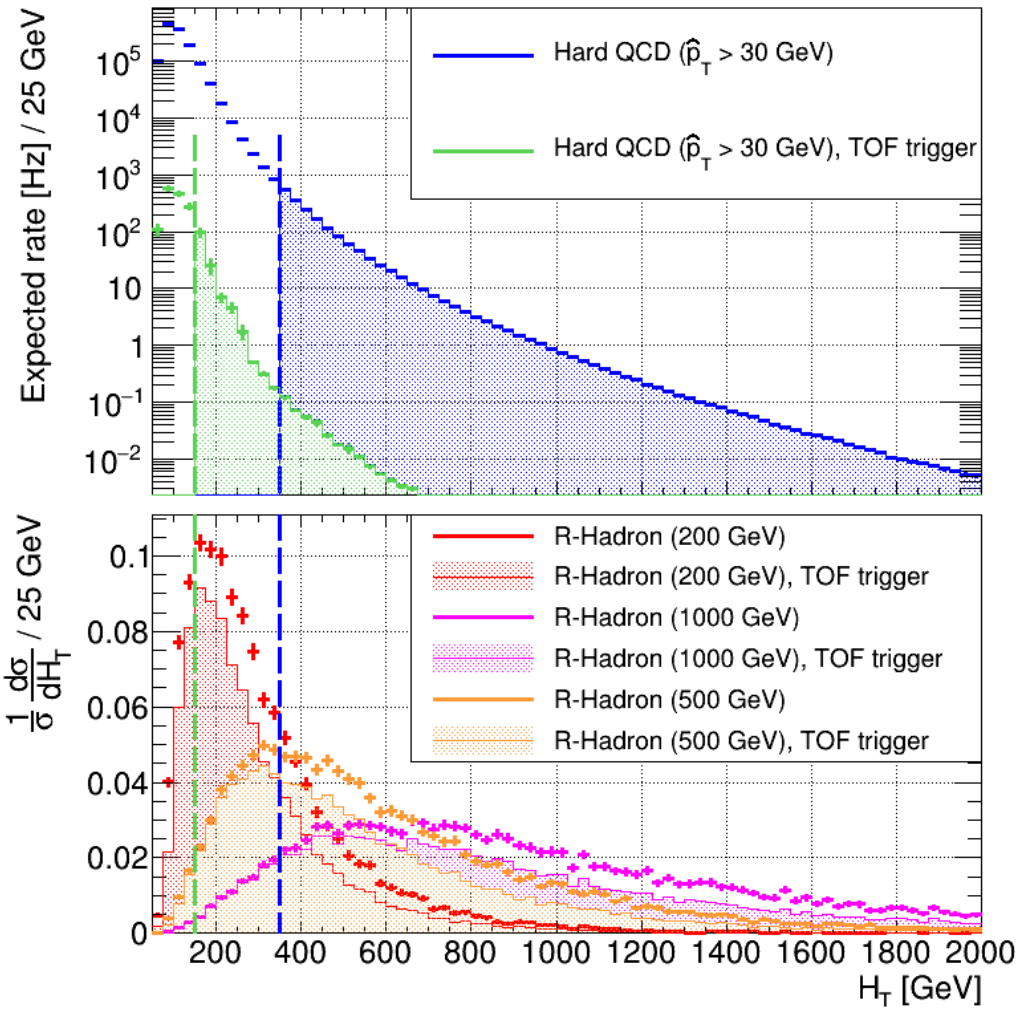


Figure 3.8: Top: Expected rate of events for hard QCD interaction as a function of track $H_T$. The full distribution from all generated events (solid blue), and the distribution after the TOF track mass trigger requirements (solid green), are shown. Bottom: The differential distribution as a function of track $H_T$ for R-hadron production events where at least one R-hadron is within the detector acceptance are shown. The solid line shows the full distribution while the shaded area shows the distribution for events that pass the TOF track mass trigger. Different colors represent different stop mass samples. In both panels, the dashed lines show the track $H_T$ cut for the baseline scenario (blue), and the scenario with the TOF track mass trigger (green).

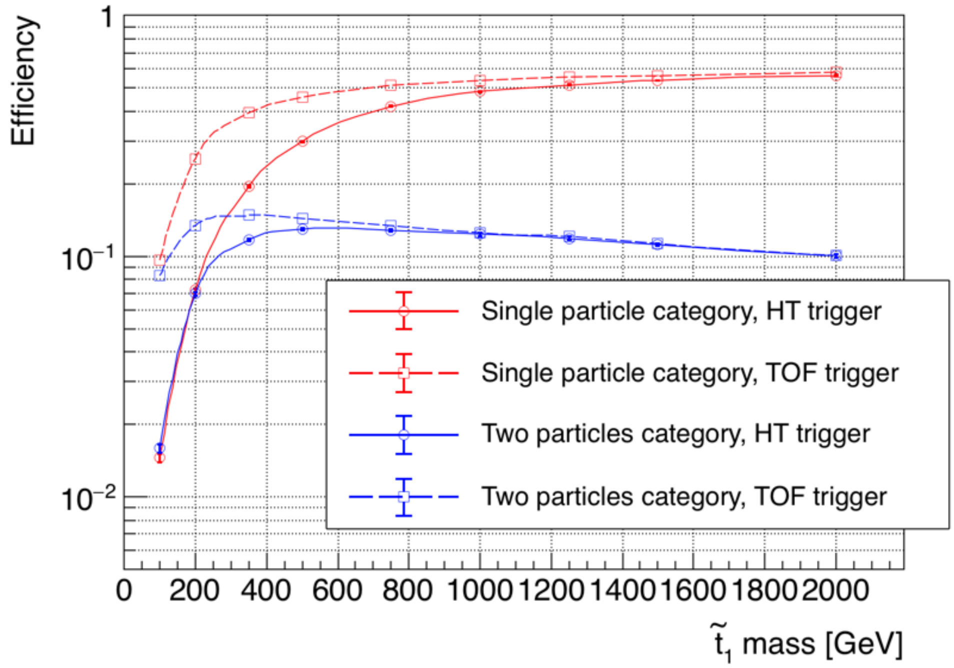


Figure 3.9: Selection efficiency of simulated R-Hadron production events in the two different categories as a function of the generated $\tilde{t}_1$ mass. Efficiency is computed as the ratio of the number of events that pass trigger and category selection over the total number of generated events. Both $H_T$ (solid line) and TOF (dashed line) scenarios are presented.

in the trigger system has a huge impact on enhancing the search sensitivity for top squark masses below 500 GeV, improving the efficiency by up to an order of magnitude.

Finally, in each search category, we perform a fit to the reconstructed R-hadron track mass to extract the signal from the falling background spectrum. For the two R-hadron category, we define $\mathrm{m_h}$ and $\mathrm{m_l}$ as the larger and smaller mass of the two R-hadron tracks, and fit to the average of the two masses. In both categories, only events with $M > 50$ GeV are considered.

**Signal and Background Modeling** The signal mass shape is modeled, in both categories, as a Gaussian with asymmetric exponential tails. The model has a total of four shape parameters that are determined by a fit to the signal Monte Carlo sample: Gaussian mean ($\mu$), Gaussian width ($\sigma$) and the two exponential tail parameters ($\alpha_L$, $\alpha_R$). In Figure 3.10 we show an example of such a fit for a signal with top squark mass of 500 GeV. For masses for which no simulated sample is available, the

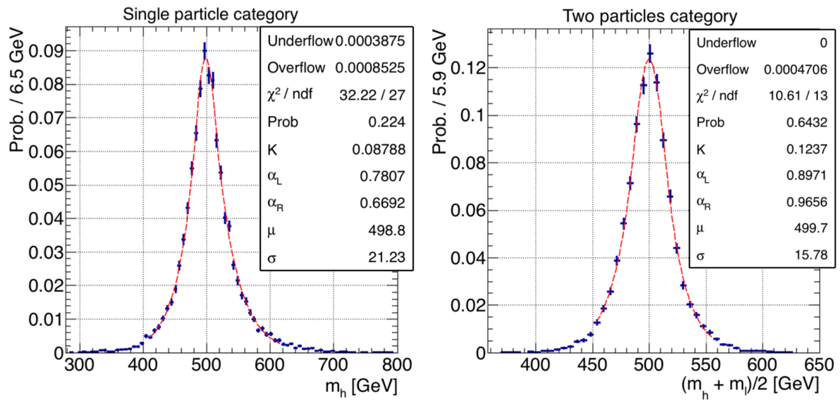


Figure 3.10: Reconstructed mass spectrum in single particle (left) and two particles category (right) for simulated R-Hadron production event assuming $H_T$ trigger. The mass set in Pythia simulation for $\tilde{t}_1$ is 500 GeV. For both categories the dashed red line shows the best fit using a Gaussian function with asymmetrical exponential tails.

value of the Gaussian parameters and the exponential tail parameters are obtained by interpolating between mass points for which simulated samples were generated.

The signal is distinguished from the background through a mass reconstruction based on the time-of-flight of a charged particle. The background is primarily composed of events where the time-of-flight of a charged particle, produced through the standard model QCD multijet production process, has been instrumentally mis-measured. Mis-measurement of the vertex time and the arrival time of the charged particle both contribute, and are dominated by the effect of the time resolution of the time-of-arrival detector. These time mis-measurements result in an exponentially decaying shape for the charged particle mass distribution for the dominant QCD multijet background.

We model the background mass spectrum by fitting an exponentially decaying analytic functional form to Monte Carlo samples of QCD multijet production. For the single R-hadron category, the following functional form is used: $P(M)dM \propto \frac{e^{-M/M^*}}{M}dM$, where $1/M^*$ is the exponential decay parameter extracted from the fit. For the two R-hadron category the following functional form is used: $P(M)dM \propto e^{-M/M^*}dM$. In Figure 3.11, we show an example of the background spectrum and the fitted functional form model for the single R-hadron category in the baseline track $H_T$ trigger scenario. The QCD multijet background Monte Carlo sample is generated in several bins of $\hat{p}_T$ in order to efficiently populate the full track mass spectrum. Considering a luminosity of $L = 12.3$ fb$^1$, the number of events passing

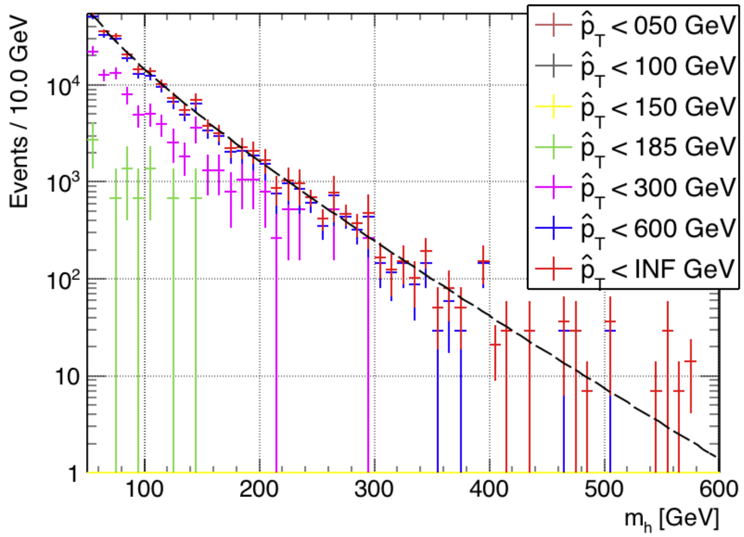


Figure 3.11: Simulated distribution of the observable $M$ (corresponding to $m_h$ in this case) in hard QCD events passing the $H_T$ trigger and the single particle category requirements. The histograms of different colors represent the contribution from different $\hat{p}_t$ bins. The dashed black lines shows the best fit function used to model the background being $M^* = 67.0 \pm 0.3$ GeV the best fit parameter.
The total number of events in the histogram is $2.24 \cdot 10^5$, normalized to a luminosity of $L = 12.3$ fb$^1$.

the cuts is $2.24 \cdot 10^5$ and the best fit parameter is $M^* = 67.0 \pm 0.3$ GeV.

**Results** Based on the signal and background models derived in Section 3.1, we generate pseudo-data for given integrated luminosity and signal cross sections. In Figure 3.12, we show an example of signal and background pseudo-data generated for 12 fb$^{-1}$ of integrated luminosity for proton-proton collisions at a center of mass energy of 14 TeV, and an assumed top squark mass of 500 GeV and production cross section of 100 fb. Fits of the signal and background in the single and two particle categories are performed simultaneously using the models described in Section 3.1.

Using the asymptotic approximation [26] we derive the 95% confidence level expected exclusion limits using the CLs method [27] for long-lived top squark production with lifetime sufficiently large that the top squark is stable over the full detector volume of the CMS detector. The expected limit for 12 fb$^{-1}$ of integrated luminosity is shown in Figure 3.13 and compared to the best existing CMS limits [13]. We show that the sensitivity of this analysis using the TOF measurement is better than limits that do not use timing information for top squark masses above about 170 GeV. The expected limit for our analysis improves more sharply as the top squark mass

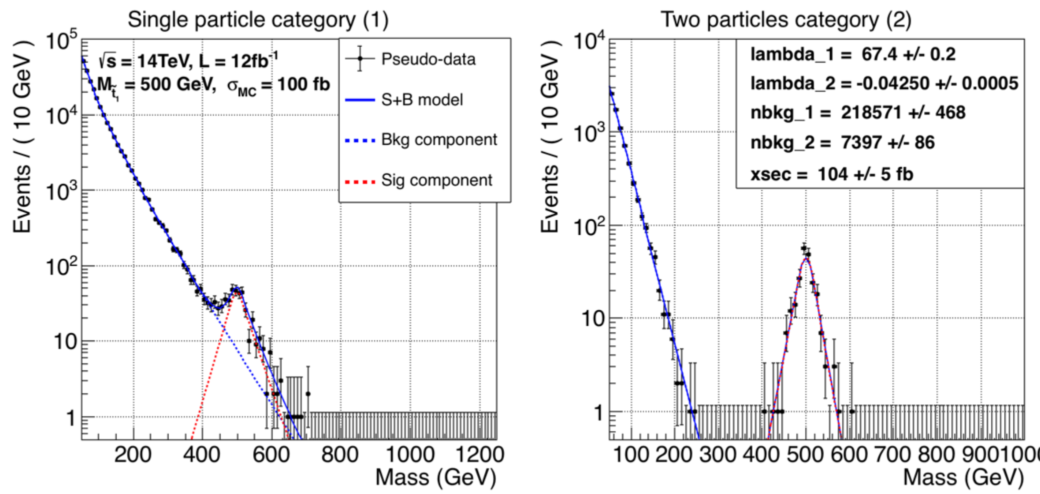


Figure 3.12: Simulated mass spectrum (black dots) in the single particle (left) and two particles (right) categories for an integrated luminosity of 12.3 $fb^{-1}$ and a stop production cross section of 100 fb. Signal (red) and background component (blue) are shown with dashed lines whereas the total fitting function is shown in solid blue. Best fit values of the fit free parameters are shown in the right panel.

increases because the larger mass results in slower velocities and increased time delay, which our analysis is sensitive to, while the sensitivity of the best existing CMS limits are less dependent on the top squark mass. Therefore, the improvement over the existing CMS limits is generally enhanced for larger top squark masses and ranges from a factor of 5 to 10. We also compare the expected limit at 1 ab$^{-1}$ of integrated luminosity between the baseline scenario using the $H_T$ trigger and the scenario where we employ TOF measurements in the trigger, and we observe that at top squark mass below 200 GeV the TOF-based trigger improves the sensitivity by a factor of 2 to 5.

**Summary**

In view of the future proposed timing capabilities of the LHC experiments during HL-HLC, we studied the impact of a time of flight measurement in performing particle identification. We computed the analytical formula for the expected performance and, given the foreseen timing resolution, we estimated the particle identification potential to be significant within SM particles up to a transverse momentum of few GeV. Similarly, we computed the expected peak width in the measurement of a heavy (500 GeV) stable particle mass with the TOF and we found it to be of the order of 10% of the mass.

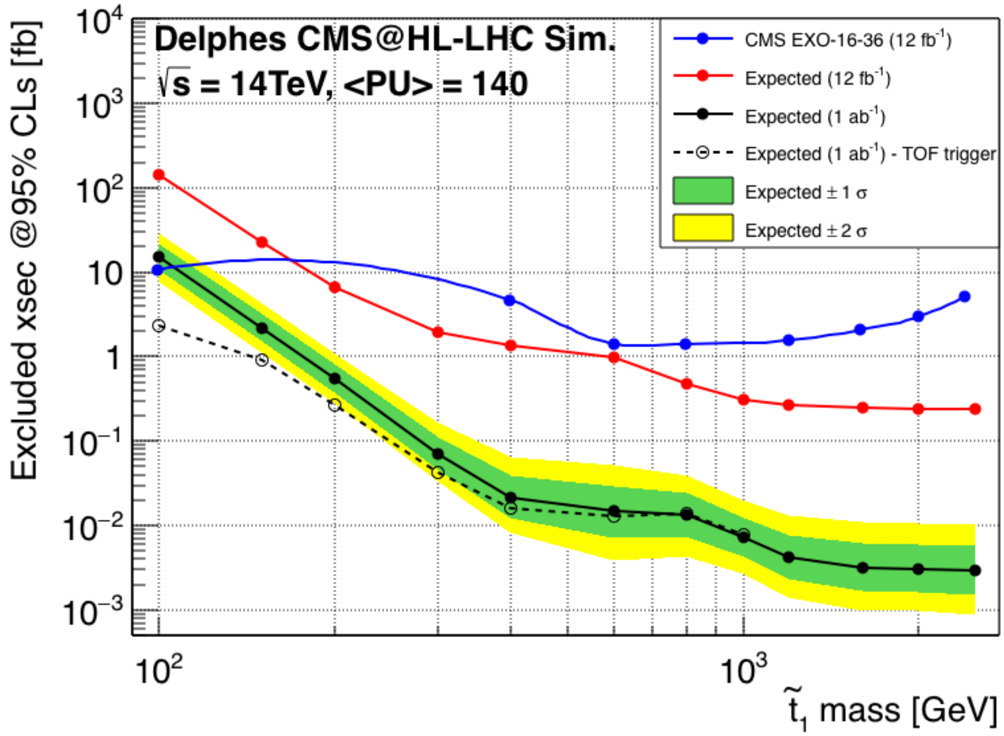


Figure 3.13: Exclusion limits on the production of R-Hadrons at LHC. The improvement in sensitivity using TOF is discussed in the text.

Using Pythia to generate the processes and Delphes to perform a fast simulation of the upgraded CMS detector for HL-LHC, we proposed an approach to perform a TOF measurement with minimal assumptions. Deploying a deterministic annealing to reconstruct vertices, we achieve a $\hat{z}$ resolution of about 50 (80) $\mu$m for PV (PU) tracks for TOF resolution of about 30 ps. These resolutions are demonstrated to be sufficient to identify both SM and BSM particles.

Using long-lived top squark pair production as a benchmark example, we have demonstrated that significant sensitivity gains in searches for long-lived particles can be made with the aid of a dedicated time-of-flight detector. We demonstrate how such a detector would enable four-dimensional vertex reconstruction and the identification of charged particles through its time-of-flight measurement. Mass resonances with good resolution can be reconstructed solely on the basis of the particle track and its time-of-flight, and can significantly enhance the rich program of searches for heavy stable charged particles. We demonstrate for our benchmark example an improvement in sensitivity of a factor of 5 to 10 for top squark masses above 300 GeV. Finally, we show that if the time-of-flight measurement could be utilized in the trigger system, an additional sensitivity improvement of a factor of 2 to 5 could be realized for top squark masses below 200 GeV. This result, along with concurrent complementary studies [25], provide good motivation for further work on the design and realization of a time-of-flight based trigger for long-lived particles.

## 3.2 Development of the Timing Layer Sensors for the CMS Phase-II Upgrade

In order to be ready for the installation at the end of the LHC run 2, the CMS collaboration started to study and develop the design of the MTD almost 10 years in advance. The Caltech Spiropulu group has always been one of the major players in the collaboration effort and, during my Ph.D., I actively participated in the development of the MTD sensors supporting the specific and evolving needs of the project as it was maturing through its phases. The focus of our R&D work has been the Barrel Timing Layer (BTL), the part of MTD that will be placed in the central part of the CMS detector up to $|\eta| < 1.6$. The expected lower radiation in this region w.r.t. the forward region allows for the use of the technology of scintillating LYSO:Ce crystals read out by silicon photomultipliers (SiPMs).

When I first started working on this hardware project, I collaborated with the group effort of developing a simulation of the light yield and timing performance of the BTL sensors [28]. In that work, we simulate different LYSO and SiPM sizes to study the impact of the aspect ratio of SiPM area and LYSO area. Different LYSO and SiPM surface properties, thickness, and LYSO and SiPM upstream (crystal facing the beamline) and downstream (SiPM facing the beamline) configurations are characterized. This study verifies and quantifies the expectations that the performance of the sensors increases as the aspect ratio of SiPM area over crystal area increases, the crystal thickness increases, and the crystal surface is polished. Also, we show that the upstream and downstream configurations are expected to have similar performance.

As a crucial part of the development of the sensor, we carried out an intense testing campaign in the past years aimed at characterizing and finalizing the design of the BTL. For this purpose, we conducted several experiments both at the laboratory on the Caltech campus laboratory (CPTLab) and at the Fermi National Laboratory (Fermilab or FNAL), whose results are discussed in the following section. An in-depth analysis of the thermal studies conducted in the Caltech laboratory is then presented in the last part of this section.

**BTL Sensor Performance in Testbeam Campaigns at FNAL**

Starting from the beginning of March 2018, I joined the effort of developing and testing the MTD sensor prototypes in preparation for the proton beams test (TB) at the FNAL test beam facility. I was the main developer of the code for processing of the pulse shape registered from the oscilloscope attached to the SiPMs. I developed a unified framework to consistently analyze the data acquired with different instruments ranging from the DRS board, the FNAL test beam setup, and the oscilloscope. I also improved the time resolution performances obtained in the analysis introducing a new way of extracting the time stamp from a waveform. This effort was one of the factors which enabled reaching a 30 ps time resolution with SiPMs at the FNAL TB of March 2018 [29, 30]. This framework underpins all the measurements during that time in the Caltech CPTLab and the FNAL TB. The pulse shape observables computed with this framework were used in all the analysis presented then by the Caltech group. Given the joint effort of the TB between BTL and ETL, the same framework and analysis tools developed were also in use for the ETL data, affecting both sensors and ASICS development.

I worked hands-on for preparation of the setup in the CPTLab. I participated in measuring the photon detection efficiency and Light collection efficiencies for different sensor and tile geometry. Aiming to explore the optical coupling between SiPMs and tiles, I was one of the crucial members of the team that developed the gluing procedure and explored the performances with different glues and greases. The results of this have been used by the Caltech group to standardize the tile and SiPMs coupling and wrapping procedure. As an additional byproduct, I also crafted prototypes of mechanical holders with a 3D printer, whose fully developed models were used at the TB.

I played a crucial role in the FNAL TB operations [31, 32]. One of my main contributions has been acting as a first response to several software issues that pops out in extreme prototyping environment like TBs. Always ready to respond, I delivered practical solutions within strict time requirements allowing to always keep the TB focus on the physics. As an example, a severe loss of packets over network transfer started to affect about 40% of the runs, resulting in corrupted binary output files and loss of sync in the trigger-tracker-digitizer system. I took care of this issue within few hours, delivering a recovery of the past and future corrupted runs with over 95% efficiency. More recently, I played a crucial role in improving the data work-flow and process at the FNAL TB facility interfacing the framework discussed above with the DAQ system. I developed a series of automatics bots and prompt loops, which significantly decrease the workload for the operator. The system has been developed with triple redundancy in order to react to systems failures, which indeed has happened and were faced with punctuality. TB data acquisition operations are significantly made easier by this contribution, which brought the number of people on shift required from three to one. As part of this project, I also integrated an online DQM system, which, using the results of the pulse processing and analysis, produces plots of interesting quantities from the latest runs within a few minutes. Those information are automatically published on a website and are available to all the interested parties within a few minutes from the run end. Those plots were used in the FNAL TB control room as a standard tool and allow for faster decision-making than before. A massive reduction of about 30 minutes of time between when the data are acquired and when decisions about operations can be taken has been enabled as a result of this contribution. Both BTL and ETL FNAL TB had a major benefit from this work.

**Cooling Performance Studies for the Barrel Timing Layer**

The sensitive elements of the Barrel Timing Layer (BTL) will need to be refrigerated in order to achieve the expected performances. The noise from the SiPM dark count rate (DCR) is an important factor in the BTL time resolution performance. Thus, it is important to achieve and maintain the lowest possible SiPM temperature with the CMS CO2 cooling infrastructure, to reduce the adverse effect of the DCR. To study the thermal performance of the BTL module, several tests were carried out at the Caltech laboratory. The first set of thermal studies used a mock-up BTL module, based on a resistor, to mimic the SiPM head [33]. The cooling system of CMS was also simulated using external liquid cooling sources operated inside a refrigerated box. This work observed that about a 1 K gradient is present between the cooling system and dummy heat load.

A second and more refined study was conducted in [34], where we performed a series of experiments to access the cooling capabilities of the different design options that were still under discussion as of Winter 2020. Our experiments, conducted on a realistic mock-up crafted in the Caltech Lab, measured the expected temperature difference between the cooling pipe and the sensors array for three different configurations. We found the tall L-shape design to have the best cooling capabilities with the array measured to be 1 K hotter than the pipe in its coolest spot and 2 K hotter than the pipe in its warmer spot.

**Introduction** As of winter 2020/2021, when the work here presented has been carried out, a few parameter choices for the realization of the MTD modules are still open and experimental studies are being conducted in order to finalize a design capable of achieving the target timing resolution of about 30 ps. As discussed in [1], the successful delivery of the appropriate cooling is one of the crucial factors to achieve the desired performance in the full-scale sub-detector when operating in CMS for the full phase II data acquisition period. Indeed, the SiPMs will have to operate at temperatures of around 240K in order to keep the dark current at levels in which the signal-to-noise ratio of the sensors will not impact the time resolution performances. The cooling of the SiPMs sensor arrays will be achieved thought the CMS cooling system based on liquid $CO_2$. Two possible design of the MTD modules cooling arrangement are currently take into consideration (Fig. 3.14). The

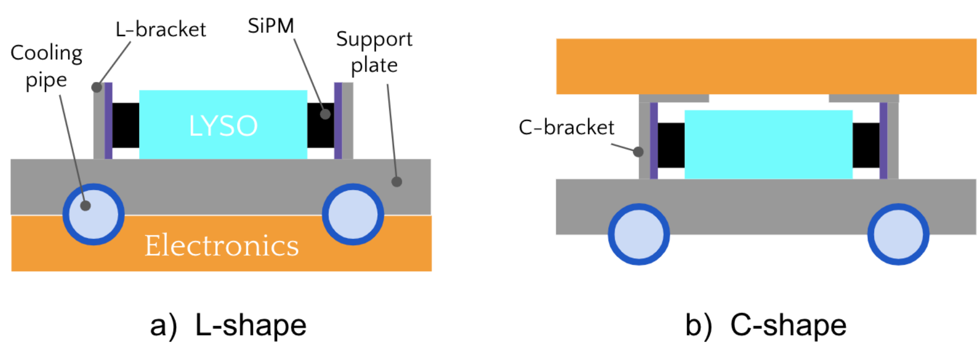


Figure 3.14: L-shaped (left) and C-shaped (right) cooling arrangement designs for the MTD modules. In both cases the support plate, made of metal, is responsible for coupling the cooling pipes to the other elements.

first one, referred to as L-shaped, foreseen the cooling pipe to be placed in the core of the module between the sensitive elements and the electronics. In this design, the SiPMs will be coupled to the cooling pipe through an L-shaped thin metal bracket and they will be thermally decoupled from the electronics. The second one, referred to as C-shaped or Z-shaped, foreseen the sensitive elements to be placed in the core of the module between the cooling pipe and the electronics. In this design, the SiPMs will be coupled to the cooling pipe through a C-shaped thin metal bracket which will share also the cooling load of the electronics. While having a possible impact on the SiPMs cooling, the C-shaped design is interesting because could allow a simpler connection between SiPMs and readout boards.

This study has two main objectives: 1) experimentally estimate the temperature difference between the cooling pipe and the SiPMs array in a regime similar to the expected in CMS close to the MTD end of life; 2) Compare the cooling performances of the two designs looking for regime temperature differences above 1K. The repository for this DN is `https://gitlab.cern.ch/tdr/notes/DN-20-013.git`.

**MTD module mockup** Since no MTD modules have been produced so far, we realized a mockup module based on [1]. To emulate the module support plate we use the 180-10-12C standard liquid cold plates from Wakefield-Vette [35] which embeds a 1 cm diameter U-shaped copper pipe into an aluminum plate. To refrigerate the pipe we connected it to the Fisher Scientific Isotemp 6200 R28 Recirculating Chiller [36] which could reach a lower temperature of 238K. Both the L and the C-shaped brackets used to hold the SiPMs array board into position have been manufactured in the lab starting from a 1 mm thick aluminum plate. The obtained brackets have an 8 mm tall vertical, an 8 mm horizontal profile(s) used to screw it to the support plate (and to support the electronics mockup for C-shaped ones), and are approximately 10 cm long. To increase thermal conductivity we applied a 0.5 mm layer of A12617-25 thermal pad [37] on both the horizontal profile of the bracket facing the support plate and the vertical profile of the bracket facing the SiPMs array.

The foreseen MTD sensors array will be made of 16 SiPMs mounted in line on a ceramic board. In our mockup we use a purple FBK prototype custom board made for the MTD in November 2019 without any SiPMs but with 15 100Ω resistances mounted in series across the SiPMs installation sites (Fig. 3.15). Indeed, for the purpose of this study, the most relevant characteristic of the SiPMs is their power consumption/dissipation which at regime is estimated [1] to be 37.5 mW per SiPM. Moreover, at this moment other factors make the usage of resistors more reasonable: the reduced SiPM prototype availability and cost, the possibility of damaging eventual SiPMs during the test, the easier control of operating conditions due to the different V-I curve, and the sensibility of SiPMs to light. The series of resistors is then supplied with a voltage of 30 V (0.2 A current) through the board connection in order to achieve the same power consumption as if 16 SiPMs at regime were present on the board. The TG-PP10-50 thermal paste [38] is applied on the board

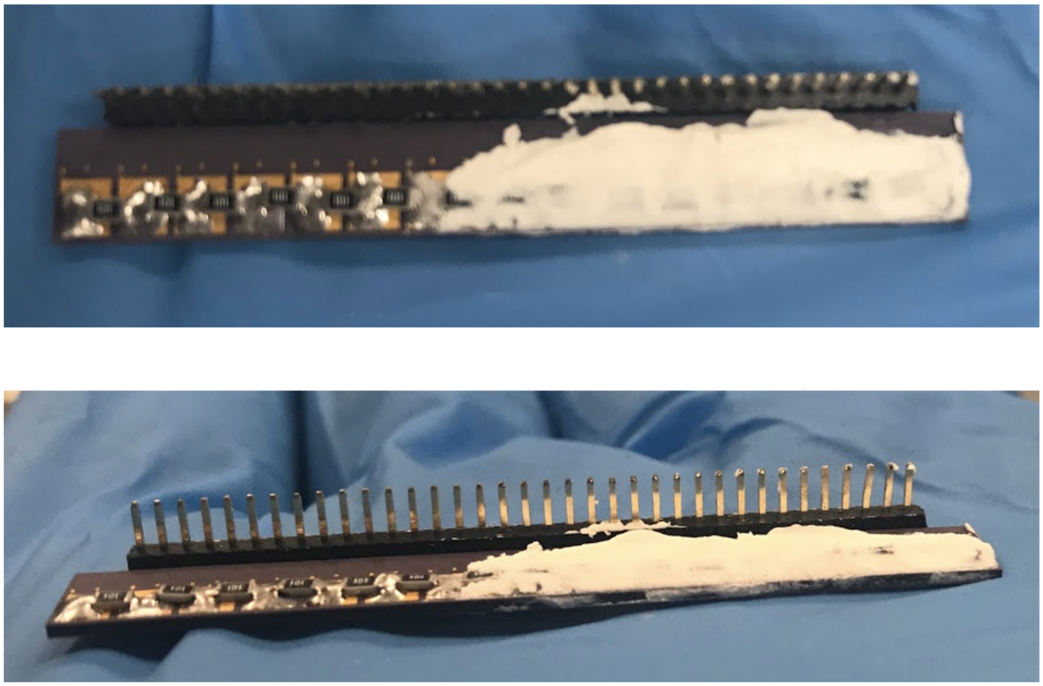

Figure 3.15: Top and side view of the purple FBK prototype board with 15 100Ω resistances mounted in series across the SiPMs installation sites and covered in thermal paste.

face with resistors. Finally, 8 thermistors and an environmental sensor are placed before enclosing the module mockup inside an insulating polystyrene foam box used to thermally decouple the setup from the lab. A full picture of the apparatus is shown in Figure 3.16. When running experiments on the C-shape design, the power consumption of the electronics board is emulated by a 150Ω resistor pad (Fig. 3.17). In order to resemble an electronic board thermal conductivity, the resistor pad is positioned on top of a fiberglass FR4 support plane which is then positioned on the bracket profile with a high conductivity double-sed tape. A total of three different design choices for the brackets are tested in the experimental runs (Fig. 3.18). The first two have an L-shape and are different for the height of the vertical profile: a) have a long vertical profile with the full height of the C-shape profile; b) have a shorter profile of 1.8 mm which ends just below the connectors pins. The last one (c) is the reference C-shape profile.

**Temperature sensors** Temperature measurements are performed using PR502J2 ultra-precision leaded thermistors by Littelfuse [39]. These sensors have been

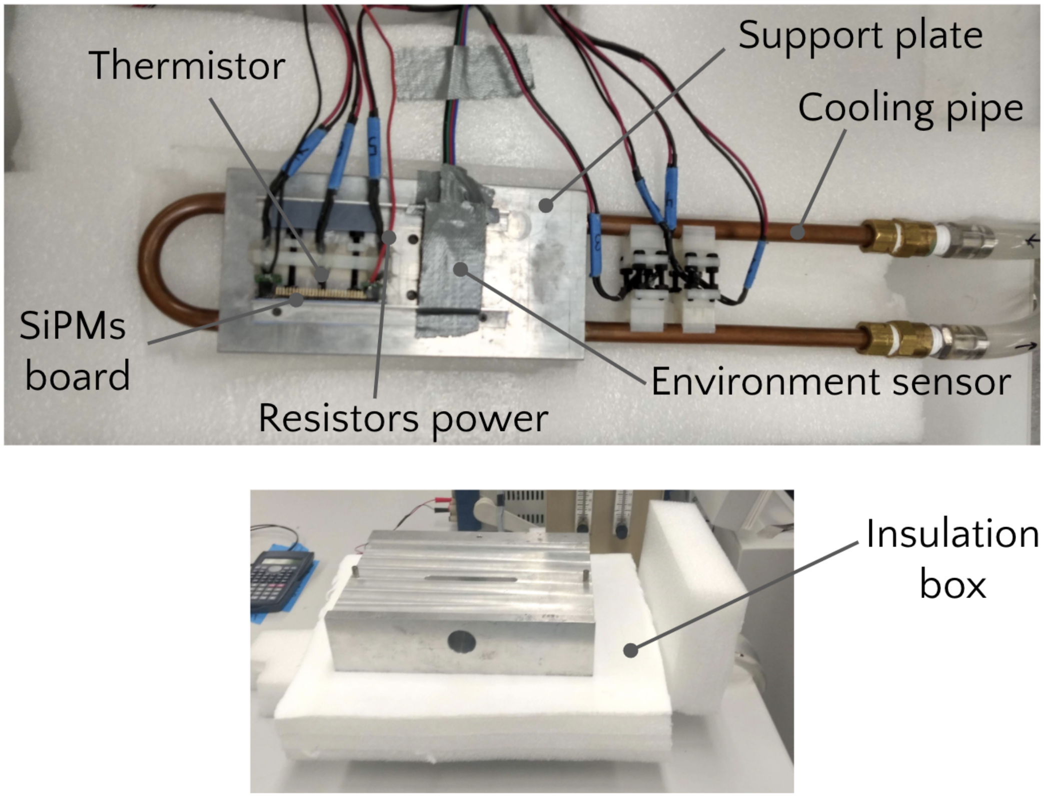


Figure 3.16: Experimental set up before (top) and after (bottom) cloasing the lid of the insultaing box.

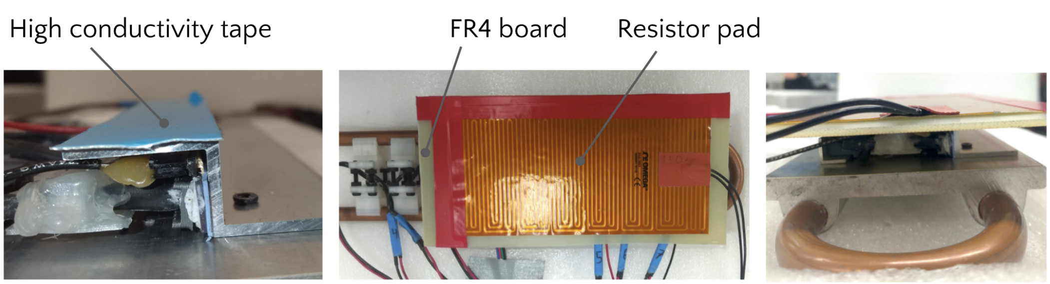


Figure 3.17: Resistor pad on FR4 support used to emulate the electronics effect in the C-shaped design.

chosen not only for their extreme accuracy of ±0.05 K but also for their reduced cost and temperature rating. The thermistors are resistance is measured through a voltage divider circuit (Fig. 3.19). The circuit voltage $V$ is read every 3 seconds by a 12-bit ADC [40] interfaced with a RaspberryPi which saves the value into

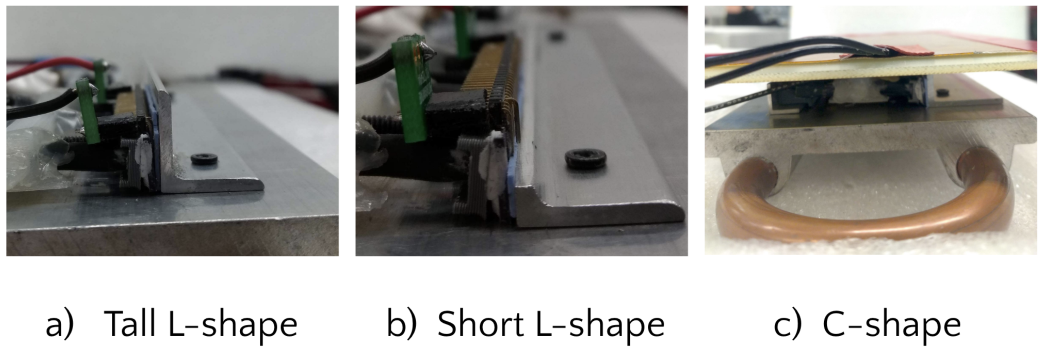


Figure 3.18: The three different brackets tested during the experiments.

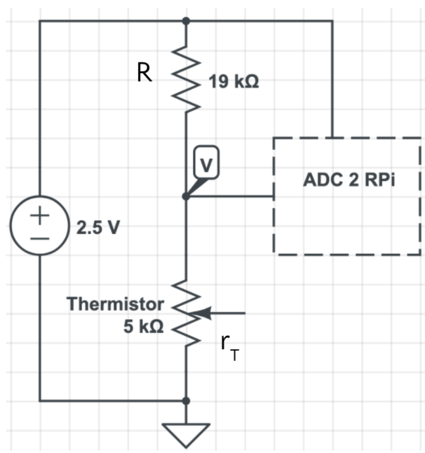


Figure 3.19: Thermistors readout circuit featuring a voltage divider circuit powered with 2.5 V and made by a fixed resistance R = 19 kΩ and the thermistor variable resistance $r_T(T)$ characterized by a value $R_0$ = 5 kΩ.

a dedicated SQL database. The software developed to control data acquisition has been stored in the repository `https://github.com/CaltechPrecisionTiming/CPTLab_ThermalStudies/tree/ocerri_dev` Given the registered ADC counts $C$, the thermistor resistance is extracted by

$$V = \frac{2.5\text{V}}{2^{12} - 1} C \quad \text{and} \quad r_T = \frac{VR}{2.5\text{V} - V}.$$

The thermistor factory accuracy provided is well beyond the resolution needed for this study, hence the measured temperature $T$ can be extracted using the Steinhart–Hart equation [41]

$$\frac{1}{T} = \frac{1}{T_0} + \frac{1}{\beta} \ln\left(\frac{r_T}{R_0}\right).$$

The factory parameters $R_0 = 5$ k$\Omega$, $\beta = -4.4$ %/K and $T_0 = 298K$ are provided with an accuracy better than few percents and hence the temperature measurement uncertainty due to the $V$-to-$T$ conversion is estimated to be smaller than 0.1 K, well beyond the goal sensitivity of this study. The ADC count uncertainty is instead propagated in the temperature estimation. As a control of the procedure we use to extract the temperature value, a thermistor has been placed at different temperatures (as measured by an external thermometer) and the ADC counts have been registered. Figure 3.20 shows how the control measurements so obtained are in agreement with

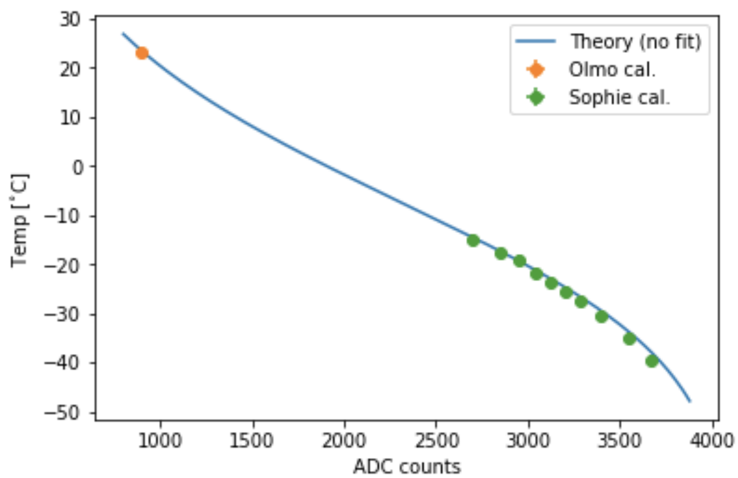


Figure 3.20: Comparison between the theoretical curve used to derive the temperature value in the following and the control measurements.

the theoretical curve that will be used in the following to derive the temperature given the ADC counts.

We monitored the environment inside the polystyrene foam box with a CC2D35-SIP humidity and temperature sensor [42]. This sensor is suspended above the support plate with adhesive tape. While the environment air temperature measurement is relatively interesting, the humidity is monitored to make sure to stay above the air dew point. The latter is crucial to avoid possible condensation which could impact the thermistor's readout. We used 8 thermistors to measure the temperature in localized spots. During preliminary tests, we observed that the accuracy and reproducibility of the temperature measurements are heavily compromised if the sensor is not carefully positioned in good contact with the desired measuring spot. For this reason, we designed and crafted with a 3D printer a set of custom plastic pieces to keep the thermistor in the correct position with a light but sizable pressure. Moreover, we applied abundant thermal paste on each thermistor tip to further enhance the thermal coupling. A total of 4 thermistors is used to monitor the temperature of the cooling pipe (Fig. 3.21). Two of these four thermistors (referred

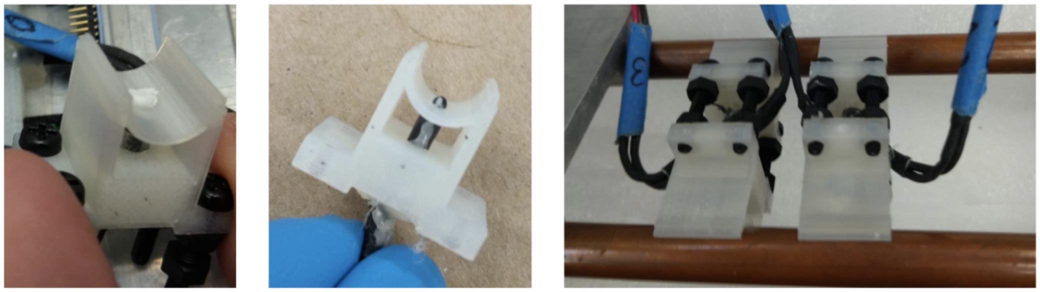

Figure 3.21: Positioning of the 4 thermistors used to monitor the pipe temperature.

to as *pipe in 0/1*) are placed on the pipe before the support plate and the other two after (referred to as *pipe out 0/1*). With this configuration not only we can monitor eventual temperature gradients along the pipe but also introduce a redundancy to make the results more robust to systematic errors due to thermistor displacing and malfunctioning. One thermistor (referred to as *Ext. pipe*) is placed on the chiller exit connector to monitor eventual oscillation due to the chiller feedback loop. The temperature measured by this thermistor will not represent an accurate measurement of the coolant for which is used the chiller internal thermometer (0.5 K uncertainty). The last three thermistor are placed on the purple FBK board against the resistors (Fig. 3.22). By means of custom supports, these three thermistors are placed at

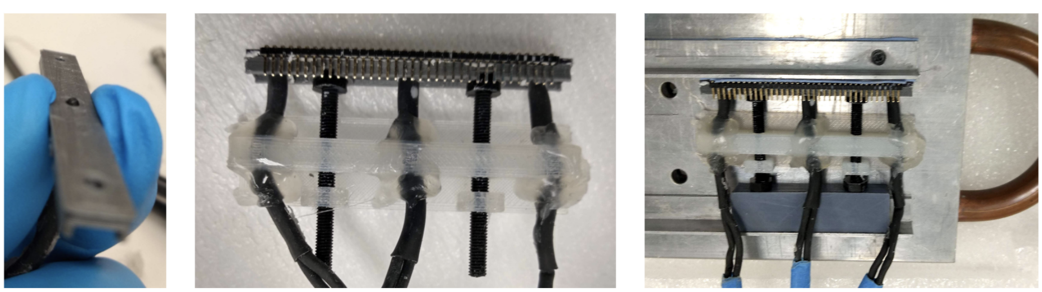

Figure 3.22: Positioning of the 3 thermistors used to measure the resistors temperature.

mid-height of the purple board on top of the resistors at three different positions along the array. One (*pkg R border*) is placed on the right of the board close to the bracket end, one (*pkg C*) in the center, and one (*pkg L*) on the left of the board close to the bracket core.

**Experimental runs** We acquired the temperature data in several experimental runs with different designs, power input, and cooling circuit temperature. After positioning the thermistors we acquired 5 test runs at room temperature to check the stability of the measurement (Fig. 3.23). In these runs, at equilibrium all the sensors where found to be in agreement better than 0.5 K with a consistency of the single sensor across the runs better than 0.1 K.

In the reasonable assumptions that at the operating temperature the heat exchanged with the surrounding air is negligible and that the material properties do not change significantly for temperatures close to the operating one, the temperature difference between the pipe and the resistors array does not depend on the temperature at which the experiment is performed. We decided to run our experiments at a temperature of 288 K. With this choice, not only do the experiments have an equilibrium time of the order of a few minutes but also the temperature is such that the insulation provided by the polystyrene foam box is sufficient and the dew point of the air is not reached. In all our experiments the humidity measured by the environmental sensor has never been observed to go above the value of 60%. For each experimental run with the different conditions we register the temperature measured by the 8 thermistors and the environmental sensor for about 20-30 minutes (Fig. 3.24).

At the beginning of each run, both the resistors (mocking the SiPMs) and the resistor pad (when present) are powered off in order to check the compatibility of the system with the initial conditions imposed by the cooling. When all cooling gradients vanish, the power to the resistors array is turned on: the new equilibrium state is reached within 2-3 minutes. When the resistor pad is present, it is turned on only after the first equilibrium is reached in order to be able to cross-check the effect of both power sources. After all the relevant sources are turned on, a stable conditions window is defined by requiring that the maximum of the average temperature drift among the thermistors placed on the resistors array is smaller than 0.05 K/min. This region, represented in Fig. 3.24 by the two magenta dashed lines, is used to perform the temperature measurements. After the stable conditions have been registered for a few minutes, all the sources are again powered off and the run is interrupted after the initial state dominated by the cooling is reached once again. As exemplified in fig. 3.24, the behavior of the temperature data registered by all thermistors is in agreement with the procedure and expectations explained above: the thermistors placed on the pipe register no variations; the thermistors placed on the resistors array follow the expected turn on/turn off pattern; and the thermistor placed on the chiller pipe (T2) shows small oscillations (probably due to the chiller fan) which have no impact on the components of the system. The data registered in each run are then processed offline. For each thermistor $i$, the data from the stable window are fitted with a first-order polynomial $T_i = \hat{T}_i + s_i(t - t_0)$ where $\hat{T}_i$ is the parameter that will be used as the temperature estimation, $s_i$ could be interpreted as the residual drift and $t_0$ is the central time of the stable conditions window (i.e. the average between the time of the two magenta dashed lines). All the fit have been found to have a p-value better than 5% and $|s_i| < 0.01$ K/min. Since no significant difference is observed between the temperature measurement of the four thermistors located on the cooling pipe, their value is averaged to obtain a reference temperature for the run which is then subtracted from all the other ones.

The results of the temperature difference ($\Delta T$) between the cooling pipe and the different parts of the resistors array in the various configurations are shown in Fig. 3.25. While the power injected in the resistor array is equivalent to 37.5 mW per SiPM, the picture summarizes the measured ($\Delta T$) as a function of the power dissipated by the resistor pad. The measurements taken with the L-shapes are displaced from 0 on the x-axis only for a more clear presentation but no resistor pad was present during those runs.

**Conclusions** In this study, we performed a series of experiments to measure the expected temperature difference between the cooling pipe and the SiPMs sensors in the MTD module. Using a mockup module made in the Caltech Lab we reproduced the operating characteristics of the detector and we used 8 thermistors to simultaneously measure the temperature across the system. When the power injected in the array is equivalent to 37.5 mW per SiPM, we measure a $\Delta T$ between 1 and 4 Kelvin degrees. The design with a tall L-shape showed the best cooling capabilities with the array measured to be 1 K hotter than the pipe in its coolest spot and 2 K hotter than the pipe in its warmer spot. The results for the design with a short L-shape are found to be 15-20% worse than the tall L-shape. Differently, the C-shape is found to underperform the other two designs when a non-negligible amount of power is injected in the electronics mockup. We measured a temperature increase of about 0.6 K/W of power injected across all the arrays. Finally, for all the designs we observe a temperature difference of about 1 K across the array. Where the warmer point is the extreme of the array placed at the edge of the bracket and the coolest point being the other edge of the array.

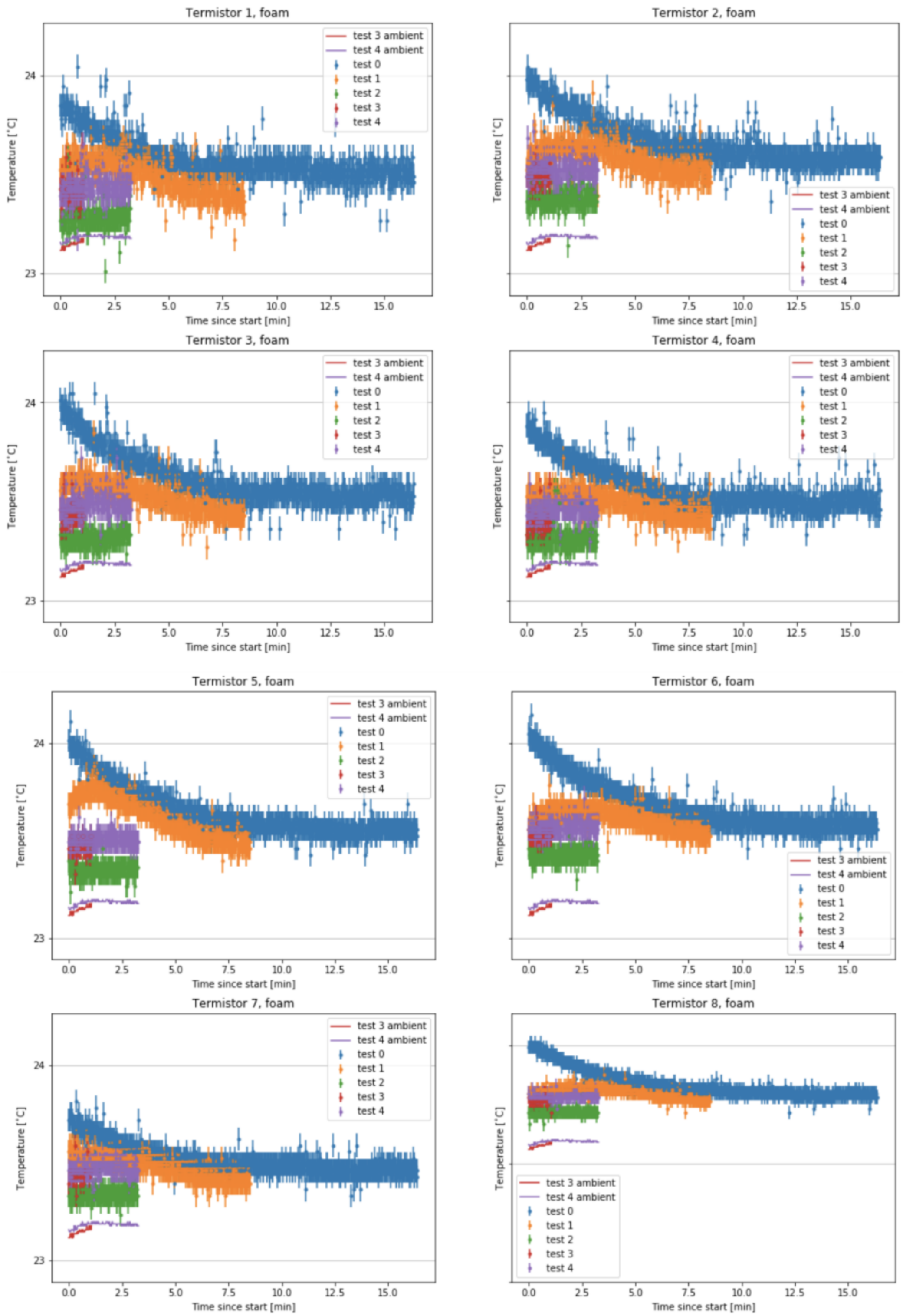


Figure 3.23: Temperature measurement for the 8 thermistors during the room temperature test runs.

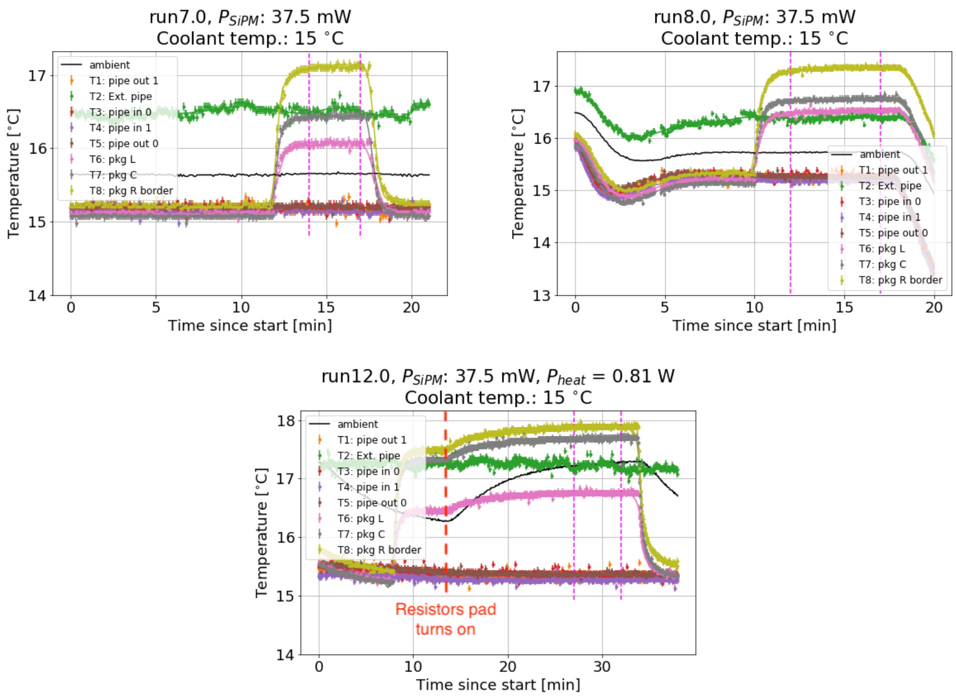


Figure 3.24: Example of experimental runs taken in January 2020. Run 7 (top), run 8 (center) and run 12 (bottom) corresponds respectively to the tall L-shape, the short L-shape, and the C-shape design. The equilibrium window used to average the temperatures is shown in dashed magenta lines.

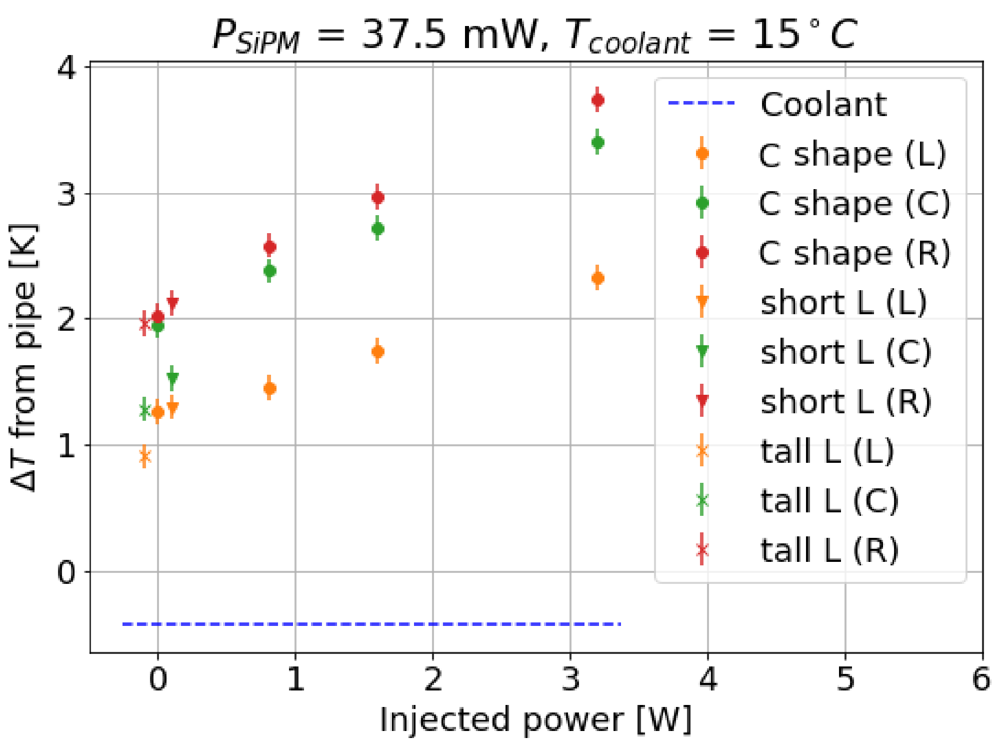


Figure 3.25: Temperature difference between the cooling pipe and the different parts of the resistors array in the various configurations. Injected power refers to the power dissipated by the electronics mockup.

## References


[1] Collaboration CMS. *A MIP Timing Detector for the CMS Phase-2 Upgrade*. Tech. rep. CERN-LHCC-2019-003. CMS-TDR-020. Geneva: CERN, 2019. URL: `http://cds.cern.ch/record/2667167`.

[2] *Technical Proposal for a MIP Timing Detector in the CMS Experiment Phase 2 Upgrade*. Tech. rep. CERN-LHCC-2017-027. LHCC-P-009. Geneva: CERN, 2017. URL: `https://cds.cern.ch/record/2296612`.

[3] S Gundacker et al. "Time of flight positron emission tomography towards 100ps resolution with L(Y)SO: an experimental and theoretical analysis". In: *Journal of Instrumentation* 8.07 (2013), P07014–P07014. DOI: `10.1088/1748-0221/8/07/p07014`. URL: `https://doi.org/10.1088/1748-0221/8/07/p07014`.

[4] D. Anderson et al. "On timing properties of LYSO-based calorimeters". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 794 (2015), pp. 7–14. ISSN: 0168-9002. DOI: `https://doi.org/10.1016/j.nima.2015.04.013`. URL: `https://www.sciencedirect.com/science/article/pii/S0168900215004829`.

[5] G. Pellegrini et al. "Technology developments and first measurements of Low Gain Avalanche Detectors (LGAD) for high energy physics applications". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 765 (2014). HSTD-9 2013 - Proceedings of the 9th International "Hiroshima" Symposium on Development and Application of Semiconductor Tracking Detectors, pp. 12–16. ISSN: 0168-9002. DOI: `https://doi.org/10.1016/j.nima.2014.06.008`. URL: `https://www.sciencedirect.com/science/article/pii/S0168900214007128`.

[6] N. Cartiglia et al. "Design optimization of ultra-fast silicon detectors". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 796 (2015). Proceedings of the 10th International Conference on Radiation Effects on Semiconductor Materials Detectors and Devices, pp. 141–148. ISSN: 0168-9002. DOI: `https://doi.org/10.1016/j.nima.2015.04.025`. URL: `https://www.sciencedirect.com/science/article/pii/S0168900215004982`.

[7] Olmo Cerri. *CMS precision timing physics impact for the HL-LHC upgrade*. Tech. rep. 2018. arXiv: `1810.00860`. URL: `http://cds.cern.ch/record/2641474`.

[8] Olmo Cerri et al. "Identification of long-lived charged particles using time-of-flight systems at the upgraded LHC detectors." In: *J. High Energy Phys.* 2019.4 (2019), p. 37. DOI: `10.1007/JHEP04(2019)037`.

[9] *A High-Granularity Timing Detector (HGTD) in ATLAS : Performance at the HL-LHC*. Tech. rep. ATL-LARG-PROC-2018-003. Geneva: CERN, 2018. URL: https://cds.cern.ch/record/2302827.

[10] Christian W. Bauer et al. “Supermodels for early LHC”. In: *Physics Letters B* 690 (2010), pp. 280–288. ISSN: 0370-2693. DOI: https://doi.org/10.1016/j.physletb.2010.05.032. URL: http://www.sciencedirect.com/science/article/pii/S0370269310006246.

[11] M. Fairbairn et al. “Stable massive particles at colliders”. In: *Physics Reports* 438 (2007), pp. 1–63. ISSN: 0370-1573. DOI: https://doi.org/10.1016/j.physrep.2006.10.002. URL: http://www.sciencedirect.com/science/article/pii/S0370157306004054.

[12] M. Drees and X. Tata. “Signals for heavy exotics at hadron colliders and supercolliders”. In: *Physics Letters B* 252 (1990), pp. 695–702. ISSN: 0370-2693. DOI: https://doi.org/10.1016/0370-2693(90)90508-4. URL: http://www.sciencedirect.com/science/article/pii/0370269390905084.

[13] *Search for heavy stable charged particles with* 12.9 $\mathrm{fb}^{-1}$ *of 2016 data*. Tech. rep. CMS-PAS-EXO-16-036. Geneva: CERN, 2016. URL: https://cds.cern.ch/record/2205281.

[14] Torbjörn Sjöstrand et al. “An introduction to PYTHIA 8.2”. In: *Computer Physics Communications* 191 (2015), pp. 159–177.

[15] Richard Corke and Torbjorn Sjostrand. “Interleaved Parton Showers and Tuning Prospects”. In: *JHEP* 03 (2011), p. 032. DOI: 10.1007/JHEP03(2011)032.

[16] J. de Favereau et al. “DELPHES 3, A modular framework for fast simulation of a generic collider experiment”. In: *JHEP* 02 (2014), p. 057. DOI: 10.1007/JHEP02(2014)057.

[17] G. Apollinari et al. “High Luminosity Large Hadron Collider HL-LHC”. In: *CERN Yellow Report* (2015), pp. 1–19. DOI: 10.5170/CERN-2015-005.1. arXiv: 1705.08830 [physics.acc-ph].

[18] S. Mersi et al. “Software package for the characterization of Tracker layouts”. In: *Proceedings, 13th ICATPP Conference on Astroparticle, Particle, Space Physics and Detectors for Physics Applications (ICATPP 2011): Como, Italy, October 3-7, 2011*. 2012, pp. 1015–1024. DOI: 10.1142/9789814405072_0153.

[19] A. Apresyan et al. “Test beam studies of silicon timing for use in calorimetry”. In: *Nucl. Instrum. Meth.* A825 (2016), pp. 62–68. DOI: 10.1016/j.nima.2016.04.031.

[20] A. Apresyan et al. "Studies of uniformity of 50 $\mu$m low-gain avalanche detectors at the Fermilab test beam". In: *Nucl. Instrum. Meth.* A895 (2018), pp. 158–172. DOI: 10.1016/j.nima.2018.03.074.

[21] K. Rose. "Deterministic annealing for clustering, compression, classification, regression, and related optimization problems". In: *Proceedings of the IEEE* 86.11 (1998), pp. 2210–2239. ISSN: 0018-9219. DOI: 10.1109/5.726788.

[22] E. Chabanat and N. Estre. "Deterministic annealing for vertex finding at CMS". In: *Proceedings of CHEP 04* 11 (2004), pp. 287–290.

[23] Rudolf Frühwirth et al. "New developments in vertex reconstruction for CMS". In: *Nucl. Instrum. Meth.* A502 (2003), pp. 699–701.

[24] D. Contardo et al. *Technical Proposal for the Phase-II Upgrade of the CMS Detector*. Tech. rep. CERN-LHCC-2015-010, LHCC-P-008, CMS-TDR-15-02. 2015.

[25] Jia Liu, Zhen Liu, and Lian-Tao Wang. "Long-lived particles at the LHC: catching them in time". In: (2018). arXiv: 1805.05957 [hep-ph].

[26] Glen Cowan et al. "Asymptotic formulae for likelihood-based tests of new physics". In: *The European Physical Journal C* 71 (2011), p. 1554.

[27] Alexander L. Read. "Presentation of search results: The $CL_s$ technique". In: *J. Phys. G* 28 (2002), p. 2693. DOI: 10.1088/0954-3899/28/10/313.

[28] Adi Bornheim et al. "Geant4 simulation of the light yield and timing performance for BTL sensors." In: *CMS DN-2018/031* (2018).

[29] Adi Bornheim et al. "BTL sensor performance in testbeam campaigns at FNAL and lab source measurements." In: *CMS DN-2018/032* (2018).

[30] Si Xie et al. "Barrel MIP timing detector testbeam setup in 2018 and 2019." In: *CMS DN-2019/042* (2019).

[31] *TimingDAQ*. https://github.com/CaltechPrecisionTiming/TimingDAQ. Accessed: 2022-07-18.

[32] *DataQualityMonitor*. https://github.com/CaltechPrecisionTiming/DataQualityMonitor. Accessed: 2022-07-18.

[33] Adi Bornheim et al. "Thermal performance of the barrel timing layer for the CMS phase 2 upgrade." In: *CMS DN-2018/039* (2019).

[34] Olmo Cerri et al. "Thermal studies on BTL prototype modules." In: *CMS DN-2020/013* (2020).

[35] *Rolled Tube Liquid Cold Plates Plates datasheet from Wakefield-Vette*. https://wakefieldthermal.com/wakefield-thermal-cold-plate-heatsink-0-041c-w-180-10-12c/. Accessed: 2020-10-10.

[36] *Fisher Scientific Isotemp Refrigerated/Heated Bath Circulators*. https://www.marshallscientific.com/v/vspfiles/specs/6200%20r28%20spec.pdf. Accessed: 2020-10-10.

[37] *Tflex 600 Series Thermal Gap Filler*. http://assets.lairdtech.com/home/brandworld/files/THR-DS-TFlex-600_07_2_14.pdf. Accessed: 2020-10-10.

[38] *TG-PP-10 Silicone Thermal Putty*. http://www.tglobaltechnology.com/uploads/files/tds/TG-PP-10.pdf. Accessed: 2020-10-10.

[39] *PR Series Ultra Precision Interchangeable Thermistors*. https://www.littelfuse.com/~/media/electronics/datasheets/leaded_thermistors/littelfuse_leaded_thermistors_interchangeable_thermistors_ultra_precision_pr_datasheet.pdf.pdf. Accessed: 2020-10-10.

[40] *AD5593R ADC datasheet*. https://www.analog.com/media/en/technical-documentation/data-sheets/AD5593R.pdf. Accessed: 2020-10-10.

[41] John S. Steinhart and Stanley R. Hart. "Calibration curves for thermistors". In: *Deep Sea Research and Oceanographic Abstracts* 15.4 (1968), pp. 497–503. ISSN: 0011-7471. DOI: https://doi.org/10.1016/0011-7471(68)90057-0. URL: http://www.sciencedirect.com/science/article/pii/0011747168900570.

[42] *CC2D35-SIP Humidity and Temperature Sensor*. http://www.farnell.com/datasheets/1855931.pdf. Accessed: 2020-10-10.

*Chapter 4*

# LEPTON FLAVOR UNIVERSALITY TEST AT CMS WITH $\mathcal{R}(D^*)$ MEASUREMENT WITH ALL-LEPTON TAU DECAYS

*“Non chi comincia, ma quel che persevera."*
Not who begins, but who perseveres.
— Leonardo Da Vinci, 1500 C.E. ca.

In this analysis we present a study for the measurement of the branching fraction ratio $\mathcal{R}(D^*) \equiv \mathcal{B}(B^0 \to D^{*-}\tau^+\nu_\tau)/\mathcal{B}(B^0 \to D^{*-}\mu^+\nu_\mu)$ with proton-proton CMS data collected in 2018 using a dedicated B parking stream. The $\tau$ lepton is tagged in the full leptonic decay channel $\tau \to \mu\nu_\tau\nu_\mu$, hence giving both numerator and denominator process the same visible final state consisting in a $\mu$ and a D$^*$ meson coming from the same displaced vertex without the presence of any additional track. The $\mathcal{R}(D^*)$ ratio is sensitive to contributions from non-standard-model particles that violate lepton flavor universality. This parameter, measured at B factories and hadron colliders, has been observed to have a tension with SM prediction of 2.5 sigma. A multidimensional fit to kinematic distributions of the reconstructed $B^0$ candidates is used to separate $\mu$ and $\tau$ decays, as well as background processes. While the outcome of the measurement is still blinded, using Asimov datasets it is estimated that the expected uncertainty on the measurement will be $\sigma$ ($\mathcal{R}(D^*)$) = 0.042. This result is competitive with state of art measurements and could impact the experimental world average. Furthermore, it can be the first measurement of this quantity at general purpose hadron collider experiments.

## 4.1 Introduction

The Standard Theory (SM) of particle physics predicts the three generations of leptons to have the same coupling to gauge bosons. This symmetry, called lepton flavor universality (LFU), is an accidental symmetry and is broken only by Yukawa interactions with the Higgs boson. State-of-the-art LFU experimental tests [1] are observed to be in partial tension with the theoretical prediction. Interactions mediated by neutral current have been studied at LEP and SLC [2]. There, direct precision tests of LFU in $Z$ couplings results are compatible with the SM prediction, with relative uncertainty values at the order of $10^{-3}$. The situation is different for interactions mediated by charged currents. On one hand, direct tests of LFU in $W$ couplings by LEP [3] present a 2.6 $\sigma$ difference between the measurement of the branching fraction of $W \to \tau\nu$ decay with respect to fractions of $W \to e\nu$ and $W \to \mu\nu$ decays (all with uncertainties of the order of $10^{-2}$). On the other, new results from the ATLAS collaboration [4] suggest a much better agreement with the SM predictions. Recently, other experiments performed LFU tests through precise measurements of heavy mesons branching fractions. Specifically, studying the ratio between the branching fraction of two different semi-leptonic decays of a meson allows an LFU test among the two leptons involved. Particular attention has been given to B mesons decays to an excited D meson and a lepton pair, resulting in the extensive scrutiny of the observable[1]:

$$\mathcal{R}(D^*) \equiv \frac{\mathcal{B}(B^0 \to D^{*-}\tau^+\nu_\tau)}{\mathcal{B}(B^0 \to D^{*-}\mu^+\nu_\mu)}, \tag{4.1}$$

where $\mathcal{B}(A \to XY)$ is the branching fraction of the state $A$ to $X$ and $Y$. Within the SM, for both $B^0 \to D^{*-}\tau^+\nu_\tau$ and $B^0 \to D^{*-}\mu^+\nu_\mu$ decays, the decay occurs via a $W^+$-mediated $b \to c\ell\nu$ interaction (Figure 4.1 left).

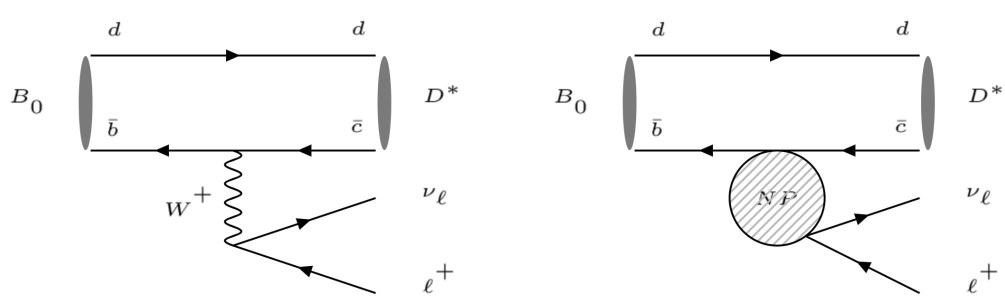


Figure 4.1: Tree level Feynman diagram for the $B^0 \to D^{*-}\ell^+\nu_\ell$ decay within the SM (left) and for new physics contributions (right).

The partial width for the decay processes can be written in the form

$$\frac{d\Gamma}{dq^2}(B \to D^*\ell\nu) = \eta_{ew}^2 g_l^2 g_q^2 |V_{cb}|^2 F(q^2), \tag{4.2}$$

[1]Expressed in the particular form which fits better the purpose of this document.

where $q^2 = (P_\mu + P_\nu)^2$ is the invariant mass squared of the lepton pair; $\eta_{ew}$ is the effect of QED corrections [5]; $g_l$ ($g_q$) is the lepton (quark) pair EW coupling; $V_{cb}$ is the relevant CKM matrix element; and $F(q^2)$ is a function of solely $q^2$ and the masses of the particles involved. Hence, in the SM, differences between the expected branching fraction of semi-leptonic decays into the three lepton families originate from the different masses of the charged leptons, either directly or through phase-space effects. In general, the ratio can then be expressed as

$$\mathcal{R}(D^*) = \frac{g_\tau^2}{g_\mu^2} \frac{\int_\tau F_\tau(q^2)}{\int_\mu F_\mu(q^2)}, \tag{4.3}$$

but it further simplifies if LFU holds true due to the assumption $g_\tau = g_\mu$. Given the cancellations which occur in the ratio, the measurement of $\mathcal{R}(D^*)$ results in a powerful and clean LFU test.

In the recent years, several percent-level SM predictions of $\mathcal{R}(D^*)$ have been published [6, 7, 8]: their results average to the value of $\mathcal{R}(D^*)_{th} = 0.258 \pm 0.005$ [9]. Considerable effort to measure $\mathcal{R}(D^*)$ has been made by experimental collaborations. At $e^+e^-$ B-factories, *BaBar* [10] and *BELLE* [11, 12, 13, 14] performed and recently updated their measurements. At hadron colliders, only the *LHCb* collaboration performed such a measurement exploiting, leptonic $\tau$ decays [15] first, and then three-prong hadronic $\tau$ decays [16]. While each single measurement has an uncertainty of about 10%, the world combination averages to $\mathcal{R}(D^*)_{exp} = 0.295 \pm 0.011 \pm 0.008$ [9]. The plot in Figure 4.2 summarizes the state of the art, comparing SM prediction with experimental measurements and estimating a tension of 2.5 $\sigma$ between them. In particular, it should be noted how the experimental uncertainty exceeds the SM prediction uncertainty by more than a factor 2. This underlines the focus needed within the experimental community to close this gap and enhance our comprehension of the LFU phenomenon.

The observed tension, which reaches $3\sigma$ when combined with $\mathcal{R}(D)$ results [9], could be a signature of physics processes beyond the SM with flavor-dependent couplings (Figure 4.1 right). Such a phenomenology can be achieved in models with an extended Higgs sector [17], leptoquarks [18, 19, 20, 21], or an extended gauge sector [22, 23] and can ultimately be explored with EFT approaches [24]. According to the authors, it is therefore of compelling and present-day interest to further extend the experimental effort and perform a first and novel $\mathcal{R}(D^*)$ measurement with a general purpose experiment like CMS.

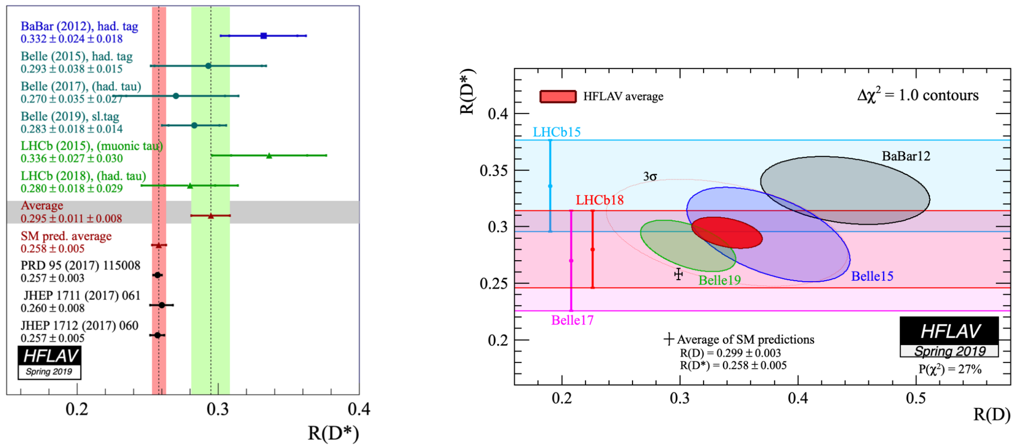


Figure 4.2: Comparison between public experimental measurements and theoretical predictions for the $\mathcal{R}(D^*)$ observable (left) and for the $\mathcal{R}(D) - \mathcal{R}(D^*)$ combination (right) as taken from [9] in March 2020.

This document showcases a measurement of $\mathcal{R}(D^*)$, using $B^0 \to D^{*-}\ell^+\nu_\ell$ with the leptonic $\tau$ decay (*i.e.* $\tau^+ \to \mu^+\nu_\mu\bar{\nu}_\tau$). The charge conjugate process $\overline{B}^0 \to D^{*+}\ell^-\bar{\nu}_\ell$ is also considered and its use is implicit in the rest of the note. While the muons are directly measured by the CMS detector, $D^*$ mesons usually decay before reaching the tracker and are therefore reconstructed starting from the decay products, targeting a specific decay chain. The decay chain $D^{*-} \to \pi^-\overline{D}_0 \to \pi^-K^+\pi^-$ has been selected because of the presence of solely charged hadrons in the final state, which allows us to reconstruct tracks and vertices for each particle and assign the particle mass given the charge without the need for additional particle identification. In addition, the analysis presented in this document will focus on the tag side, meaning that the muon from the $B$ must be the object that triggered the event. Hopefully, more documents exploring the probe side will follow.
As a consequence of the chosen $\tau$ decay, both B semi-leptonic decays involved will present a similar visible final state consisting of a $D^{*-}$ and a $\mu^+$ forming a vertex displaced from the primary interaction. A template fit to kinematic observables of the visible system will be used to distinguish between the events in which the $\mu^+$ is produced directly from the $B^0$ decay and those in which the $\mu^+$ is produced in the subsequent decay of a produced $\tau^+$. Consistent with the literature, the three kinematic observables chosen are: $q^2$, $M^2_{miss}$ and $E^*_\mu$ (a detailed discussion is present in Section 4.2).

Several uncertainties affect the measurement. The biggest uncertainty for the similar measurement done at *LHCb* [15] is the simulation statistical uncertainty. Systematic uncertainties due to the meson form factors, $B$ mesons transverse momentum ($p_T$) spectrum, efficiencies, and simulation statistic are considered among the others. Since the final state of both interesting processes is only partially visible, a non-negligible contamination from other background processes is expected, even after a meticulous selection of the events. There are two main expected sources of background. First, events where a $B$ meson decays to a lepton pair and a double excited charmed meson ($B \to D^{**}\ell^+\nu$) can be confused with signal events when the additional pion(s) produced in the $D^{**}$ decay are not reconstructed (either because neutrals or because of detector effects). Second, events where a $D^{*-}$ and a $\mu^+$ are produced through a $B$ meson decay chain but are not direct daughters of the meson itself. This is most likely to happen when a $B$ meson decays into two charmed mesons, where one is a $D^{*-}$ and the other decays into a $\mu^+$, with a decay like $B \to D^{*-}H_c$ and $H_c \to \mu^+ + X$.

The software developed to pursue this analysis, on top of the CMS software, is grouped in three repositories: Monte Carlo simulation cards, details, and utilities (`https://github.com/ocerri/BPH_CMSMCGen`); software to analyze event data model (EDM) files and produce flat ntuples (`https://github.com/ocerri/BPH_RDntuplizer`); and the analysis scripts, functions, and documentation (`https://github.com/ocerri/BPH_RD_Analysis`). A twiki gathering information and comments on the analysis can be found at `https://twiki.cern.ch/twiki/bin/view/CMS/RDstTagParkingData2018AN19162`.

## 4.2 Data and Simulation Samples

A complete catalog of the samples used in the analysis with their location in the CMS Data System (CDS) can be found at `https://github.com/ocerri/BPH_RDntuplizer/blob/master/production/samples.yml`.

### Data

The data (data, for real data) used in this analysis have been collected by the CMS detector during *proton-proton* collisions at LHC in 2018. Specifically, the data were recorded in a special high rate ($\sim$ 3 kHz) stream which, due to resource management, did not undergo the prompt reconstruction step and is known as *ParkingBPH* [25], consistently with previous CMS studies [26]. This stream groups a set of trigger paths firing on events in which a displaced muon is present, hence targeting with high purity [25] the production of B mesons which decay with a muon in the final state. The HLT paths present in ParkingBPH are named `HLT_MuX_IPY`, where `X` and Y represent respectively the threshold in transverse momentum ($p_T$) in GeV (requested at L1 and HLT) and transverse impact parameter significance (IP, HLT only) applied on the muon at trigger level. The only other requirement applied at trigger level is the L1 $\mu$ trigger logic cut on the muon pseudo-rapidity ($\eta$) at $|\eta| \leq 1.5$. The triggers $p_T$ threshold varies from 7 GeV to 12 GeV and the IP one from 4 to 6. In order to meet CMS requirements, these paths, which are always present in the trigger menu, are artificially and independently switched off in portions of the LHC fill imposing a pre-scale factor of 0. As a result, each HLT path has a different integrated luminosity: the set of active triggers evolves allowing non-zero prescale for looser paths as the instantaneous luminosity falls. Each path is furthermore divided into $n = 5$ (6) parts for the C and D (A and B, the switch from 5 to 6 happened halfway through era B) eras by the creation of $n$ copies of the same trigger path (appending `_partZ` at the end of the name, `Z`=1, ... $n$). Each copy is prescaled of a factor $n$ in order to retain the totality of triggered events. Given the conditions difference in the 2018 acquisition eras and the fact that almost 80% of the luminosity has been integrated in the last period, only the era D is used for this analysis. The HLT paths used in the analysis are `HLT_Mu7_IP4`, `HLT_Mu9_IP6` and `HLT_Mu12_IP6`. The input format for data are the centrally produced MINIAOD datasets `/ParkingBPH*/Run2018D-05May2019promptD-v1/MINIAOD`, which are then further analyzed using the global tag `102X_dataRun2_v11` and the certified luminosity mask `https://cms-service-dqm.web.cern.ch/cms-service-dqm/CAF/certification/Collisions18/13TeV/ReReco/Cert_314472-325175_13TeV_17SeptEarlyReReco2018ABC_PromptEraD_Collisions18_JSON.txt`. The integrated luminosity for each dataset, as obtained crossing the information from brilcalc and from the report of processed luminosity sections by Crab, is reported in Table 4.1.

Table 4.1: Integrated luminosity in $fb^{-1}$ for each dataset used in the analysis as obtained with the brilcalc utility [27].

| Primary dataset | Processed dataset | Mu7_IP4 | Mu9_IP6 | Mu12_IP6 |
|---|---|---|---|---|
| ParkingBPH1 | Run2018D-05May2019promptD-v1 | 1.28 | 4.12 | 5.25 |
| ParkingBPH2 | | 1.28 | 4.15 | 5.29 |
| ParkingBPH3 | | 1.28 | 4.13 | 5.29 |
| ParkingBPH4 | | 1.28 | 4.14 | 5.29 |
| ParkingBPH5 | | 1.28 | 4.15 | 5.29 |
| Total Run2018D-05May2019promptD-v1 | | 6.4 | 20.7 | 26.4 |

**Monte Carlo simulations**

The Monte Carlo simulation (MC) samples used in this analysis have been produced using the modules existing in CMSSW. The Pythia 8.230 [28] generator is used to simulate proton proton interactions at $\sqrt{s} = 13$ TeV through the CMSSW `Pythia8GeneratorFilter` module. Beside the CMS standard common settings and CP5 tuning settings, Pythia is run with the settings:

```
SoftQCD:nonDiffractive = on
PTFilter:filter = on
PTFilter:quarkToFilter = 5
PTFilter:scaleToFilter = 5.0
```

The first setting is intended to include the total cross-section of $pp$ collisions for inelastic, non-diffractive events. The last three settings include a filter that rejects events without at least a $b$ quark with a transverse momentum of 5 GeV. This threshold has been chosen by repeating the generation of a $B \to D^{*}\mu\nu$ sample with different values of the scale to filter. Starting from a value of 0 and scanning towards higher values we observed that there was no generation efficiency loss for events with muons with $p_T$ above 6.5 GeV for filter values below 5 GeV. Hence this value has been selected to increase the efficiency of the MC generation without losing phase space relevant to the analysis. The Pythia estimated cross section for the generated $pp \to b\bar{b}$ process is $\sigma_{b\bar{b}} = 4.5657 \pm 0.0003 \cdot 10^{11}$ fb. In a previous iteration of the MC generation, we also tried to use the `HardQCD:hardbbbar`. However, we realized that this subcategory does not include soft QCD initiated b-quark production which is still relevant. The B-physics oriented MC software EvtGen 1.3.0 [29] is used as an external library to simulate the decay chains of the relevant B mesons. The standard particle property file[2] is used unless differently specified and the standard decay table[3] is amended in a customized way to optimize the production of each specific sample as explained in the following subsections. When calling the CMSSW EvtGen plugin, the `list_forced_decays` option is used to ensure that one and only one particle decay chain is enforced in each event while letting all the other ones decay unbiased. In order to optimize the resource requirements of the generation process, the option `operates_on_particles` is

[2] `https://github.com/ocerri/BPH_CMSMCGen/blob/master/GeneratorInterface/EvtGenInterface/data/evt_2014.pdl`

[3] `https://github.com/ocerri/BPH_CMSMCGen/blob/master/GeneratorInterface/EvtGenInterface/data/DECAY_2014_NOLONGLIFE.DEC`

used to call EvtGen only in the decay of the relevant B hadrons. Unless differently specified, charge conjugation is implied for all the generated MC samples.

In the process of generating MC samples, most of the time is spent running steps representing the interaction of the particles with the detector, the simulation of the detector readout electronics, and the reconstruction of the event. Before running these parts, which take about 30 seconds per event, events are selected based on their MC event content in order to avoid wasting resources in proceeding with the generation for events that will not pass further selections. In particular, the specific selection may vary from sample to sample but they are aimed at selecting events where the muon from the B decay has at least 6.7 GeV of transverse momentum and $|\eta| \leq 1.5$.

When creating the MC configurations, the following parameters are passed to `cmsDriver`:

```
conditions  102X_upgrade2018_realistic_v15
beamspot  Realistic25ns13TeVEarly2018Collision
geometry  DB:Extended
era  Run2_2018
```

When necessary, pileup vertices are added during the DIGI step from the GEN-SIM dataset:
`/MinBias_TuneCP5_13TeV-pythia8/RunIIFall18GS-102X_upgrade2018_realistic_v9-v1/GEN-SIM`.
Samples with different pileup conditions have been generated to better suit dedicated tasks. Samples with Poisson distributed pileup are generated with the option `pileup "AVE_25_BX_25ns,{'N': m}"`, whereas custom pileup profiles are created directly setting the probability distribution in the CMSSW configuration file.

The code and the MC fragments used to run the generation are collected in the following repository: `https://github.com/ocerri/BPH_CMSMCGen`.

**Signals simulation samples**

In this section, *signals* is used to intended the two processes that enter in the $\mathcal{R}(D^*)$ ratio: $B^0 \to D^{*-}\mu^+\nu_\mu$ and $B^0 \to D^{*-}\tau^+\nu_\tau$. When generating signal samples the decay of all $B^0$ mesons is forced to the $D^{*-}\ell^+\nu_\ell$ final state. The decay kinematic is generated according to the ISGW2 form factors (FF) model [30]. Despite more recent works have been published, this model has been chosen for simplicity in

the synchronization with the HAMMER library [31] used to reweigh the events to different FF schemes. A smaller sample of $B^0 \to D^{*-}\mu^+\nu_\mu$ has been generated with the HQET2 FF scheme for testing purposes of the reweighing procedure. The values used for the HQET2 free parameters used are: $\rho^2 = 1.205 \pm 0.026$ (`rho2`), $R_1(1) = 1.404 \pm 0.032$ (`R1_1`) and $R_2(1) = 0.854 \pm 0.020$ (`R2_1`) from [9]; $h_{A1}(1) = 0.921 \pm 0.024$ (`hA1_1`) from [32]; and $R_0(1) = 1.14 \pm 10\%$ (`R0_1`) from [33].
Final state radiation (FSR) in $B^0$ decays is modeled interfacing EVTGEN with the PHOTOS [34] library. All $D^{*-}$ mesons from $B^0$ are forced to decay to $\overline{D}^0\pi^-$ though the `VSS` option which decays a vector particle into two scalars with an amplitude $A = \epsilon^\mu v_\mu$, where $\epsilon$ is the polarization of the vector ($D^*$) and $v$ is the velocity of one of the two daughters. $\overline{D}^0$ produced from those $D^{*-}$ are forced to decay into the final state $K^+pi^-$ through a generic 2-bodies phase space (`PHSP`). Finally, $\tau^+$ from $B^0$ are forced to decay to $\mu^+\nu_\nu\nu_\tau$ with the option `TAULNUNU`, consisting essentially in a V-A interaction.

Using data from [1], the branching ratios of the full decay chains are $6.75 \pm 0.20 \cdot 10^{-4}$ for $B \to D^*\mu\nu$ and $2.35 \pm 0.07 \cdot 10^{-4}$ for $B \to D^*\tau\nu$, assuming $\mathcal{R}(D^*) = 1$. The cuts applied at generator level are: presence of at least one $B^0$ meson; and presence of at least one muon from the $B^0$ decay chain with $p_T > 6.7$ GeV and $|\eta| < 1.6$.
The generator level cuts efficiency is $4.454 \pm 0.003 \cdot 10^{-3}$ for $B \to D^*\mu\nu$ and $0.857 \pm 0.002 \cdot 10^{-3}$ for $B \to D^*\tau\nu$. These efficiencies are the results of two factors. The first factor is the probability of having a $B^0$, estimated by PYTHIA to be 0.4, and is shared by both processes. The second factor is the probability that a $B^0$ will produce a muon within the phase space requirements ($p_T$ and $\eta$) and it is different for the two processes. The kinematic distributions at generator level for the muon produced in the $B$ decay are shown in Fig. 4.3. Unless differently specified, the sample considered in the following for $B \to D^{*-}\mu^+\nu$ process is the one produced with the custom PU profile `c2` (Sec. 4.4) since it is also the sample used in the fit to data.

#### Backgrounds simulation: $B_{(s)} \to D^{**}_{(s)}\ell\nu$

These samples include decays that are similar to the signal ones, except that a double excited charmed meson ($D^{**}$) is produced instead of a single excited one ($D^*$). A $B \to D^{**}\mu\nu$ decay can appear as a background if the $D^{**}$ is produced in association with a $\mu^+$ and it decays into a $D^{*-}$ (plus additional pions). Semi-leptonic $B$ decays

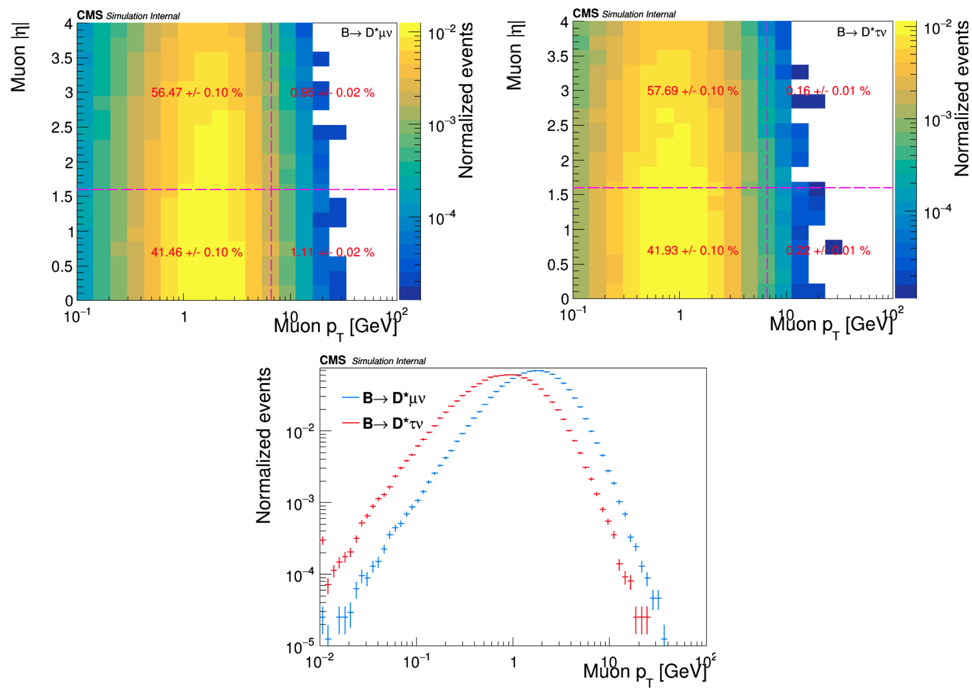


Figure 4.3: Generator level kinematic distribution of the muon produced in $B \to D^*\mu\nu$ (left), $B \to D^*\tau\nu$ (center) and combined (right). The relative efficiency is shown in red for each of the four regions defined by the cuts, the region that is selected for the analysis is the one at high $p_T$ and low $\eta$.

into $D^{**}$ have been studied in B factories [35, 36]. Several double excited states of charmed mesons exists (Fig. 4.4), all which decay through strong interaction into $D$ and $D^*$ mesons with the emission of one or more light unflavored mesons. It should be noted that the state $D_1^*$ is also referred to as $D_1'$ and corresponds to the particle $D_1(2430)$ (pdg id 20423 and 20413) in Pythia. Also, the state $D'^*$ does not exist in Pythia, but is referred to as $D^*(2S)$ in EvtGen and $D^*(2640)$ in [1]. The branching fractions of decays with the same number of pions are related to each other by the isospin symmetry. As a result, a decay is suppressed by a factor $1/2$ for each $\pi_0$ with respect to the decay with the same number of charged pions. The processes considered in this analysis, divided into 10 MC samples, are shown in Tab. 4.2.

The division into the 10 samples is represented in the table by the integer ID which corresponds to an MC card in the MC gen github. When needed, a breakdown of the sub-processes present in the sample is also present. The branching fraction reported in the second column does not include the factor due to the desired $D^*$ decay chain but includes a 0.17 factor from $\tau \to \mu\nu\nu$ when needed. For the backgrounds with a

Table 4.2: Processes simulated for $B_{(s)} \rightarrow D^{**}_{(s)}\ell\nu$ background. The first two rows with ID 0.1 and 0.2 represent for comparison the signal MC.

| ID | Sample | BR [$10^{-3}$] | $\epsilon_{gen}/\epsilon_{\mu}$ | Notes |
|---|---|---|---|---|
| 0.1 | $B^0 \rightarrow D^{*-}\mu^+\nu$ | 50.5 (±1.4) | 1 | From PDG. |
| 0.2 | $B^0 \rightarrow D^{*-}\tau^+\nu$ | 2.67 (±0.15) | 0.23 ± 0.07 | From PDG. |
| 1 | $B^+ \rightarrow D^{*-}\pi^+\mu^+\nu$ | 6.0 ± 0.4 | 0.99 ± 0.20 | From PDG. Compatible with the sum of below. |
| 1.1 | via $D_1$ | 3.03 ± 0.20 | - | From PDG $\Gamma_{13}$. |
| 1.2 | via $D_1'$ | 2.70 ± 0.60 | - | From PDG $\Gamma_{14}$. |
| 1.3 | via $D_2^*$ | 1.01 ± 0.24 | - | From PDG $\Gamma_{15}$. |
| 1.4 | non-resonant | 0.45 ± 0.45 | - | Not observed, from EvtGen default. |
| 2 | $B^0 \rightarrow D^{*-}\pi^0\mu^+\nu$ | 3.0 | 0.90 ± 0.11 | Not measured. Half of [1] (isospin). Compatible with CC mode from PDG $\Gamma_{12}/2$ (isosping check). Uncertainty correlated to [1], breakdown also compatible (see below). |
| 2.1 | via $D_1$ | 1.40 ± 0.14 | - | From PDG CC mode divide by 2. |
| 2.2 | via $D_1'$ | 1.55 ± 0.45 | - | From PDG CC mode divide by 2. |
| 2.3 | via $D_2^*$ | 0.34 ± 0.06 | - | From PDG CC mode divide by 2. |
| 2.4 | non-resonant | 0.06 ± 0.06 | - | Not observed, from EvtGen default. |
| 3 | $B^0 \rightarrow D^{*-}\pi\pi\mu^+\nu$ | 1.20 | - | See Tab. 4.3. |
| 4 | $B^+ \rightarrow D^{*-}\pi^+\pi^0\mu^+\nu$ | 0.48 | - | Isospin from [3.1]. Uncertainty correlated to [3.1]. |
| 5 | $B^+ \rightarrow D^{*-}\pi^+\tau^+\nu$ | 0.20 | 0.21 ± 0.06 | From [1] estimating $R(D^{**}) = 0.2$, same breakdown as [1]. Uncertainty correlated to [1]. |
| 6 | $B^0 \rightarrow D^{*-}\pi^0\tau^+\nu$ | 0.10 | - | Not measured, isospin from [5]. Uncertainty correlated to [5]. |
| 7 | $B^0 \rightarrow D^{*-}\pi\pi\tau^+\nu$ | 0.041 | - | Sum of [7.1] and [7.2]. |
| 7.1 | $B^0 \rightarrow D^{*-}\pi^+\pi^-\tau^+\nu$ | 0.033 | - | From [3.1] estimating $R(D^{**})$. Unc. correlated to [3.1]. |
| 7.2 | $B^0 \rightarrow D^{*-}\pi^0\pi^0\tau^+\nu$ | 0.008 | - | Isospin from [7.1], unc. correlated. |
| 8 | $B^+ \rightarrow D^{*-}\pi^+\pi^0\tau^+\nu$ | 0.016 | - | Isospin from [7.1], unc. correlated. |
| 9 | $B_s \rightarrow D^{*-}K^0\mu^+\nu$ | 5.9 ± 1.5 | 0.20 ± 0.03 | Double the sum of [9.1] and [9.2] because those only consider $K_S$. |
| 9.1 | via $D_{s1}' \rightarrow D^{*-}K_S^0$ | 2.70 ± 0.7 | - | From PDG. |
| 9.2 | via $D_{s2}^* \rightarrow D^{*-}K_S^0$ | 0.25 ± 0.25 | - | From PDG using $\Gamma_9 \cdot \Gamma_8/\Gamma_7$ times 0.3 (exclusion upper limit from here). |
| 10 | $B_s \rightarrow D^{*-}K^0\tau^+\nu$ | 0.21 | 0.05 ± 0.02 | From [9] estimating $R(D_s^{**}) = 0.2$, unc. correlated with [9]. Same breakdown as [9]. |

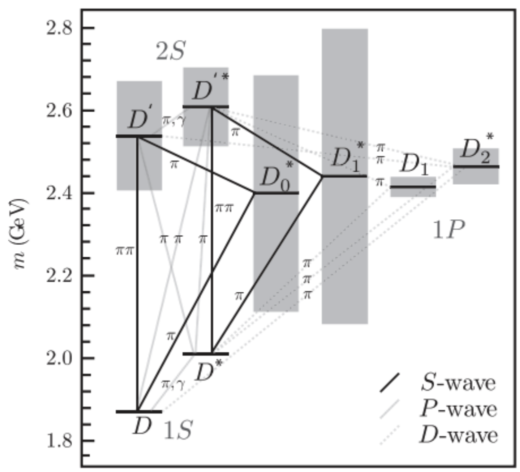


Figure 4.4: Picture from [37]. Strong decays of $D^{**}$ mesons into $D$ and $D^*$ with the emission of one or two pions. The gray band around the mass value of each states represent its width.

$\tau$ in the final state, a value of $R(D^{**}_{(s)}) = 0.2$ is estimated. The third column reports the generator efficiency relative to the $B^0 \to D^{*-}\mu^+\nu$ signal sample as measured on small test generation before the full-scale generation. From the same samples, we estimate that the additional charged pions from $D^{**}$ decays have about 65-75 % probability of being reconstructed in CMS. As additional remarks to the decay table, the following should be noted. Pion decays are not forced, hence the $\pi^0 \to ee\gamma$ process is included. Light mesons come from strong decays which do not violate flavor, hence there is no Cabibbo suppressed decay mode. Changing the K with a $\pi$ is not a weak vertex but implies creating a $s\bar{s}$ pair instead of a light unflavored quark pair, hence the charm mesons should become charmed-strange. The process $D^{**} \to KD^*$ can not happen because the $D^{**}$ decays via the creation of a $q\bar{q}$ quarks couple. The $D^{**} \to KD_s^{(*)}$ is forbidden by energy conservation. The process $D_s^{**} \to \pi D^*$ can not happen and the $D_s^{**} \to \pi D_s^*$ will not have the $D^*$ in the final state. $D_s^{**}$ resonances do not have enough mass to decay into $D^*K$ with additional pions because $D^*K$ saturates the phase space.

Decays with the production of one single pion ($B \to D^{**}\ell\nu$, with $D^{**} \to D^*\pi$) are simulated via a proper mixture of $D_1$, $D_1^*$, $D_2^*$ and non-resonant contribution. The resonant decay $B \to D^{**}\ell\nu$ is simulated using the ISGW2 FF scheme, with the $D^{**} \to D^*\pi$ decays simulated using the appropriate S-wave ($D_1^*$) or D-wave ($D_1$, $D_2^*$) amplitude. The most updated PDG value is used for the mass and width of these resonances. The non-resonant contribution is instead modeled with the

EvtGen corrected Goity-Roberts model [38]. Particular attention is dedicated to simulating decays with 2 pions in the final state. A first simple simulation of $B \to D^{**}\ell\nu$ with $D^{**} \to D^*\pi\pi$ where only the $D^*(2S)$ was considered highlighted how a simple model was not enough to describe the complex phase space observed in the control region of the analysis. As an upgrade, we increased the number of decays considered in our simulation to include $D^{**} \to D^*\pi\pi$ from all the relevant $D^{**}$ resonances and chained decays of the form $B \to D^{**}\pi\ell\nu$ with $D^{**} \to D^*\pi$. A full list of the processes simulated for the $B \to D^*\pi\pi\ell\nu$ can be found in Tab. 4.3. Similarly to the $B \to D^*\pi\mu\nu$ sample, the $B \to D^{**}\ell\nu$ decays are simulated using the

Table 4.3: Explicit breakdown of sample 3 from Tab. 4.2. The mass (M) and width ($\Gamma$) of the $D^{**}$ resonances involved in the decays is reported together with the branching ratio.

| ID | Sample | BR $[10^{-3}]$ | $\Gamma$ | M | Notes |
|---|---|---|---|---|---|
| 3 | $B^0 \to D^{*-}\pi\pi\mu^+\nu$ | 1.20 | | | Using $R^*_{\pi^+\pi^-} = 19 \pm 6 \cdot 10^{-3}$ from [39] times [0.1]. Sum of contributions below are re-scaled to 1 by EvtGen. |
| 3.1 | $D^*(2S)^-\mu^+\nu$ | 20 | 400 | 2640 | 80% $D^{*-}\pi^+\pi^-$, 20% $D^{*-}\pi^0\pi^0$. Only one in old MC. |
| 3.2 | $D_1^-\mu^+\nu$ | 3 | 25 | 2420 | 80% $D^{*-}\pi^+\pi^-$, 20% $D^{*-}\pi^0\pi^0$ |
| 3.3 | $D_1'^-\mu^+\nu$ | 2.7 | 385 | 2430 | 80% $D^{*-}\pi^+\pi^-$, 20% $D^{*-}\pi^0\pi^0$ |
| 3.4 | $D_2^{*-}\mu^+\nu$ | 1.0 | 46 | 2460 | 80% $D^{*-}\pi^+\pi^-$, 20% $D^{*-}\pi^0\pi^0$ |
| 3.5 | $D_1^-\pi^0\mu^+\nu$ | 0.75 | 25 | 2420 | 100% $D^{*-}\pi^0$, 1/4 of [1.8] (isospin) |
| 3.6 | $D_1'^-\pi^0\mu^+\nu$ | 0.67 | 385 | 2430 | 100% $D^{*-}\pi^0$, 1/4 of [1.9] (isospin) |
| 3.7 | $D_2^{*-}\pi^0\mu^+\nu$ | 0.25 | 46 | 2460 | 100% $D^{*-}\pi^0$, 1/4 of [1.10] (isospin) |
| 3.8 | $D_1^0\pi^-\mu^+\nu$ | 3 | 25 | 2420 | 100% $D^{*-}\pi^+$. |
| 3.9 | $D_1'^0\pi^0\mu^+\nu$ | 2.7 | 385 | 2430 | 100% $D^{*-}\pi^+$. |
| 3.10 | $D_2^{*0}\pi^0\mu^+\nu$ | 1 | 46 | 2460 | 100% $D^{*-}\pi^+$. |

ISGW2 FF scheme while $B \to D^{**}\pi\ell\nu$ decays are simulated via a flat phase-space decay.

As for the signal samples, $D^{*-}$ meson produced in these decays is forced to decay to $\pi^-K^+\pi^-$ final state. Given the presence of multiple resonances or ambiguities arising from a particle missing in Pythia, the generator cuts applied have small differences from the signal ones but are designed to be equivalent.

**Backgrounds simulation:** $B \to D^*H_c(\mu X)$

This set of samples includes hadronic decays of $B$ mesons in which the $D^{*-}$ or the $\mu^+$ are not produced directly from the $B$ meson decay but as a result of a secondary decay. This happens almost exclusively when the $B$ decays into two charmed mesons, one of which undergoes a semi-leptonic decay involving a muon in the final state and the other one is a $D^{*-}$. It is in principle also possible for the $D^{*-}$ to be a product of a secondary decay (e.g. the $D^{*-}$ is produced in the decay of a $D^{**}$ meson), but this is practically only relevant for decays of the form $D_s \rightarrow D^* K$. Instead, it is not uncommon for the charmed mesons which produce the muon to come from an excited $D$ meson promptly produced in the $B$ decay (e.g. $B^0 \rightarrow D^{*-} D^{*+} K^0$, with $D^{*+} \rightarrow \pi^+ D^0$ and $D^0 \rightarrow \mu + X$). The major contribution to this sample is given by diagrams with only one Cabibbo suppressed vertex (the one that mediates $b \rightarrow Wc$). As a consequence, it is expected the sample to be dominate by two kind of decays: $B \rightarrow D^{*-} D_s^{(*)}$ where the $W$ materialize as a charmed-strange meson (e.g. fig. 4.5 left); $B \rightarrow D^{*-} D^{(*)} K^{(*)}$ where an additional $q\bar{q}$ pair is radiated during the hadronization (e.g. fig. 4.5 right). Since the excited charmed mesons decay strong,

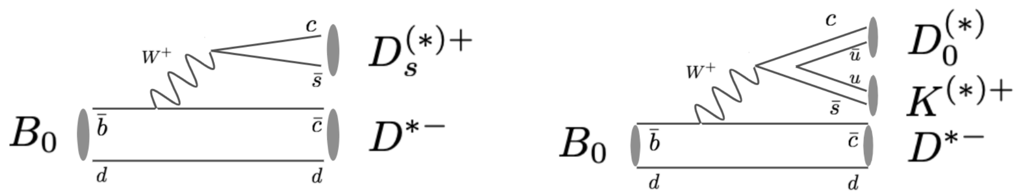


Figure 4.5: Example tree level Feynman diagram for the $B \rightarrow D^{*-} D_s^{(*)}$ (left) and $B \rightarrow D^{*-} D^{(*)} K^{(*)}$ (right) processes.

only the non-excited $D$ and $D_s$ mesons that decay via an electroweak interaction can produce a muon in the final state. All the considered decays of a charmed meson with a muon in the final state are reported in Tab. 4.5 accompanied by a note on how the branching fraction has been estimated. As a remark, when considering the possible decay chains of the charmed mesons with a $\mu$ in the final state we include also $D_{(s)}^{(*)} \rightarrow \tau + X$ with $\tau \rightarrow \mu\nu\nu$. This includes particularly the $D_s \rightarrow \tau\nu$ decay which, considering the requirement of a leptonic $\tau$ decay, has branching fraction of $0.053 \cdot 0.174 = 0.0093$. We also report in Tab. 4.4 the breakdown of the $D^*$ decays considered in the generation of this set of backgrounds.

In order to efficiently generate $B \rightarrow D^* H_c$ events, the different processes have been divided into several groups of MC samples. Tab. 4.6 shows a comprehensive list of the decays included for this kind of background divided per sample ID in a similar manner as Tab. 4.2.

Table 4.4: Branching fraction of $D^*$ mesons.

| Meson | BR [$10^{-2}$] | Notes |
|---|---|---|
| $D^{*0}$ | - | |
| $D^0\pi^0$ | $64.7 \pm 0.9$ | From PDG. |
| $D^0\gamma$ | $35.3 \pm 0.9$ | From PDG. |
| $D^{*+}$ | - | |
| $D^0\pi^+$ | $67.7 \pm 0.5$ | From PDG. |
| $D^+\pi^0$ | $30.7 \pm 0.5$ | From PDG. |
| $D^+\gamma$ | $1.6 \pm 0.4$ | From PDG. |
| $D_{s0}^{*+}$ | - | |
| $D_s^+\pi^0$ | 100 | From PDG and EvtGen. |
| $D_s^{*+}$ | - | |
| $D_s^+\gamma$ | $93.5 \pm 0.7$ | From PDG. |
| $D_s^+\pi^0$ | $5.8 \pm 0.7$ | From PDG. |
| $D_s^+e^+e^-$ | $0.67 \pm 0.16$ | From PDG. |

Table 4.6: Relevant processes included in the analysis as $B \to D^*H_c$ background.

| ID | Sample | BR [$10^{-3}$] | $\epsilon_{gen}/\epsilon_\mu$ [%] | Notes |
|---|---|---|---|---|
| 0.1 | $B^0 \to D^{*-}\mu^+\nu$ | 50.5 (±1.4) | 100 | From PDG. |
| 0.2 | $B^0 \to D^{*-}\tau^+\nu$ | 2.67 (±0.15) | 23 ± 7 | From PDG, includes $\tau \to \mu\nu\nu$ (implicit in the following). |
| 1 | $B^0 \to D^{*-}D^0(\mu X)Y$ | $2.25 \pm 0.12$ | $7 \pm 0.5$ | Sum of below 36.9±1.3 times [C1]. |
| 1.1 | $D^{*-}D^0K^+$ | $2.47 \pm 0.21$ | - | PDG $\Gamma_{174}$. |
| 1.2 | $D^{*-}D^0K^{*+}$ | 1.24 | - | Half of above. |
| 1.3 | $D^{*-}D^{*0}K^+$ | $10.6 \pm 0.9$ | - | PDG $\Gamma_{175}$. |
| 1.4 | $D^{*-}D^{*0}K^{*+}$ | 5.3 | - | Half of above. |
| 1.5 | $D^{*-}D^{*+}(D^0\pi^+)K^0$ | $5.43 \pm 0.47$ | - | PDG $\Gamma_{178}$, includes 0.67 from $D^{*+} \to D^0\pi^+$. |
| 1.6 | $D^{*-}D^{*+}(D^0\pi^+)K^{*0}$ | 2.7 | - | Half of above. |
| 1.7 | $D^{*-}D^{*+}(D^0\pi^+)$ | $0.54 \pm 0.04$ | - | PDG $\Gamma_{168}$, includes 0.67 from $D^{*+} \to D^0\pi^+$. |
| 1.8 | $D^{*-}(\overline{D}^0\pi^-)D^{*+}K^0$ | 5.43 | - | Like [1.5] with swapped decays. |

| ID | Sample | BR [$10^{-3}$] | $\epsilon_{gen}/\epsilon_{\mu}$ [%] | Notes |
|---|---|---|---|---|
| 1.9 | $D^{*-}(\overline{D}^0\pi^-)D^{*+}K^{*0}$ | 2.7 | - | Like [1.6] with swapped decays. |
| 1.10 | $D^{*-}(\overline{D}^0\pi^-)D^{*+}$ | 0.54 | - | Like [1.7] with swapped decays. |
| 2 | $B^0 \to D^{*-}D^+(\mu X)Y$ | 3.07 ± 0.13 | 5.6 ± 0.4 | Sum of below 19.34 ± 0.74 times [C2]. |
| 2.1 | $D^{*-}D^+K^0$ | 3.2 ± 0.25 | - | Half of PDG $\Gamma_{177}$. |
| 2.2 | $D^{*+}D^-K^0$ | 3.2 ± 0.25 | - | Half of PDG $\Gamma_{177}$. |
| 2.3 | $D^{*-}D^+K^{*0}$ | 1.6 | - | Half of [2.1]. |
| 2.4 | $D^{*+}D^-K^{*0}$ | 1.6 | - | Half of [2.2]. |
| 2.5 | $D^{*-}D^{*+}(D^+X^0)K^0$ | 2.67 ± 0.23 | - | PDG $\Gamma_{178}$, includes 0.33 from $D^{*+} \to D^+X^0$. |
| 2.6 | $D^{*-}D^{*+}(D^+X^0)K^{*0}$ | 1.33 | - | Half of above. |
| 2.7 | $D^{*-}D^{*+}(D^+X^0)$ | 0.26 ± 0.02 | - | PDG $\Gamma_{168}$, includes 0.33 from $D^{*+} \to D^+X^0$. |
| 2.8 | $D^{*-}(D^-X^0)D^{*+}K^0$ | 2.67 | - | Like [2.5] with swapped decays. |
| 2.9 | $D^{*-}(D^-X^0)D^{*+}K^{*0}$ | 1.33 | - | Like [2.6] with swapped decays. |
| 2.10 | $D^{*-}(D^-X^0)D^{*+}$ | 0.26 | - | Like [2.7] with swapped decays. |
| 2.11 | $D^{*+}D^-$ | 0.61 ± 0.15 | - | PDG $\Gamma_{170}$ |
| 2.12 | $D^{*-}D^+$ | 0.61 | - | CC of above. |
| 3 | $B^0 \to D^{*-}D_s^+(\mu X)Y$ | 2.05 ± 0.13 | 18.1 ± 1.3 | Sum of below 27.2±1.5 times [C3]. |
| 3.1 | $D^{*-}D_s^+$ | 8.0 ± 1.1 | - | From PDG $\Gamma_{83}$. |
| 3.2 | $D^{*-}D_s^{*+}$ | 17.7 ± 0.14 | - | From PDG $\Gamma_{85}$. |
| 3.3 | $D^{*-}D_{s0}^{*+}$ | 1.5 ± 0.6 | - | From PDG $\Gamma_{95}$. |
| 4 | $B^+ \to D^{*-}D^0(\mu X)Y$ | 1.42 ± 0.14 | 4.4 ± 1.3 | Sum of below 23.3±2.0 times [C1]. |
| 4.1 | $D^{*+}\overline{D}^0K^0$ | 3.8 ± 0.4 | - | PDG $\Gamma_{195}$ |
| 4.2 | $D^{*+}\overline{D}^0K^{*0}$ | 1.9 | - | Half of above. |
| 4.3 | $D^{*+}\overline{D}^{*0}K^0$ | 9.2 ± 1.2 | - | PDG $\Gamma_{196}$ |
| 4.4 | $D^{*+}\overline{D}^{*0}K^{*0}$ | 4.6 | - | Half of above. |

| ID | Sample | BR [$10^{-3}$] | $\epsilon_{gen}/\epsilon_{\mu}$ [%] | Notes |
|---|---|---|---|---|
| 4.5 | $D^{*-}D^{*+}(D^0\pi^+)K^+$ | $0.88 \pm 0.12$ | - | PDG $\Gamma_{204}$, includes 0.67 from $D^{*+} \to D^0\pi^+$ |
| 4.6 | $D^{*-}D^{*+}(D^0\pi^+)K^{*+}$ | 0.44 | - | Half of above. |
| 4.7 | $D^{*-}(\overline{D}^0\pi^-)D^{*+}K^+$ | 0.88 | - | Like [4.5] with swapped decays. |
| 4.8 | $D^{*-}(\overline{D}^0\pi^-)D^{*+}K^{*+}$ | 0.44 | - | Like [4.6] with swapped decays. |
| 4.9 | $D^{*+}\overline{D}^0$ | $0.39 \pm 0.05$ | - | PDG $\Gamma_{190}$ |
| 4.10 | $D^{*+}\overline{D}^{*0}$ | $0.81 \pm 0.17$ | - | PDG $\Gamma_{188}$ |
| 5 | $B^+ \to D^{*-}D^+(\mu X)Y$ | $0.34 \pm 0.03$ | - | Sum of below $2.11 \pm 0.20$ times [C2]. |
| 5.1 | $D^{*-}D^+K^+$ | $0.60 \pm 0.12$ | - | PDG $\Gamma_{203}$ |
| 5.2 | $D^{*-}D^+K^{*+}$ | 0.3 | - | Half of above. |
| 5.3 | $D^{*+}D^-K^+$ | $0.63 \pm 0.11$ | - | PDG $\Gamma_{202}$ |
| 5.4 | $D^{*+}D^-K^{*+}$ | 0.31 | - | Half of above. |
| 5.5 | $D^{*-}D^{*+}(D^+X^0)K^+$ | $0.44 \pm 0.06$ | - | PDG $\Gamma_{204}$, includes 0.33 from $D^{*+} \to D^+X^0$ |
| 5.6 | $D^{*-}D^{*+}(D^+X^0)K^{*+}$ | 0.22 | - | Half of above. |
| 5.7 | $D^{*-}(D^-X^0)D^{*+}K^+$ | 0.44 | - | Like [5.5] with swapped decays. |
| 5.8 | $D^{*-}(D^-X^0)D^{*+}K^{*+}$ | 0.22 | - | Like [5.6] with swapped decays. |
| 6 | $B_s^0 \to D^*D_s(\mu X)Y$ | $2.33 \pm 2.0$ | $0.8 \pm 0.4$ | Sum of below 30.9 times [C3] with some added 100% uncertainty. |
| 6.1 | $D^{*+}D_s^{*-}K^0$ | 15 | - | Not observed, EvtGen default. |
| 6.2 | $D^{*+}D_s^{*-}K^{*0}$ | 5 | - | Not observed, EvtGen default. |
| 6.3 | $D^{*+}D_s^-K^0$ | 5 | - | Not observed, EvtGen default. |
| 6.4 | $D^{*+}D_s^-K^{*0}$ | 2.5 | - | Not observed, EvtGen default. |

| ID | Sample | BR [$10^{-3}$] | $\epsilon_{gen}/\epsilon_\mu$ [%] | Notes |
|---|---|---|---|---|
| 6.5 | $D^{*-}D_s^+$ | 1.7 | - | Not observed, EvtGen default. |
| 6.6 | $D^{*-}D_s^{*+}$ | 1.7 | - | Not observed, EvtGen default. |
| 7 | $B^0 \to D^* D_{s1}$ | $0.79 \pm 0.3$ | - | Sum of below. |
| 7.1 | $D^{*-}D_{s1}^{(2460)+}(\mu X)$ | $0.70 \pm 0.17$ | - | PDG $\Gamma_{101}$ times [C3], ending in $D_s \to \mu Y$. |
| 7.2 | $D^-(\mu X)D_{s1}'^{(2536)+}(D^* Y)$ | $0.02 \pm 0.01$ | - | PDG $\Gamma_{103}/2$ times [C2]. |
| 7.3 | $D^{*-}(\mu X)D_{s1}'^{(2536)+}(D^* Y)$ | $0.02 \pm 0.01$ | - | PDG $\Gamma_{106}/2$ times $D^* \to \mu X$. |
| 7.4 | $D^{*-}D_{s1}'^{(2536)+}(\mu X)$ | $0.04 \pm 0.02$ | - | PDG $\Gamma_{106}/2$ times $D'_{s1} \to \mu X$. |
| 8 | $B^+ \to D^{(*)}D'_{s1}(2536)$ | $0.45 \pm 0.45$ | - | Sum of below times 0.5 ($D'_{s1} \to D^* K^0$) times [C1]. |
| 8.1 | $\overline{D}^0(\mu X)D_{s1}'^+(D^{*+}K^0)$ | 10 | - | Not measured. |
| 8.2 | $\overline{D}^{*0}(\mu X)D_{s1}'^+(D^{*+}K^0)$ | 5 | - | Not measured. |
| 9 | $B \to D^* D X X$ | $5 \pm 5$ | - | Not observed nor present in EvtGen. |
| 9.1 | $B^0 \to D^{*-}D^0(\mu X)\pi^+\pi^0$ | 1 | - | Not measured. |
| 9.2 | $B^0 \to D^{*-}D^+(\mu X)\pi^+\pi^-$ | 1 | - | Not measured. |
| 9.3 | $B^0 \to D^{*-}D^0(\mu X)\rho^+$ | 1 | - | Not measured. |
| 9.4 | $B^0 \to D^{*-}D^+(\mu X)K^+\pi^-$ | 1 | - | Not measured. |
| 9.5 | $B^+ \to D^{*-}D^0(\mu X)\rho^+ K^+$ | 1 | - | Not measured. |
| 9.6 | $B^+ \to D^{*-}D^+(\mu X)\pi^+\overline{K}^0$ | 1 | - | Not measured. |
| 9.7 | $B^+ \to D^{*-}D^+(\mu X)\rho^+$ | 1 | - | Not measured. |

The interesting decays from $B^0$ have been divided into the first three blocks based on the meson which is the mother of the muon ($D^0$, $D^+$ or $D_s$). The second three groups collect instead the similar interesting decays from the charged mesons $B^\pm$ and the charmed-strange meson $B_s$. The final three blocks instead contain rare processes involving excited $D_s$ or unmeasured multi-particles final states. All the $D^{*-}$ produced in the relevant B decays and not followed by a round bracket are forced to the $\pi^- K^+ \pi^-$ final state. Similarly, the proper charmed meson is forced to decay in a semi-leptonic final state using the default decay models in EvtGen. The

Table 4.5: Collection of charmed mesons decays producing a muon in the final state. The branching fraction of decays including a $\tau$ lepton factors in already the 0.17 fraction due to the $\tau \to \mu\nu\nu$ decay rate.

| ID | Process | BR [$10^{-3}$] | Notes |
|---|---|---|---|
| C1 | $D^0 \to \mu X$ | 60.8 ± 2.5 | Sum of below, compatible with PDG inclusive. |
| C1.1 | $K^-\mu^+\nu$ | 34.1 ± 0.4 | From PDG. |
| C1.2 | $K^{*-}\mu^+\nu$ | 18.9 ± 2.4 | From PDG. |
| C1.3 | $\pi^-\mu^+\nu$ | 2.67 ± 0.12 | From PDG. |
| C1.4 | $\rho^-\mu^+\nu$ | 1.50 ± 0.12 | From PDG similar channel with electron (link). |
| C1.5 | $K_1(1270)^-\mu^+\nu$ | 0.76 ± 0.3 | From PDG similar electron channel (link). |
| C1.6 | $\overline{K}^0\pi^-\mu^+\nu$ | 0.77 ± 0.16 | Non resonant. From electron paper. |
| C1.7 | $K^-\pi^0\mu^+\nu$ | 0.39 | Isospin from above. |
| C1.8 | $K_2^{*-}\mu^+\nu$ | 0.3 ± 0.3 | In EvtGen but no trace elsewhere. |
| C2 | $D^+ \to \mu^+ X$ | 158.8 ± 2.7 | Sum from below, consistent with PDG inclusive. |
| C2.1 | $\overline{K}^0\mu^+\nu$ | 87.6 ± 1.9 | From PDG $\Gamma_{17}$. |
| C2.2 | $\overline{K}^{*0}\mu^+\nu$ | 52.7 ± 1.5 | From PDG $\Gamma_{29}$. |
| C2.3 | $\pi^0\mu^+\nu$ | 3.5 ± 0.15 | From PDG. |
| C2.4 | $\overline{K}_1^0\mu^+\nu$ | 2.77 ± 0.40 | From PDG $\Gamma_{31}$, smilar electron channel times BR. |
| C2.5 | $\overline{K}_2^{*0}\mu^+\nu$ | 1.0 ± 1.0 | In EvtGen but no trace elsewhere. |
| C2.6 | $\rho^0\mu^+\nu$ | 2.18 ± 0.20 | From PDG similar electron channel. |
| C2.7 | $\omega\mu^+\nu$ | 1.69 ± 0.11 | From PDG similar electron channel. |
| C2.8 | $\eta\mu^+\nu$ | 1.11 ± 0.07 | From PDG similar electron channel. |
| C2.9 | $\eta'\mu^+\nu$ | 0.20 ± 0.04 | From PDG similar electron channel. |
| C2.10 | $\pi^-\pi^+\mu^+\nu$ | 0.63 ± 0.05 | From PDG similar electron channel. |
| C2.11 | $K^-\pi^+\mu^+\nu$ | 1.9 ± 0.5 | Non resonant, from PDG $\Gamma_{28}$. |
| C2.12 | $\overline{K}^0\pi^0\mu^+\nu$ | 0.95 | Isospin from above. |
| C2.13 | $\mu^+\nu$ | 0.37 ± 0.02 | From PDG. |
| C2.14 | $\tau^+\nu$ | 0.20 ± 0.05 | From PDG, includes $\tau \to \mu\nu\nu$. |
| C3 | $D_s^+ \to \mu^+ X$ | 75.40 ± 2.06 | From sum of below, not present in PDG. |
| C3.1 | $\phi\mu^+\nu$ | 23.9 ± 1.6 | From PDG $e$ ($\Gamma_{24}$), more precise than $\mu$ ($\Gamma_{25}$). |
| C3.2 | $\eta\mu^+\nu$ | 23.2 ± 0.8 | From PDG similar electron channel. |
| C3.3 | $\eta'\mu^+\nu$ | 8.0 ± 0.7 | From PDG similar electron channel. |
| C3.4 | $\overline{K}^0\mu^+\nu$ | 3.4 ± 0.4 | From PDG similar electron channel. |
| C3.5 | $\overline{K}^{*0}\mu^+\nu$ | 2.15 ± 0.28 | From PDG similar electron channel. |
| C3.6 | $\tau^+\nu$ | 9.31 ± 0.39 | From PDG, includes $\tau \to \mu\nu\nu$. |
| C3.7 | $\mu^+\nu$ | 5.49 ± 0.16 | From PDG. |

table presents several unmeasured decay modes with a $K^*$ in the final state which are anyways present in EvtGen and are here estimated as half of the similar $K$ mode as EvtGen does. In order to account for this uncertainty, a 50% uncertainty on those contributions, uncorrelated among those decays, will be added when fitting the data. Furthermore, it is interesting to notice that it is not possible to chance sign/conjugate to the strange meson where present because its flavor is determined by the $b$ quark. The generator cuts requirements applied are:

- At least one muon with $|\eta| < 1.6$ and $p_T > 6.7$ GeV.
- At least one $D^{*-}$, coming from the relevant $B$ meson and decaying to a $\overline{D}^0\pi^-$. $\overline{D}^0$ and $\pi^-$ are required to have $|\eta| < 2.5$ and, respectively, $p_T > 0.5$ GeV and $p_T > 0.25$ GeV.
- At least one $\overline{D}^0$, coming from a $D^{*-}$ and decaying to a $K^+\pi^-$. Both $K^+$ and $\pi^-$ are required to have $|\eta| < 2.5$ GeV and $p_T > 0.5$ GeV.

All the samples have been generated with the custom `c2` pileup profile.

### $B \to J/\psi K^{(*)}$ simulation

These samples are used to extract the trigger efficiency scale factors as explained in Sec. 4.4. When generating $B^0 \to J/\psi K^*$ samples, the decay of all $B^0$ mesons is forced to the $J/\psi K^{*0}$ final state using the amplitude option `SVV_HELAMP` with the standard parameters specified in the default decay table. $J/\psi$ produced in $B^0$ decays are forced to decay to a muon pair according to the amplitude option `VLL` and using Photos to simulate FSR. $K^{*0}$ produced in $B^0$ decays are forced to decay to $K^+\pi^-$ thought the `VSS` amplitude option. The product of the branching ratios enforced is $6.00 \pm 0.16 \cdot 10^{-5}$. A similar generation is set up to produce $B^+ \to J/\psi K^+$ events, which have a product of the branching ratios enforced equally to $5.96 \pm 0.30 \cdot 10^{-5}$. The cuts applied at the generator level are:

- At least one $B$ meson decaying into a $J/\psi$ and a $K^{(*)}$.
- Where present, $K^{*0}$ decaying to $K^+\pi^-$, with both final state mesons satisfying $p_T > 0.5$ GeV and $|\eta| < 2.5$.
- The above $J/\psi$ decaying to a muon pair, with both muons satisfying $p_T > 1$ GeV and $|\eta| < 2.5$.

- At least one muon with $p_T > 6.7$ GeV and $|\eta| < 1.55$.

This sample was initially also mean to be used as control for the B mesons kinematic but internal studies are still trying to establish the feasibility of this calibration.

## 4.3 Candidate Selection

As discussed in Section 4.1, both $B \to D^* \ell \nu_\ell$ signal processes involved in $\mathcal{R}(D^*)$ present a similar visible final state consisting of a opposite charge $D^*$ and $\mu$ pair. Events from both processes are selected by a single procedure, creating a set of *candidates*, whose different populations will be separated by a fit to final state kinematic observables. Candidates are reconstructed by looking for the $D^*$ mesons decaying solely in the specific chain $D^{*-} \to \pi^- D^0 \to \pi^- K^+ \pi^-$ (charge conjugation is intended here and will be implicit in the following unless differently specified). It has to be noticed that final state particles are such that all pions have the same charge of the opposite sign with respect to the kaon one which is forced to have the same charge sign as the muon. This observation about the charges is crucial to making a mass hypothesis of the tracks involved in the decay without using additional particle ID techniques once the charge of the muon is fixed. Muons, charged pions, and kaons are directly measured by the CMS detector and reconstructed by the CMS software [40]. In this section, the procedure followed to select $B$ candidates and compute relevant quantities is described. As a reminder, the events corresponding to the charge conjugate process are also selected by repeating the same procedure with the opposite charge signs.

Given the data stream chosen [25], each event can be pictured as a $pp$ interaction where at least two bottom hadrons are produced and at least one of them undergoes a decay with a muon in the final state. In this analysis is chosen to look for $B^0 \to D^{*-} \ell^+ \nu_\ell$ decays on the *tag* side, defined as the one containing the muon which triggered the event. This side has been chosen over the other one (*probe* side) for two main reasons: to increase the number of events and to select $B^0$ with a higher $p_T$. The first one, which is important to reduce the statistical uncertainty, is achieved because the tag side has, with high purity, a $B$ semi-leptonic decay which otherwise would represent only about 10% of the total decays in the probe side. Moreover, the muon from the $B$ decays, since has to satisfy the trigger requirements, will more likely satisfy also the requirements to be identified as a muon by the CMS reconstruction software. For the same trigger requirements, the muons from tag side $B$ mesons have a harder $p_T$ spectrum, hence their selection also bias the $B$

(and the other mesons produced) towards the harder side of the spectrum increasing their acceptance and reconstruction rate. It has also to be noticed that it is easier to produce an effective MC for the tag side since the cuts can directly and without ambiguities be applied to the side in which the interesting decay is forced to happen. Unless specified, the procedure and selections described in the following are applied in the same way by the same code to both MC and data.

**MINIAOD Processing and Preliminary Selections**

Events are analyzed in CMSSW 10.2.3 starting form the MINIAOD [41] format using the configurations and plugins from `https://github.com/ocerri/BPH_RDntuplizer`. The processing of each event starts with the creation of a set of possible vertexes where the original $pp$ interactions have happened. To do this we start from the `offlineSlimmedPrimaryVertices` collection present in the MINIAOD and, using the adaptive vertex fitting [42] algorithm, we refit the vertex position excluding all the tracks that do not pass the tight PV assignment requirements [43] or have a deterministic annealing weight below 0.5. Only the vertexes selected by requiring a converging fit and more than 4 degrees of freedom will be considered in the following as possible primary vertexes (PV). Events with no selected vertex are discarded.

After the creation of the PV set, a set of possible muons is created starting from the collection `slimmedMuons` present in MINIAOD. To be accepted in this set, a muon has to satisfy the following requirements:

- $|\eta| < 1.5$, consistently with the L1 trigger requirements.
- Have a valid match with a track in the inner pixel tracker, crucial to have a meaningful vertex fit with the eventual $D*$ candidate.
- Match a trigger object from at least one of the relevant ParkingBPH HLT paths. Practically, the function `pat::Muon::triggered` has to return true when evaluated with the string `HLT_Mu*_IP*_part*_v*`.
- Pass the Medium ID [44, 45], which is a selection optimized to select prompt muons and muons from heavy flavor decay while rejecting in-flight mesons decay through a kink-finding algorithm.

In the following of the selection process description, we will assume for better clarity that the charge of the muon is positive but a charge conjugate process is applied for negatively charged muons. For each accepted muons, the particle flow (PF) candidates from `packedPFCandidates` collection of MINIAOD are looped through 3 times to find all the possible combinations of charged mesons that can create a $B^0$ candidate with that $\mu^+$. During these loops, a re-fit, called *vertex fit*, involving the selected tracks will be performed several times in order to test their provenance from a common vertex, find the most likely decay vertex position, and decrease the track parameters uncertainty. This refit is done using the CMSSW KinematicVertexFit module [46, 47] which, starting from transient tracks built out of the best track four-momentum(`pat::Muon::muonBestTrack` for muons, `pat::PackedCandidate::bestTrack` for $\pi$ and $K$), perform a pure geometrical fit to the linearized tracks with the additional constraint that they have to come from a common vertex. When the fit converges (`KinematicTree::isValid`), the results include the $\chi^2$ of the fit, the common vertex and its uncertainty, and a post-fit state of all the particles involved. It has to be noticed that KinematicVertexFit would provide also an option to add a mass constrain in the fit. However, this option has not been used in order to retain the discriminating power of the candidate invariant mass and because non-physical behavior of the mass constraints implementation has been observed on several occasions (but has not been chased down because out of the scope of this analysis). The implementation of the kinematic fits used can be found at `https://github.com/ocerri/BPH_RDntuplizer/blob/master/plugins/VtxUtils.cc`.

First, two loops try to reconstruct all the possible $\overline{D}_0 \to \pi^- K^+$ candidates near the muon. The first loop looks for the $K^+$ meson, selecting PF candidates that satisfy:

- Positive charged hadrons PF id (211) with track details.
- Have a valid match with a track in the inner pixel tracker.
- $p_T > 0.6$ GeV, to avoid fake tracks.
- $\Delta z < 1.5$ cm between the muon and $K^+$ tracks at the closest approach to the beam spot line (see `https://github.com/ocerri/BPH_RDntuplizer/blob/master/plugins/B2DstMuDecayTreeProducer.cc#L169`).
- $\Delta R < 2$ from the muon to reduce the combinatorics background[4].
- Transverse impact parameter significance from the PV $d_{xy}/\sigma_{d_{xy}} > 2$, in order to require the displacement originated by the $B$ and $D$ lifetimes.

[4] $\Delta R = \sqrt{\Delta\varphi^2 + \Delta\eta^2}$.

If a PF candidate passes this selection, is called a $K^+$ candidate and a mass $m_K = 0.493677$ GeV is assumed. In the second loop, nested in the previous one, we look for the $\pi^-$ meson. It aims to select PF candidates that satisfy the same requirements of the $K^+$ selection, with the exception of the PF id which is required to be the one of a negatively charged hadron (-211). Accepted PF candidates are called $\pi^-$ candidate and a mass $m_\pi = 0.13957018$ GeV is assumed. Before accepting the $\pi^- K^+$ pair, a vertex fit is performed. The $\pi^- K^+$ pair is accepted and a $\overline{D}_0$ candidate is created if:

- The fit $\chi^2$ probability is above 0.5%.
- $|m_{\pi K} - m_{D^0}| < 50$ MeV, where $m_{\pi K}$ is the pair invariant mass after the fit and $m_{D^0} = 1.864$ GeV is the world average for the $D^0$ meson mass measurements [1].

It has to be noticed that the requirement on $m_{\pi K}$, not only rejects combinatorial $\pi K$ pairs but also discards most of the cases where the particle ID/mass hypothesis is wrong (e.g. cases in which the positive track is not a kaon). To reconstruct the $D^{*-} \rightarrow \overline{D}_0 \pi^-$ decay, a third and last nested loop over PF candidates is done to look for a second $\pi^-$. Since this last $\pi$ is expected to be softer than the first one, due to the low $q^2$ of the $D^*$ decay, it will be referred to as *soft* pion. The same selection as the one for the first $\pi$ is applied to select soft $\pi$. For each selected candidate, a vertex fit to the $\overline{D}_0 \pi^-$ and $\overline{D}_0 \pi^- \mu^+$ systems is performed. The following cuts are applied:

- The $\overline{D}_0 \pi^-$ vertex fit $\chi^2$ probability is above 0.5%.
- $|m_{\overline{D}_0 \pi} - m_{D^{*+}}| < 150$ MeV, where $m_{D^{*+}} = 2.01026$ GeV is the world average for the charged $D^*$ meson mass measurements [1].
- $\Delta m - (m_{D^{*+}} - m_{D^0})| < 3$ MeV, where $\Delta m = m_{\overline{D}_0 \pi} - m_{\pi K}$
- The $\overline{D}_0 \pi^- \mu^+$ vertex fit $\chi^2$ probability is above 0.5%.
- $m_{\overline{D}_0 \pi \mu} < 8$ GeV, in order to reduce the combinatorial background.

If those requirements are satisfied, the $\overline{D}_0 \pi^- \mu^+$ is called a $B^0$ candidate. For each candidate, the origin PV is chosen as the primary vertex with the higher pointing defined as the cosine between the visible system 4-momentum and the line going from the primary vertex to the B decay vertex. No further cuts are applied at this level. This selection procedure then identifies zero or more $B^0$ candidates per event, each made from a set of visible particles which differs from the other ones by at least one reconstructed particle.

Once the $B^0$ candidate is selected, the PF collection is looped once again (avoiding the already selected particles) to look for the possible additional track(s) that might be compatible with the $\overline{D}_0\pi^-\mu^+$ vertex. The presence of these tracks can be used to discriminate background processes such as $B \to D^{**}\mu\nu$. Additional requirements will be applied in the following to suppress background contributions in the signal region of the final fit. A PF candidate $q$ is considered a possible additional track if:

- Charged hadrons PF id (±211) with track associated with inner pixel hits.
- $p_T > 0.3$ GeV.
- $\Delta z < 1.5$ cm between the muon and candidate tracks at the closest approach to the PV.
- $\Delta R < 2$.
- Vertex fit to the $\overline{D}_0\pi^-\mu^+q$ system with $\chi^2$ probability greater than 0.5%.
- $m_{\overline{D}_0\pi^-\mu^+q} < 8$ GeV.

Furthermore, the hadronic mass for the $B$ candidate with the candidate $q$ is defined as $m_{had} = \text{mass}(\overline{D}_0\pi^-q)$. For each $B^0$ candidate, when processing MC samples, the additional information including details of the true generated event at the MC level is also stored.

**Final State Observables**

The three final state observables used in the fit to the signal region are: $q^2$, $m^2_{\text{miss}}$ and $E^*_\mu$. Considering a $D^*\mu$ pair coming from the decay of a $B$, the quantities are defined as follow:

$$q^2 = (P_B - P_{D^*})^2$$
$$m^2_{\text{miss}} = (P_B - P_{D^*} - P_\mu)^2$$
$$E^*_\mu = \left[\Lambda_{(\vec{\beta})}P_\mu\right]_0$$

where $P$ are the 4-momenta measured in the CMS experiment frame, $\Lambda$ is the Lorentz transformation matrix, $\vec{\beta} = \vec{P}_B/m_B$ is the $B$ velocity, and $E^*_\mu$ is, in words, the energy of the $\mu$ in the rest frame of the $B$.

It is clear that to compute all three of these quantities it is necessary to know the $B$ four-momentum. One can assume that the mass of the $B$ meson is $m_B = 5.2796$ GeV since its width is negligible compared to the experimental resolution. However, given only the $B$ production vertex, the *visible* system $D^*\mu$ and knowing that at least one other invisible particle is produced in the decay, there are not enough constraints to define a unique solution for $\vec{p}_B$. The $B$ meson flight direction, $\hat{n}_B = \vec{p}_B/p_B$, is measured directly from the relative position of the PV ($\vec{r}_{\rm PV}$) and the visible system vertex ($\vec{r}_{\rm vis}$) as:

$$\hat{n}_B = \frac{\vec{r}_{\rm vis} - \vec{r}_{\rm PV}}{|\vec{r}_{\rm vis} - \vec{r}_{\rm PV}|}.$$

On the contrary, the momentum magnitude $p_B$ can not be measured directly and will be estimated using the following *transverse approximation*. Let $P$ and $P^*$ be the 4-momenta of the visible system respectively in the lab frame and in the $B$ rest frame, then $P = \Lambda(-\vec{\beta})P^*$. Let's consider a frame rotated in the transverse plane defined by the versors $(\hat{a}_T, \hat{a}_\perp, \hat{a}_z)$, such that the first axis $\hat{a}_T$ is aligned with $\vec{p}_T(B)$. In this frame, the previous relation takes the explicit form

$$\begin{bmatrix} E \\ P_T \\ P_\perp \\ P_z \end{bmatrix} = \begin{bmatrix} \gamma & \gamma\beta n_T & 0 & \gamma\beta n_z \\ \gamma\beta n_T & 1 + (\gamma-1)n_T^2 & 0 & (\gamma-1)n_T n_z \\ 0 & 0 & 1 & 0 \\ \gamma\beta n_z & (\gamma-1)n_T n_z & 0 & 1+(\gamma-1)n_z^2 \end{bmatrix} \begin{bmatrix} E^* \\ P_T^* \\ P_\perp^* \\ P_z^* \end{bmatrix} \tag{4.4}$$

where $\vec{n} = \vec{\beta}/\beta$ and $\gamma$ is the Lorentz factor of the $B$ meson. Given the $q^2$ of the $B$ decay and the masses of the particles, the visible mass $m_V$ is consistently bigger than $P^*$. As shown in Figure 4.6, the relation $\delta = \dfrac{|\vec{P}^*|}{m_V} \sim 10^{-1}$ holds true. Considering that $\gamma\beta n_T = P_T^{(B)}/m_B$ and expanding $E^*$ at the first order in $\delta$, the second row can be rewritten as

$$\frac{P_T}{m_V} = \frac{P_T^{(B)}}{m_B} + \frac{P_T^{(B)}}{2m_B}\delta^2 + \left[1 + (\gamma-1)n_T^2\right] k_T\delta + (\gamma-1)n_T n_z k_z \delta \tag{4.5}$$

where $\vec{k} = \vec{P}^*/|\vec{P}^*|$. Dropping all the terms in $\delta$, we obtain

$$P_T^{(B)} = P_T \frac{m_B}{m_V}. \tag{4.6}$$

With Eq. 4.6 and the measured $B$ flight direction, it is possible to estimate the momentum of the $B$ meson. It has to be noticed that $P_T$ represents the transverse momentum of the visible system along the $B$ transverse momentum. However, given the small $\delta$ approximation the two are aligned (Figure 4.6 right) and the total

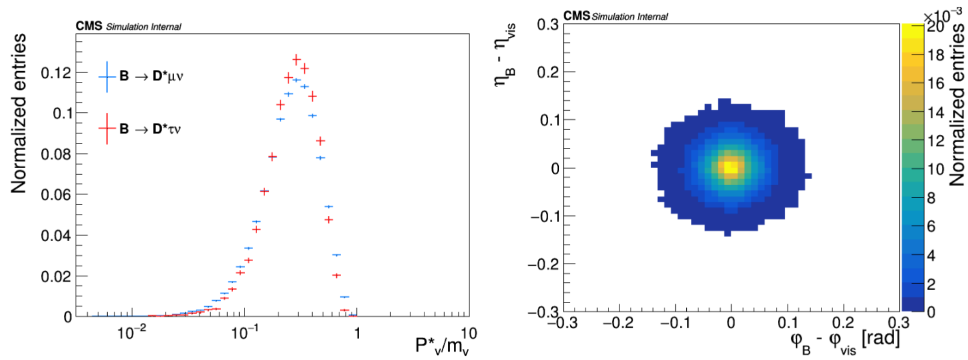


Figure 4.6: Distribution of $P^*/m_V$ for simulated signal sample (left) and angular distance between $B$ and visible momentum (right). Both distributions are for MC truth level variables after the selection for the low category (see Section 4.3), which represent the worse case scenario.

transverse momentum can be used. The transverse approximation presented above differs from the one used by LHCb [15]. Figure 4.17 shows the resolution on the $B$ meson flight direction and the goodness of the transverse approximation in MC events.

In the special case of $B \to D^*\ell\nu_\ell$, with possibly $\ell \to \mu\nu\nu$, the quantities can be re-written as: $q^2 = (P_\ell + P_{\nu_\ell})^2$ and $m^2_{\text{miss}} = (\sum P_\nu)^2$. It is then possible to understand how: $B \to D^*\tau\nu$ events will populate the region with $q^2 \gtrsim 4$ GeV, due to the non-negligible mass of the charged lepton; Events from $B \to D^*\mu\nu$, will have $m^2_{\text{miss}} \approx 0$ with deviations due to experimental effects; and, $B \to D^*\tau\nu$ will have a lower $E^*_\mu$ due to the presence of additional neutrinos.

**Categories and Final Selection**

After the MINIAOD events are analyzed and stored in flat ntuples, a second selection is applied to further increase the purity and enhance background rejection. This refined selection is postponed to this later stage in order to better control the tuning of the cuts. The following requirements are applied to the $B^0$ candidates selected during the processing of MINIAOD:

- p-value of the $\pi K$ vertex $\chi^2$ greater than 5%
- For both $\pi$ and $K$ forming the $\overline{D}_0$:
  - $p_T > 0.8$ GeV
  - $|\eta| < 2.4$

  - $\Delta R > 10^{-3}$ from $\mu$, to rejected split tracks
- $d_{xy}/\sigma_{d_{xy}} > 2$ for the $\pi K$ vertex w.r.t. the PV
- For the soft $\pi$:
  - $|\eta| < 2.4$
- p-value of the $\overline{D}_0\pi$ vertex $\chi^2$ greater than 5%
- $|m(\overline{D}_0\pi) - m_{\pi K} - (m_{D^{*+}} - m_{D^0})| < 2$ MeV
- p-value of the $\overline{D}_0\pi\mu$ vertex $\chi^2$ greater than 5%
- $\cos(\Delta\alpha_{PV}^{\text{vis}}) > 0.99$ [5], where $\Delta\alpha_{PV}^{\text{vis}}$ is the angle between the visible system momentum and the displacement vector from the PV to the visible system vertex (i.e. the $B$ flight direction).
- $-2\ \text{GeV}^2 < q^2 < 12\ \text{GeV}^2$
- $M^2_{miss} > -2.5\ \text{GeV}^2$
- $m(\overline{D}_0\pi\mu) = m_V < 5.4$ GeV
- Additional tracks are considered "good" if:
  - p-value of the $\overline{D}_0\pi q$ vertex $\chi^2$ greater than 5%
  - $p_T > 0.55$ GeV
  - $|\eta| < 2.4$
  - The visible mass with the additional track is $m(\overline{D}_0\pi\mu q) < m_{B^0}$
  - The pointing to the PV above 0.95.

The number and charge of good tracks are stored and later used to divide the events into regions. While the signal processes are going to populate mostly the region with no good additional tracks (signal region), the regions (control regions) with additional track(s) will be a major constraint to normalizing the background processes. The selected $B^0$ candidates are further divided into three different and disjoint categories. The categories are defined by requiring a specific trigger and consistent kinematics cuts on the muon. Table 4.7 shows the categories definition details and, later on, Fig. 4.15 shows how data are distributed among the categories.

Table 4.7: Definition of the three categories used in the analysis. Each category is specified by a required HLT path for the muon matching trigger object, a minimum transverse impact parameter significance (min $\mathrm{IP}_\mu$), and a muon transverse momentum ($\mu p_T$) range. The requirements on min $\mathrm{IP}_\mu$ and $\mu$ $p_T$ are chosen in order to avoid the turn-on effects of the trigger. It has to be noticed that categories have no overlap because the $p_T$ requirements are disjoint. What is more, since the trigger paths are turned on progressively, this also ensures that no event is wasted.

| Cat. name | Low | Mid | High |
|---|---|---|---|
| HLT path | `HLT_Mu7_IP4` | `HLT_Mu9_IP6` | `HLT_Mu12_IP6` |
| min $\mathrm{IP}_\mu$ | 5 | 7 | 7 |
| $\mu p_T$ [GeV] | $7.2 < p_T < 9.2$ | $9.2 < p_T < 12.2$ | $p_T > 12.2$ |

At this stage, the trigger muon is also required to match a muon reconstructed in the L1 trigger with a $p_T$ above the category HLT path requirement. This matching is necessary in order to ensure an accurate simulation of the trigger process. In the MC, indeed, the L1 seeds and the HLT path are all always on while in data present a turn-on schedule dependent on the LHC instant luminosity. The muons are also required to have $|\eta| < 0.8$ in order to exclude the ones reconstructed in the endcap region for which additional validation is required. After this final selection, the total number of selected candidates in data is about 500k, 300k, and 150k respectively for the high, mid, and low category. A non-negligible number of events coming from $B \to D^*\mu\nu$ decays are expected to still be present in the phase space where $B \to D^*\tau\nu$ decays group. The expected sensitivity for the different categories presents a tradeoff between the higher signal-to-noise ratio in the low category (Figure 4.7) and the higher statistic of the high category due to the greater integrated luminosity.

### MC and Data Corrections

As a first test after the final selection, it is useful to compare data and MC distributions of invariant masses for fully reconstructed charmed mesons. $B^0 \to D^{*-}\mu\nu$ is used as a reference for MC since no difference is expected for the charmed mesons mass shape between the different simulated processes, it has the highest expected population, and it has the highest generated statistics. Figure 4.8 shows the distributions of reconstructed mass for the $D^0$ meson (i.e. the $\pi^-K^+$ invariant mass), the $D^{*-}$ meson (i.e. the $\pi^-K^+\pi^-$ invariant mass) and their difference (i.e. $m(\pi^-K^+\pi^-) - m(\pi^-K^+)$).

[5] The name *pointing cosine* will be used in the rest of this document to refer to such quantity.

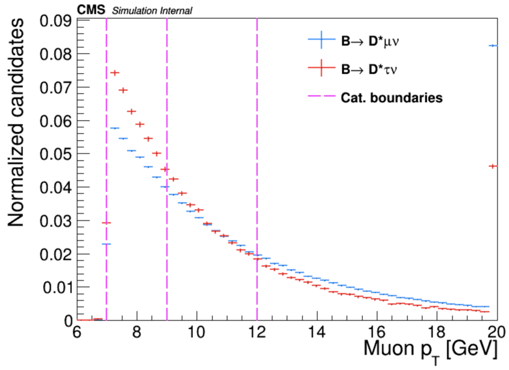


Figure 4.7: Distributions of muon $p_T$ in MC events for $B^0 \to D^*\mu\nu$ and $B^0 \to D^*\tau\nu$ before category selection.

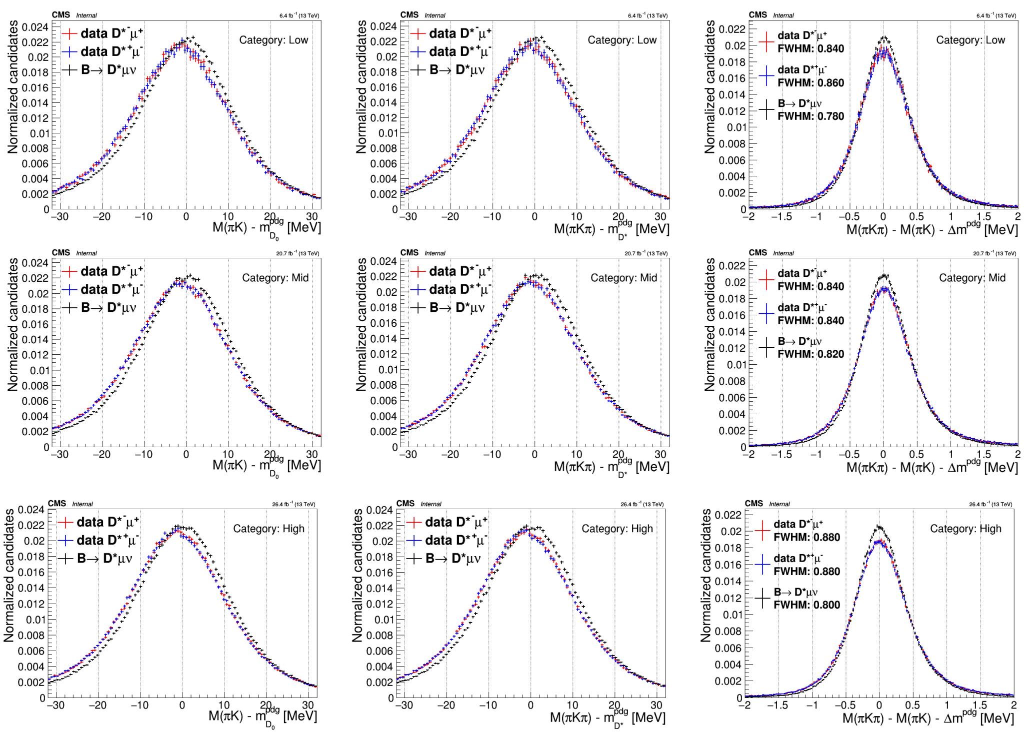


Figure 4.8: Distributions of reconstructed mass for the $D^0$ meson (i.e. the $\pi^- K^+$ invariant mass), the $D^{*-}$ meson (i.e. the $\pi^- K^+ \pi^-$ invariant mass) and their difference (i.e. $m(\pi^- K^+ \pi^-) - m(\pi^- K^+)$) in data and MC. The best estimation of the parameters for each particle is used corresponding to the final state of $D^0$ vertex fit between for $\pi^-$ and $K^+$, and to the final state of the $D^0\pi\mu$ vertex fit for the soft $\pi^-$.

Remembering that the $D$ mesons mass is around 2 GeV and that $\Delta m^{pdg} = m_{D^*} - m_{D^0} = 145$ MeV, it is clear that both $m(\pi^- K^+)$ and $m(\pi^- K^+ \pi^-)$ present a relative shift between data and MC of the order $10^{-3}$. At the same time, all of them present a difference in width of the order $10^{-3} - 10^{-4}$. It is possible to demonstrate [48] that none of these effects can be due to a reasonable efficiency mis-modeling or difference in the particles $p_T$ and $\eta$ spectrum. Considering the phase space of the tracks entering in the reconstruction (Figure 4.9), it possible to associate the observed discrepancies with a mis-modeling of the momentum scale.

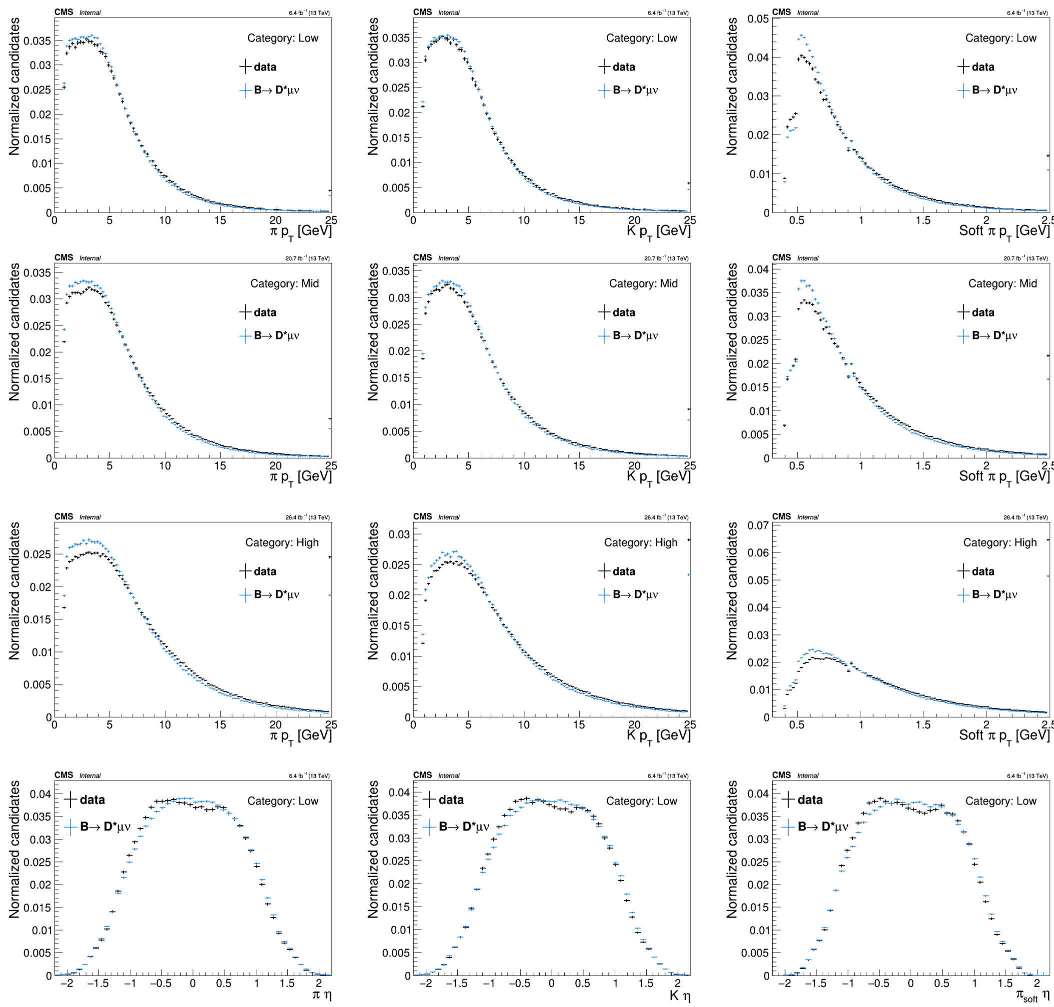


Figure 4.9: Distribution of $p_T$ and $\eta$ spectrum for $\pi^-$, $K^+$ and soft $\pi^-$ for selected $B$ candidates in data and MC for the three categories.

Three main effect can impact the momentum scale: tracker modules misalignments, magnetic field mis-modeling and energy loss mis-modeling. Given the relatively low $p_T$ of the tracks, it is possible to neglect misalignment's effects. A naive intuition can be gained by observing that our tracks have a curvature radius $\rho$ of the order 10 m or smaller. This means that, in the CMS tracker, their sagitta $s = L^2/8\rho$ is not smaller than a few millimeters whereas the tracker modules are aligned by the CMS standard algorithm [49] better than 10 $\mu$m. The magnetic field mis-modeling exists when the magnetic field value used during the reconstruction is different from the real one. This can only happen in data because in MC the same magnetic field map is used to propagate particles and reconstruct tracks. Given that $p_T$ [GeV] = $0.3B\rho$ [Tm], a small mis-modeling of the magnetic field $B \to (1+\varepsilon)B$ will affect the momentum measurement likewise $p_T \to (1+\varepsilon)p_T$. Since the invariant mass of two particles is roughly proportional to $p_T$, when the $B$ field mis-modeling is a dominant effect, one would expect approximately constant shift as a function of the tracks $p_T$. This is indeed what can be seen in Figure 4.10 left where the average $\pi K$ pair mass is plot against the tracks $p_T$. More systematically, considering the decay of a particle

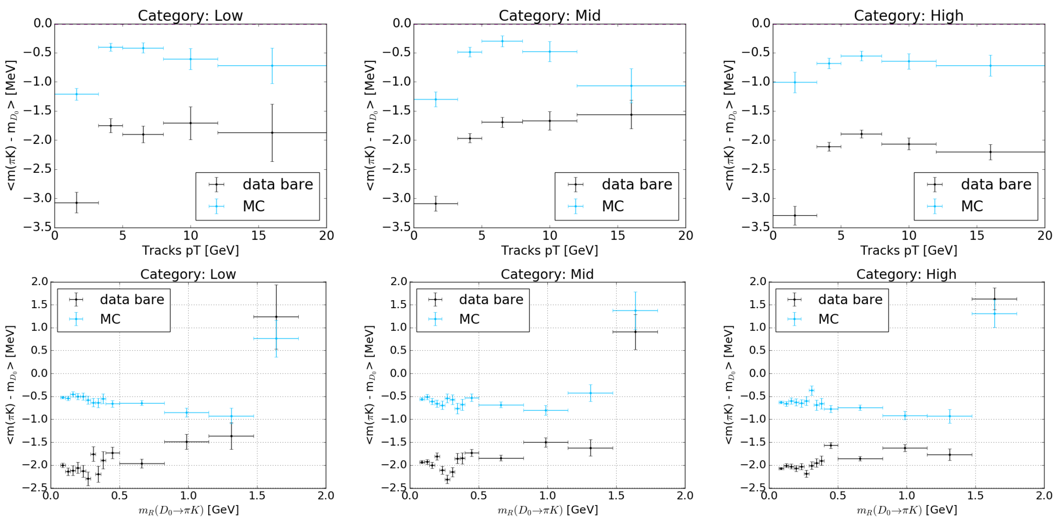


Figure 4.10: Top: Average $\pi K$ mass in data and MC as a function of tracks $p_T$. Bins are defined such that both tracks $p_T$ lies withint the bin boundaries.
Bottom: Average $\pi K$ mass in data and MC as a function $m_R$. The value of $q^2/M$ for $D^0 \to \pi K$ is equal to 1.72 GeV.

with mass $M$ into two particles with 4-momenta $p_1^\mu$ and $p_2^\mu$, it is possible to compute what is the effect of $\varepsilon$ on the reconstructed invariant mass. Expanding at the first order in $\varepsilon$, it is possible to demonstrate that

$$M(\varepsilon) = M - \varepsilon \left( m_R - \frac{q^2}{M} \right)$$

where $q^2 = M^2 - m_1^2 - m_2^2$ and

$$m_R \equiv \frac{1}{M} \left( m_1^2 \frac{E_2}{E_1} + m_2^2 \frac{E_1}{E_2} \right).$$

As a consequence, if the $B$ field mis-modeling is dominant, a plot of the average $m(\pi K)$ versus $m_R$ will show a linear behavior vanishing at $q^2/M = 1.72$ GeV. This is observed in Figure 4.10 right. What is more the slope of the data would be a measure of the magnetic field shift. A linear fit with the vanishing constraint performed on data estimate $\varepsilon$ to be around $6.8 \pm 0.2 \cdot 10^{-4}$ with a p-val = 0.18. This demonstrate that the B field mis-modeling is the dominant effect.
It is now possible to understand why the mass shift seems not to affect the $m(\pi^- K^+ \pi^-) - m(\pi^- K^+)$ distribution. The expression for $M(\varepsilon)$ can be rearranged as $M(\varepsilon) = M(1 + \varepsilon\alpha_1(p_1^\mu, p_2^\mu))$, where

$$\alpha_1 = \frac{\left(\beta_1^2 + \beta_2^2\right) E_1 E_2 - 2\vec{p}_1\vec{p}_2}{M} \xrightarrow{\beta \to 1} \frac{q^2}{M^2}.$$

The effect on the mass can then be expressed as $M(\varepsilon) \sim M(1 + \varepsilon\frac{q^2}{M^2})$. For the $D^0 \to \pi K$ decay, $q^2/M^2 \sim 1$, the effect is expected to be of the order $10^{-3}$ in agreement with the observation. In the $D^* \to \pi D^0$ decay, the $D^0$ mass is dominant and the same reasoning applies. However, the mass difference at the first order in $\varepsilon$ can be expressed as

$$m_{D^*}(\varepsilon) - m_D(\varepsilon) = m_{D^*} - m_D + \varepsilon\left[\alpha_1^{\pi K}\left(\frac{m_D^2}{m_{D^*}} - m_D\right) + \alpha_1^{\pi D^0} m_{D^*}\right].$$

In the approximation $m_D \sim m_{D^*}$, the contribution from the $D^0$ (term proportional to $\alpha_1^{\pi K}$) cancel out and the remaining contribution is proportional to

$$\alpha_1^{\pi D^0} \sim \frac{q^2_{(D^* \to \pi D^0)}}{m^2_{D^*}} \sim 10^{-1}$$

which results to be one order of magnitude smaller than the one on $m(\pi K)$ or $m(\pi K \pi)$.
In order to correct the effect of the magnetic field, we considered that, as reported in [50], the standard magnetic field map used in CMSSW is a 2D ($B$ as a function of $z$ and $r$) approximation of a more sophisticated 3D measurement ($B$ as a function of $z$, $r$ and $\phi$) performed on the CMS detector before lowering it down in the LHC cave. Figure 4.11 shows the ratio between the two maps as a function of $\eta$ and $\phi$. Besides the modulation in $\phi$ due to the return yoke of the magnet, the average value in the central region (where most of our tracks are) is about $1 - 7 \cdot 10^{-3}$ , in agreement with what we found in the previous analysis. As a consequence, the appropriate value retrieved from the 2D ratio histogram has beed decided to be used

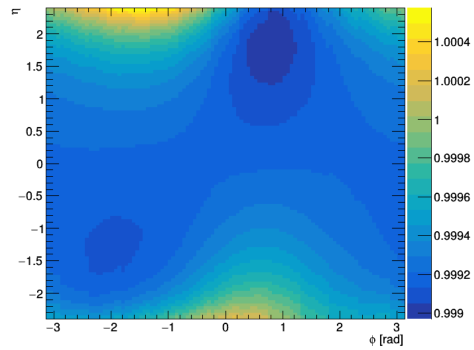


Figure 4.11: Ratio between the 2D default map used in CMSSW and the original 3D map measured.

as a corrections to all the tracks in data by scaling the $p_T$ and then recomputing all the relevant variables.

In order to mitigate the distribution width observed, an additional smearing has been applied on the MC tracks. An effective additional gaussian smearing of $6 \cdot 10^{-3}$ ($3 \cdot 10^{-3}$) has been applied to all MC PF candidates with $p_T$ below (above) 1 GeV.

Figure 4.12 shows the mass distributions after the corrections have been applied, whereas Figure 4.13 shows the average mass as a function of relevant variables. The agreement achieved after these corrections is at the limit of the statistical precision. Small deviations from expected values are well modeled in MC and the agreement between data and MC is well beyond fractions of MeV or resolution. The corrections discussed above are applied as part of the final selection, before selecting the events. Unless differently specified the corrections are applied in the following. Despite their importance to understand the datasets and their impact on resonant masses, these corrections have a very limited effect on the observables used in the fit. Fig. 4.14 shows an example comparison of the distributions of the kinematic variables used in the fit for data and MC, before and after the corrections.

In order to study further corrections, effect of energy loss mis-modeling have also been explored. These effects depends on $\beta\gamma = P/m$ and hence affect generally pions and kaons in different ways. The considered effects are:

- Effect on kaons tracks due to the fact that the CMS reconstruction software subtract an estimation of the energy loss to the track energy and recomputes the momentum assuming the pion mass.

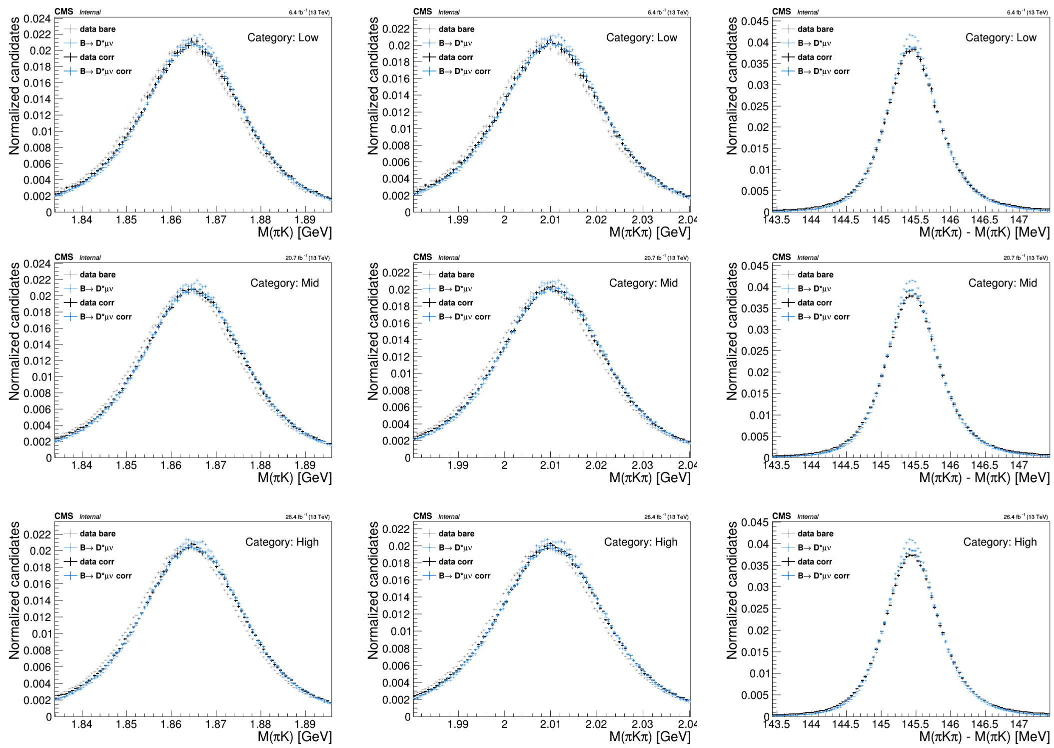

Figure 4.12: Masses distributions in data (before and after corrections) and MC (corrections do not affect the mean) as a function of tracks.

- Effects on both $\pi$ and $K$ due to a miscount of the material budget.
- Effect on $K$ due to the different energy loss when passing through materials with the respect of pions.

The approach used has been similar to the one used for the magnetic field, where the correction on the particle 4-momentum has been propagated at the first order to an effect on the mass. No relevant trend has been observed and no further study has been pursued.

**Relevant Distributions**

This subsection present several plots showing the general features of the reconstruction and selection above discussed.

First of all, we show the distribution in data of transverse momentum vs impact parameters for the $\mu$ belonging to the candidate (Fig. 4.15). Since in the analysis this muon is also required to be the triggering muon, this distribution is also a clear way to see how events are partitioned into the three categories: low, mid, and high.

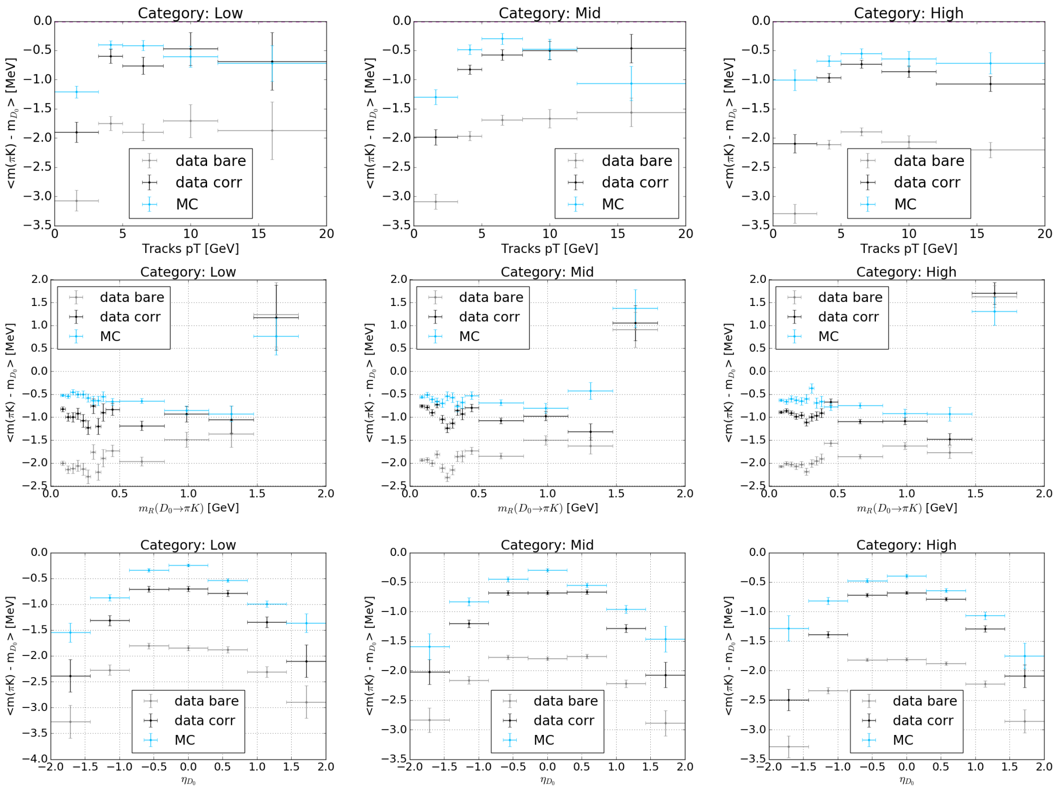


Figure 4.13: Average $\pi K$ mass in data (before and after corrections) and MC (corrections do not affect the mean) as a function of tracks $p_T$, $m_R$, and $\eta$.

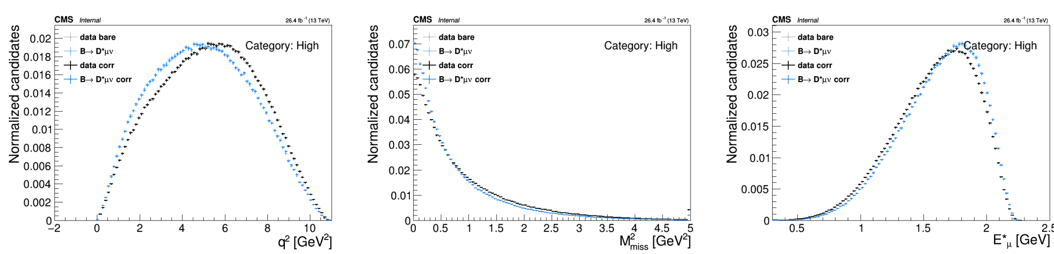


Figure 4.14: Comparison of the distributions of the kinematic variables used in the fit for data and MC, before and after the corrections. The impact of the corrections on these variables is very small and the distribution before and after the corrections overlap almost perfectly.

It is clear from the picture how the phase space region with $5 < IP < 7$ and $p_T > 9.2$ has a negligible amount of relevant candidates. During the reconstruction procedure

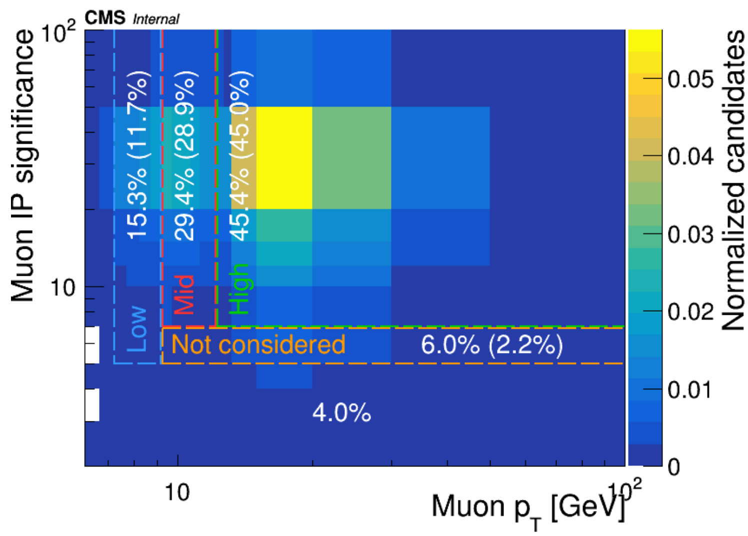


Figure 4.15: Data distribution transverse momentum vs impact parameters for the $\mu$ belonging to the candidate, no trigger requirement is applied to the muon for this specific plot. However, the relative population of each region is reported together with the fraction of candidates also satisfying the relevant trigger requirements. The latter is the number in brackets.

we fitted a secondary vertex between $D^0$, $\pi$ and $\mu$ under the assumption that they all are produced near the $B$ decay vertex. However, in the case of muons produced in the $B \to D^*\tau\nu$ decay an additional displacement is present due to the $\tau$ lifetime and the above procedure holds only thanks to the experimental resolution. In order to justify this treatment, we show in Fig. 4.16 left the distribution of the $\chi^2$ for the fit of the $D^0$, $\pi$, and $\mu$ vertex. The $\chi^2$ for the $B \to D^*\tau\nu$ sample results close to the one of a $\chi^2$ with the degrees of freedom of the fit. In Fig. 4.16 right the muon impact parameter (IP) from B decay vertex in MC $B \to D^*\tau\nu$ events is also shown. The true MC generated values is compared to the IP estimated in the reconstruction considering the muon track before any re-fitting with the vertex obtained fitting $D^0\pi$ (1) and $D^0\pi\mu$ (2). In approximation (1) the experimental resolution dominates the distribution resulting in a shift towards higher values. In (2) the bias coming from including the muon in the vertex fit leads to an underestimation of the IP.

The goodness of the $B$ meson approximation discussed in Section 4.3 can be validated using MC events. The resolution obtained in $B^0 \to D^*\mu\nu$ events is shown in Figure 4.17. The resolution on angular variables, depending mainly on the vertexing precison, is estimated to be slightly below $10^{-2}$. The relative $p_T$ resolution is esti-

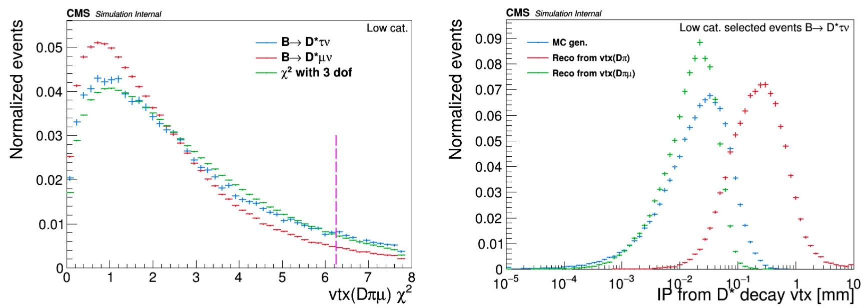


Figure 4.16: Left: Distribution of the $\chi^2$ for the fit of the $D^0$, $\pi$ and $\mu$ vertex in MC signal samples. The magenta dashed lines represent the selection requirement of a p-value better than 10%, corresponding to the quantile of the distribution with 3 degrees of freedom (same number of dof of the vertex refitting).
Right: Muon impact parameter from B decay vertex in MC $B \to D^{*}\tau\nu$ events. The MC true generated values is compared to the reconstructed value.

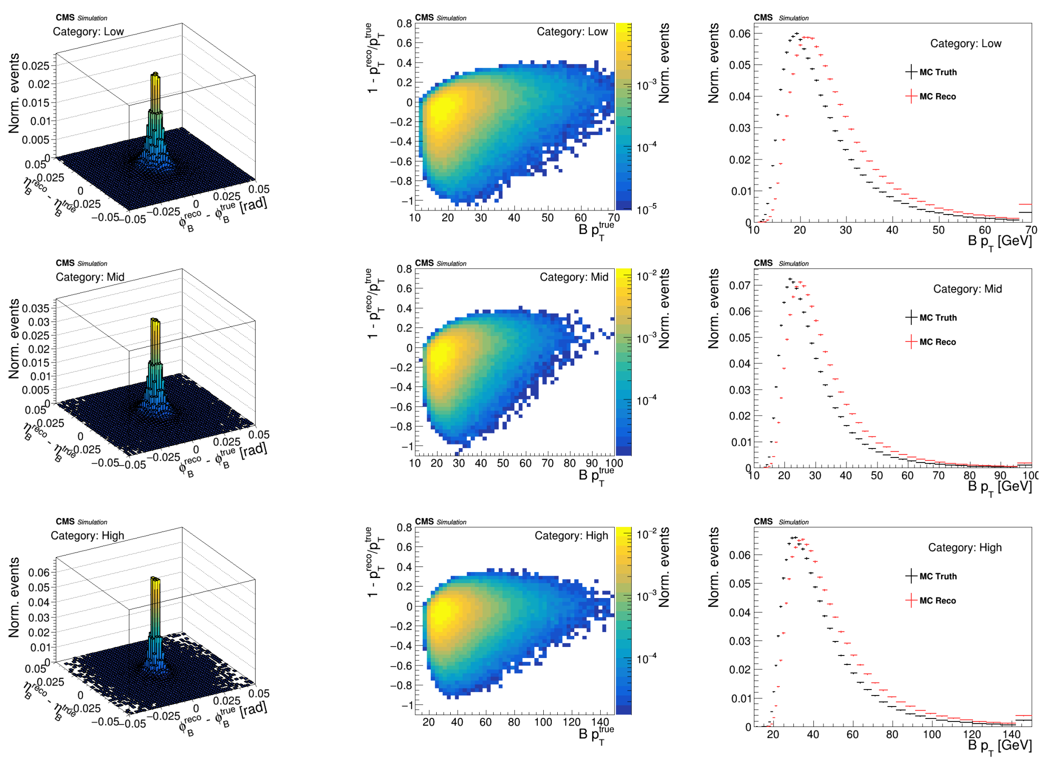


Figure 4.17: MC distribution from the $B \to D^{*}\mu\nu$ sample of: angular distance between reconstructed and true B direction (left); relative error of reconstructed B $p_T$ in bins of true B $p_T$ (center); and $p_T$ spectrum of B meson in MC truth and reconstruced (right).

mated to be of the order of 10%, driven by the dominant effect of the approximation. To get a naive intuition of the separation power, it is interesting to visualize what is the distribution of the kinematics variables that will be used to distinguish the two signal processes. Figure 4.18 shows the distribution of $q^2$, $m^2_{\text{miss}}$ and $E^*_\mu$ for

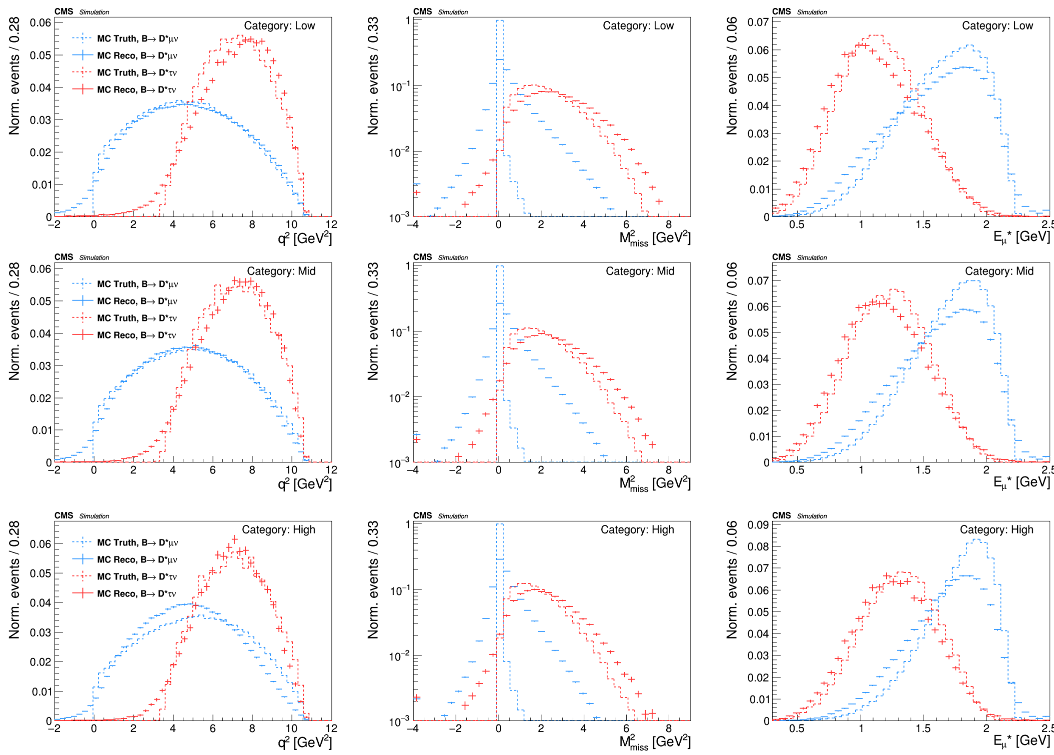

Figure 4.18: Distribution of $q^2$, $m^2_{\text{miss}}$ and $E^*_\mu$ for $B \to D^*\mu\nu$ and $B \to D^*\tau\nu$ MC samples using true (dashed) and reconstructed (solid) information.

$B \to D^*\mu\nu$ and $B \to D^*\tau\nu$ MC samples using true and reconstructed information. It can be noticed that the achieved resolution is enough not to introduce major distortion in the kinematic variables distributions. The only exception is the missing mass distribution for the $B \to D^*\mu\nu$ which in the MC truth is a delta at 0 with a small tail due to FSR. No major difference among the different categories have been observed in the plots shown above beside a moderate increase of the resolution in the higher categories of the order 10%. The distributions of muon impact parameter respect the D and $D^*$ reconstructed decay vertex for MC events with no cut on the $D^*\mu$ vertex p-value is shown in Fig. 4.19. It is clear how the separation power of this observable is negligible compared to the kinematic observables used in this analysis.

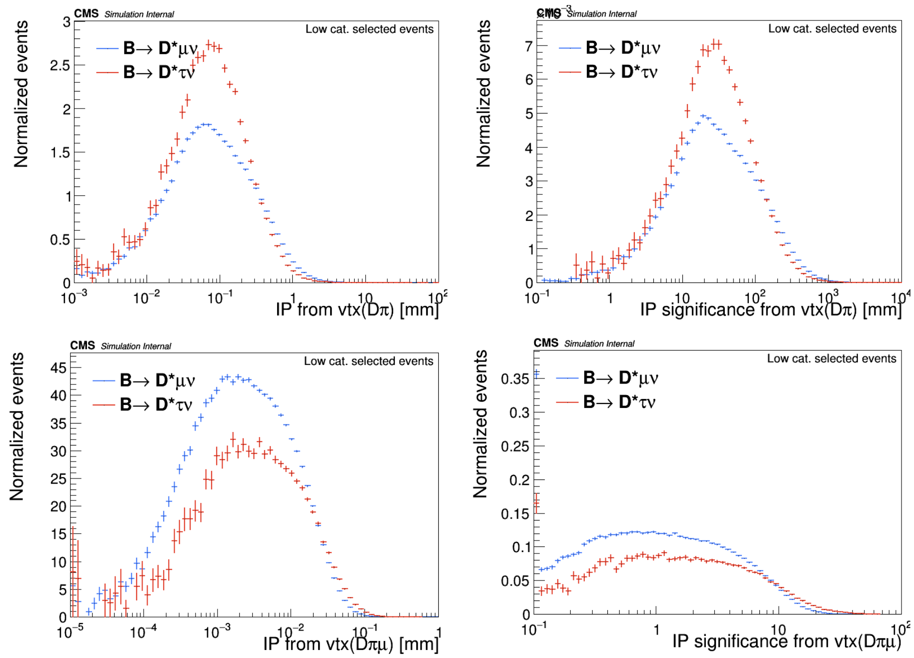


Figure 4.19: Distribution muon impact parameter respect the D (top) and $D^*$ (bottom) reconstructed decay vertex for MC events with no cut on the $D^*\mu$ vertex p-value.

Both charge conjugate configurations are considered in data and are analyzed in the same way with the obvious differences in the charge requirements. The distributions of the two configurations in data are found to be statistically compatible in all the observables. As an example we report in Fig. 4.20 a comparison the distributions for the kinematic observables used in the signal region fit. In the following data made by $D^{*-}\mu^+$ and $D^{*+}\mu^-$ candidates will be displayed together in the same histogram by simply adding the two contributions.

### Combinatorial Background

With *combinatorial background* we refer to the possibly that random tracks produced during the bunch crossing will satisfy the selection criteria even if not coming from the same specific process. The yield of this kind of background is reduced by quality requirements on the vertexing procedure and with cuts on observables like invariant masses and pointing angles. The remaining contribution is here estimated. The combinatorial background source we consider are three: two random tracks to be considered a $D^0$ meson which then get selected with a $\pi$ and a $\mu$ as a $B$ candidate; a true $D^0$ meson selected with a $\mu$ and a random $\pi$; and a true $D^*$ selected with a random muon.

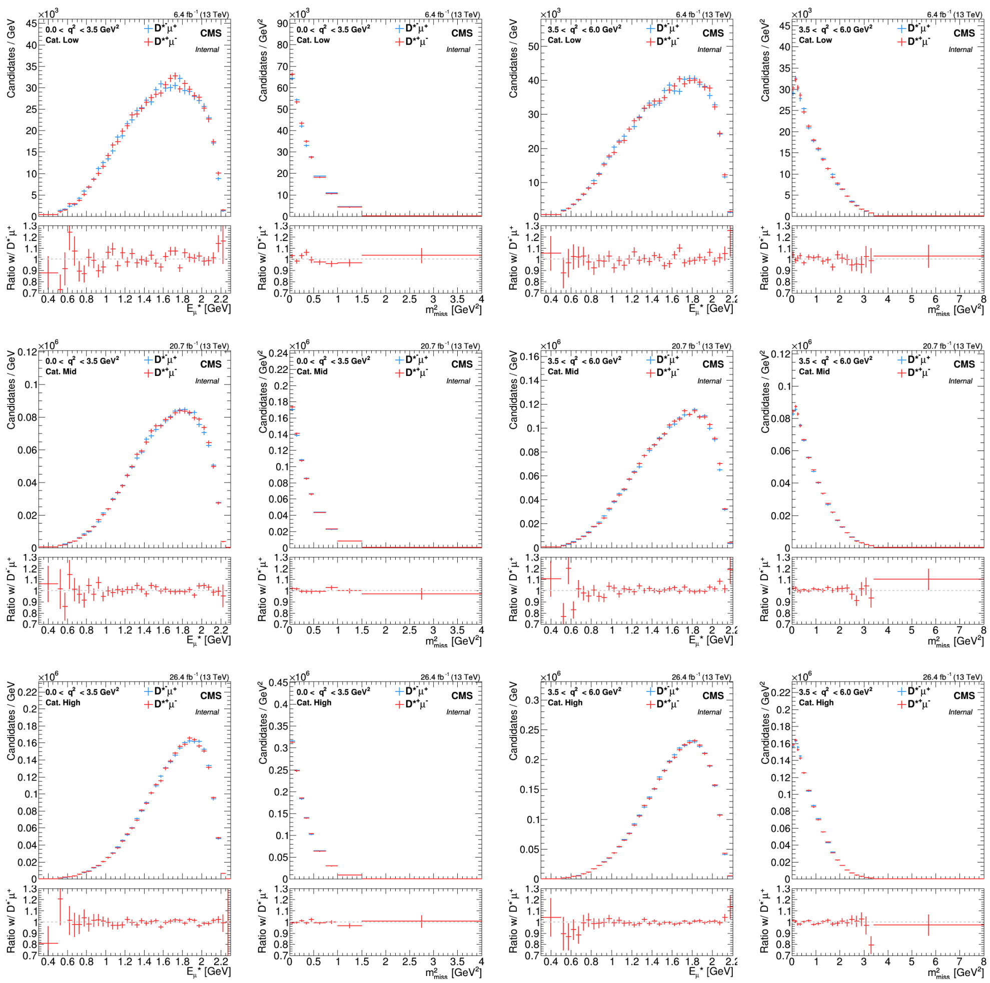

Figure 4.20: Distribution of the kinematic observables used in the signal region fit for $D^{*-}\mu^{+}$ and $D^{*+}\mu^{-}$ real data candidates.

Assessing the level of agreement of the plots in Figure 4.12 it is possible to conclude that no additional presence of the first two kind of combinatorial background can be observed. The shape agreement on those plots allow us to safely neglect these two backgrounds.

In order to study the presence of the third kind of combinatorial background, a special data sample is selected by inverting the charge requirement on the trigger muon such that the supposed $B$ candidate is made by a same sign couple ($D^{*-}\mu^{-}$ or $D^{*+}\mu^{+}$). While there is no relevant physical process that can produce such a final state without the presence of other additional charged particles, it is possible to use the yield and the shape of the distributions in this sample to accesses the background in the signal region. Figure 4.21 shows the mass distribution for the opposite sign (signal) and same sign (backgrounds) candidates selected in data. The contamination from this background is estimated to be of the order of $10^{-4}$

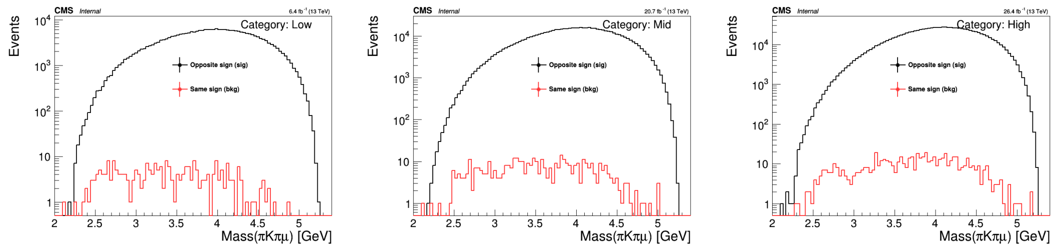


Figure 4.21: Distribution of visible mass in $D^{*-}\mu^{+}$ (correct sign) and $D^{*-}\mu^{-}$ (wrong sign) data sample.

in all the categories. Given the extremely small number of same sign candidates with the respect of the opposite sign candidates, this last source of combinatorial background is neglected as well. As a side note, it has to be noticed that a non negligible contribution from splitted tracks that was forming a peak at 2.7 GeV has been removed from the wrong sign sample by requiring $\Delta R(\mu, \pi) > 10^{-3}$.

**Muon Mis-ID**

This subsection discuss the possible background due to events where a particle is falsely identified as a muon. To study this background the data in each category are analyzed again following the same procedure explained above but inverting (vetoing) the medium muon ID requirement. In this sample then candidates are then made by a $D^*$ and a PF candidate of the appropriate charge sign, matched with the trigger object and failing the medium muon ID requirements (also not matching any muon reconstructed in the events passing the medium ID). A total of about 12k candidates are so selected among all the categories and both charge signs.

We consider two main processes that may be present in this data sample. First, events where $B \to D^*\mu + X$ decay actually happened but the muon simply fails to pass the ID. As per muon POG data[6], this false negative probability for muons in the relevant phase space ($p_T \sim 10$ GeV, $\eta \sim 0.5$) is about $p_{FN} = 1\%$. Secondly, events where $B \to D^* + \mathrm{n}h$ happened and one of the hadrons $h$ satisfies the trigger requirements. It is useful to consider the decays $B \to D^*K$ and $B \to D^*\pi$ which

[6] `https://gitlab.cern.ch/cms-muonPOG/MuonReferenceEfficiencies/-/blob/master/EfficienciesStudies/2018/Jpsi/jsonfiles/RunABCD_data_ID.json`

belong to this class of processes and have a branching fraction of $2.2 \cdot 10^{-4}$ and $2.7 \cdot 10^{-3}$.

Fig. 4.22 show the distribution of the visible mass for events with no additional tracks (signal region) when the PF particle matched with the trigger object (which makes the visible system together with the $D*$) is interpreted as a $\pi$ (top), $K$ (middle) and $\mu$ (bottom). It is clear that neither in the sample with the requirement of the

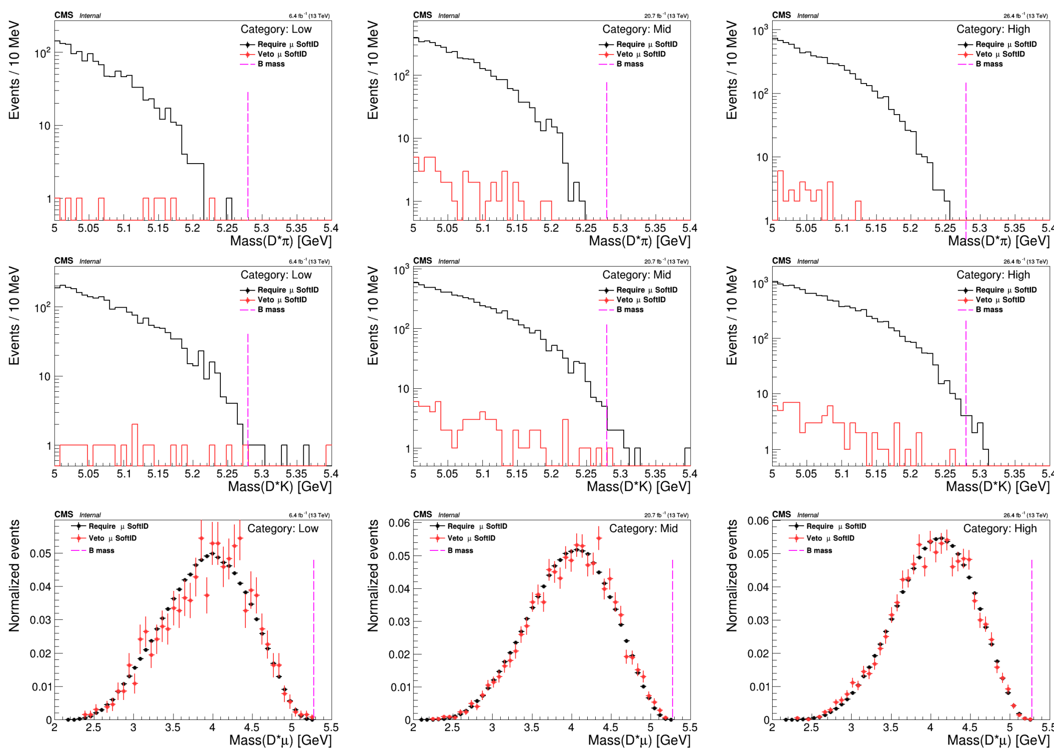


Figure 4.22: Distribution of visible mass in data when the PF particle matched with the trigger object is given the $\pi$ (top), $K$ (middle) and $\mu$ (bottom) mass hypothesis. The distribution is shown for the three categories, from left to right, and for both the sample with the required and vetoed muon medium ID. In the case of the $\mu$ mass hypothesis the distributions are normalized in order to better highlight the shape comparison.

medium ID nor in the one with the veto on the medium ID it is present a resonance on $m(D^*K)$ or $m(D^*\pi)$ around the value of the $B$ meson mass. Hence, we can then conclude that no $B \to D^*K$ or $B \to D^*\pi$ events are selected. Considering that the biggest contribution to the sample is from $B \to D^*\mu\nu$, we can set an upper bound on the false positive rate ($\epsilon_{mis-ID}$) for hadrons to be identified as muon in control region candidates. Since no events where observed, at 95% c.l. up to about $\lambda_{max} = 3$

events are compatible with data (assuming a poisson distribution). Estimating from MC that the ratio of the probability to produce a particle in the trigger phase space is

$$r_{GEN} = \frac{\epsilon_{GEN}(B \to D^* \mu \nu)}{\epsilon_{GEN}(B \to D^* K)} \approx \frac{1.5 \cdot 10^{-2}}{4.5 \cdot 10^{-3}} = 3$$

and that the ratio of the branching fractions is $r_{Br} \approx 10^{-2}$, we can set an upper limit of

$$\epsilon_{mis-ID} \leq \frac{\lambda_{max}}{N_{\text{tot}} r_{Br} r_{GEN}} \approx 5 \cdot 10^{-5}$$

where $N_{\text{tot}} \sim 2 \cdot 10^6$ is the total number of candidates observed in data. As a consequence a negligible amount of background from hadrons misidentified as muon is expected. What is more, the visible mass distribution under the muon hypothesis when the medium ID is vetoed is compatible with the one when the medium ID is required. This can leads to the interpretation that the dominant process in the sample with the veto on the medium ID are the false negative $B \to D^* \mu \nu$ events. This is also compatible with the observed yield which is indeed the same order of $p_{FN} \cdot N_{\text{tot}}$.

## 4.4 Systematic Uncertainties

### Pileup

Since the CMS data parking stream is made of different trigger paths that are switched on progressively during the fill, the PU distribution is different for each trigger (and hence category). Figure 4.23 left shows the pileup distribution at trigger level for the three different triggers considered in the analysis. For the MC simulations a custom pileup profile has been derived in order to allow a reweighing to all the categories while minimizing at the same time the statistical loss. Although the Deterministic Annealing primary vertex reconstruction has been shown to be efficient and well-behaved up to relatively high levels of pileup, the final distribution for the number of reconstructed primary vertices is still sensitive to the details of the primary vertex reconstruction and to differences in the underlying event in data vs MC. Additionally, there is the potentially larger effect that the distribution for the number of reconstructed vertices can be biased by the offline event selection criteria and even by the trigger. In order to factorize these effects, instead of reweighting the Monte Carlo by the number of reconstructed vertices, we reweight instead the number of pileup interactions from the simulation truth. The target pileup distribution for data is derived by using the per bunch-crossing-per-luminosity section instantaneous

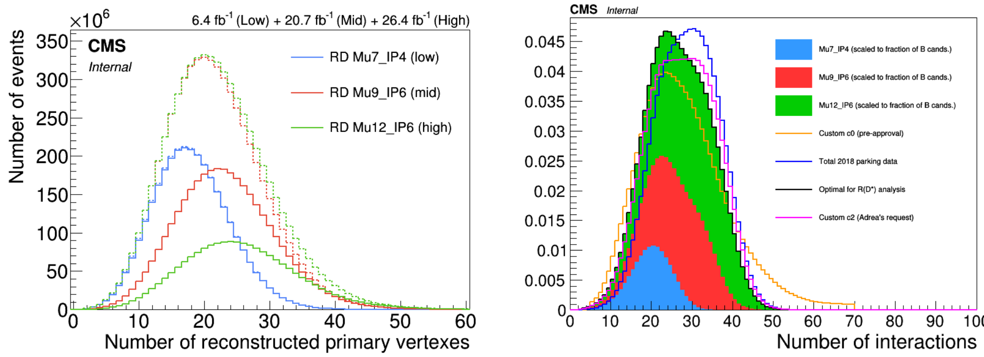


Figure 4.23: Left: Distribution of the number of vertexes in data for event where the triggers used in the analysis were active (dashed line) and fired (solid line). Right: Distribution of the number of interactions in data for the three different categories represented in a stack full histogram. The considered custom pileup spectra (c0 and c2), the statistically optimal spectrum and the total in data are also shown.

luminosity together with the total pp inelastic cross-section to generate an expected pileup distribution.

The custom *c2* pileup spectrum used in the simulation of the main samples for this analysis is defined by an ad-hoc distribution shown in Figure 4.23 right. Since a reweighing will anyways be needed to reproduce the pileup distribution in each category, it has to be noticed that although it is not necessary to achieve a closure between data and MC at this stage However, the better agreement the lower the waste of statistic due to pileup reweighing will be. Once the events are selected in the various categories they are reweighed to have a matching pileup spectrum between data and MC. Figure 4.24 shows a comparison of the pileup spectrum in data and

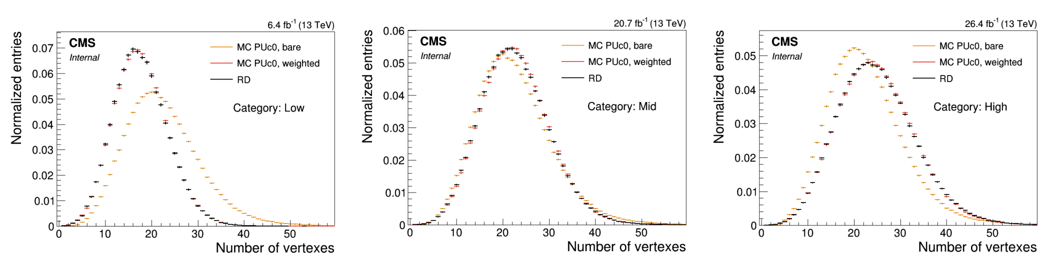


Figure 4.24: Data and MC distribution of the number of vertexes in the three categories.

MC for all the three categories, before and after the pileup weights are applied. An agreement at the level of the statistical uncertainty is achieved in all the categories between the data and MC pileup distributions in selected events. No systematic uncertainty is assigned to the pileup modeling.

### Muon ID Efficiency

The official CMS muon POG scale factors (SF) for the muon ID efficiencies [51] for the 2018 data are used to correct the MC and estimate the Muon ID uncertainty. The SF from $J/\psi$ tag-and-probe for low $p_T$ muons have been used: `https://gitlab.cern.ch/cms-muonPOG/MuonReferenceEfficiencies/blob/master/EfficienciesStudies/2018/Jpsi/rootfiles/RunABCD_SF_ID.root`. The effect of the uncertainty of muon ID

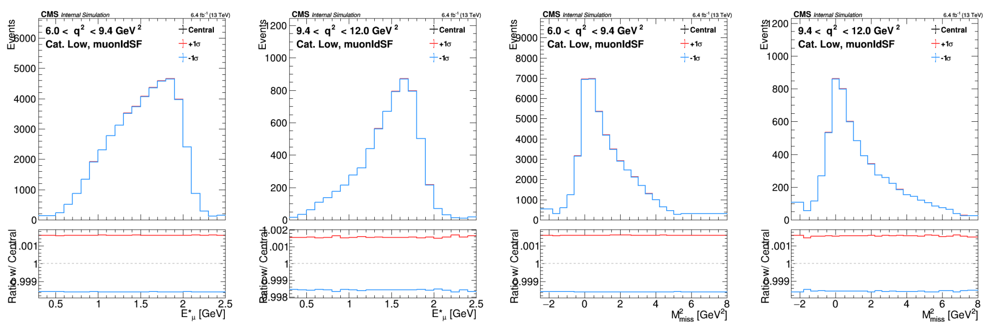


Figure 4.25: Variation on the total expected number of events due to the uncertainty on the muon ID scale factors.

scale factor is found to be completely degenerate with a overall scale uncertainty of about 0.15% (Fig. 4.25). Hence it is not propagate because it is perfectly absorbed by the luminosity uncertainty which is assigned a freely floating overall scale factor.

### Trigger Efficiency

In order to evaluate the systematic uncertainty due to the trigger modeling, the triggers efficiency have been measured in data and MC. The ratio of the data efficiency over MC efficiency, here referred as *scale factor*, is extracted with its uncertainty and than used to weight the MC samples in the fitting procedure. The parking stream trigger paths, named `HLT_Mu*_IP*`, require constraints only on a muon in order to fire. For each of the three trigger used in the analysis (see Section 4.3), the efficiency is measured in bins of the triggering muon $p_T$, $\eta$ and impact parameter significance (IP). Following the guidelines in [51], the trigger efficiency are measured given the request of the Medium muon ID ($\varepsilon_{\text{trg|ID}}$). A tag and probe method is used to measure the efficiencies as the ratio of the number of numerator muons over the number of denominator muons, defined as follows. A muon is counted in the denominator if the following conditions are met:

- The trigger under study has a non zero prescale in the event
- The muon has an associated inner track and pass the muon ID.
- It exists in the event a different muon, called *tag*, that has opposite charge of the probe muon, has triggered a parking HLT path, and has inner track details.
- The angular distance between two muons is $\Delta R > 0.35$
- The secondary vertex refit of the two muons has a p-value > 5% and the muons pair refit mass is within 0.1 GeV form the $J/\Psi$ mass.

A muon is counted in the numerator if is in the denominator, has a matching L1 muon satisfying the trigger requirements and has fired the trigger under study. This same procedure is applied to data and the $B \to J/\psi K^*$ MC sample. All the data coming from the luminosity sections used in the analysis are used to measure trigger efficiencies.

Figure 4.26 and Figure 4.27 show the measured trigger efficiency respectively in data and MC. The quoted uncertainty is the statistical one estimated with the root TEfficiecncy class using 68% Clopper-Pearson confidence intervals [52].

**Meson Decay Form Factors**

To estimate the uncertainty due to the decays form factors (FF), the matrix element reweighing library Hammer [31] has been used. Given the true MC kinematic variables, this library provides the user with the matrix element ratio between two different FF schemes. This ratio can be directly used as a weight to be applied to the selected candidates in order to emulate a change in the FF scheme. In the particular case of the signal samples, it has been decided to generate the events using the ISGW2 FF scheme because the Hammer author explicitly synchronized and cross tested the ISGW2 matrix element in Hammer with the one in EvtGen. The CLN [53] parametrization, obtained by reweighing the original events with Hammer, is used instead for fit procedure. This parametrization is based on Heavy-quark effective theory (HQET) and it includes both short-distance and $1/m_Q$ corrections to provide relations between the FF near zero recoil. It has to be noticed that more recent and detailed parametrization of semi-leptonic B decays [54, 55] exist in literature. However, the CLN approach has been chosen as the main scheme for its simplicity and for consistency with existing experimental $\mathcal{R}(D^*)$ measurements. In particular, this approach is equivalent to the one used

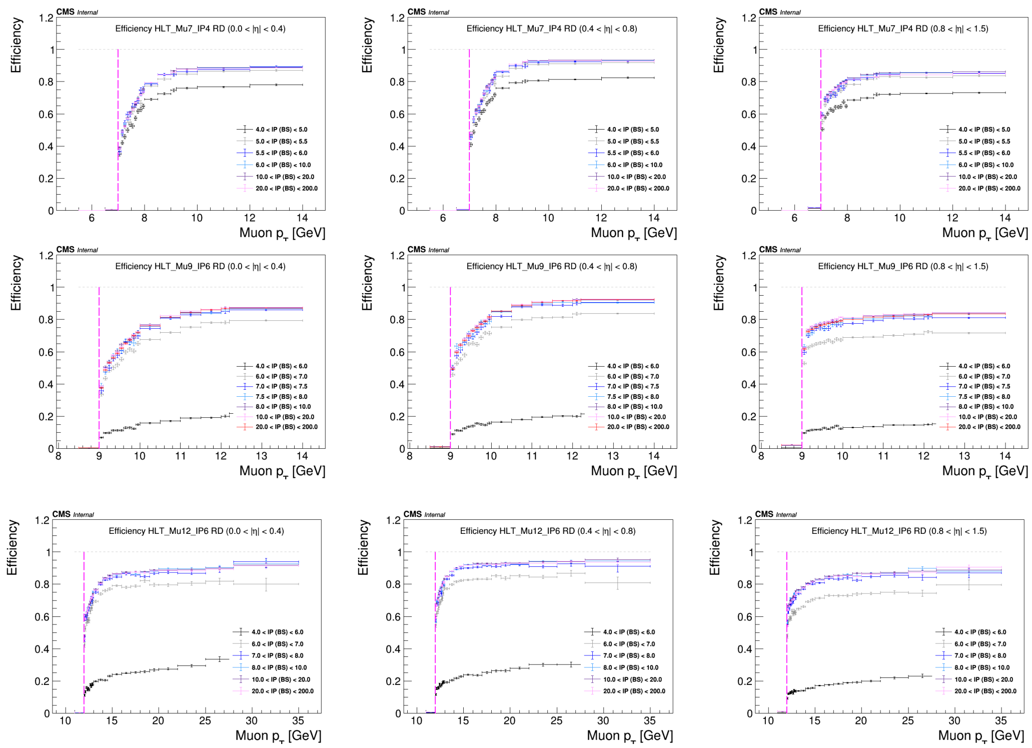


Figure 4.26: Trigger efficiencies measured in data. Different triggers are organized by row, $\eta$ bins by column, IP bins by different colors and $p_T$ bins as the X axis of the graph.

by LHCb in [15]. In the HAMMER implementation, the matrix elements computed with CLN parametrization are varied propagating the uncertainty on 4 parameters: $\rho^2$, $R_1$, $R_2$, and $R_0$. Experimental measurements of the first three parameters are available in literature. Central values and uncertainties are taken from [56]: $\rho^2 = 1.122 \pm 0.024$, $R_1 = 1.270 \pm 0.026$ and $R_2 = 0.852 \pm 0.018$. The parameter $R_0$, which quantify the helicity suppressed FF, is of crucial importance for $B \to D^* \tau \nu$ decays. Due to the absence of experimental measurements, the theoretical estimation from [33] is used with the provided uncertainty of 10%: $R_0 = 1.14 \pm 0.11$. While the uncertainty on $R_0$ is clearly uncorrelated by the other ones, the uncertainty of the three parameters experimentally measured have non-negligible correlations: $\sigma_{\rho^2,R_1} = 0.566$, $\sigma_{\rho^2,R_2} = -0.824$ and $\sigma_{R_1,R_2} = -0.715$. In order to take into account linear correlations, independent variations are computed by diagonalizing the covariance matrix (`https://github.com/ocerri/BPH_RD_Analysis/blob/master/scripts/FormFactorsDiag_BtoDstLNu.ipynb`). Uncertainty of FF parameters will be considered in the fit by considering $1\sigma$ variations of the covariance matrix eigenvectors.

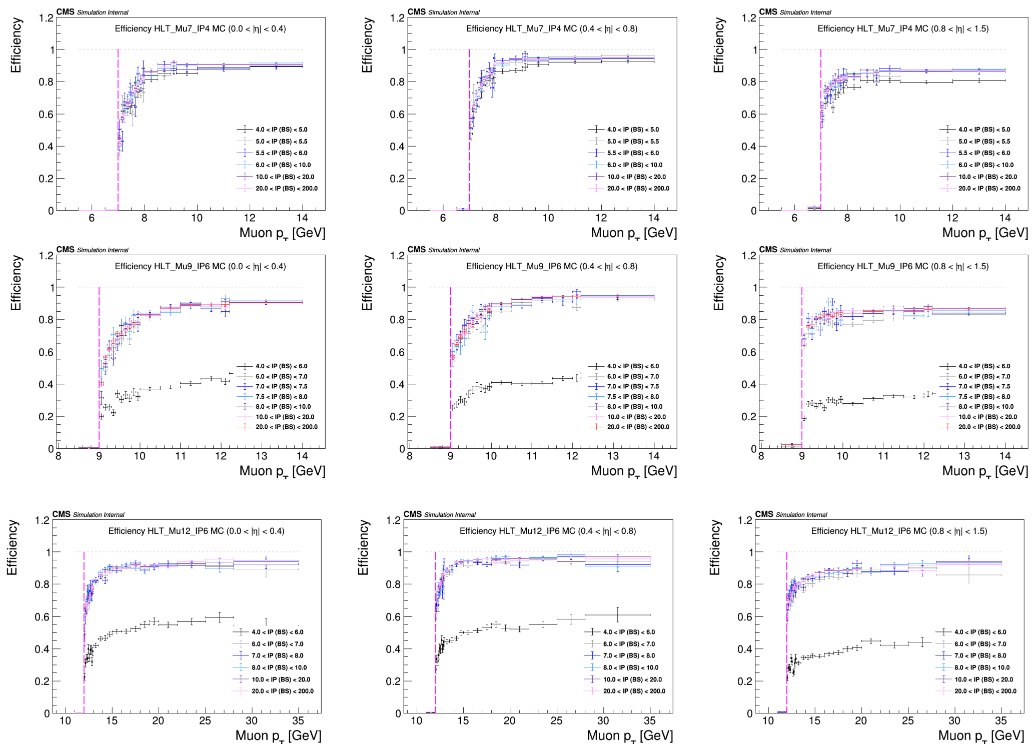


Figure 4.27: Trigger efficiencies measured in MC. Different triggers are organized by row, $\eta$ bins by column, IP bins by different colors and $p_T$ bins as the X axis of the graph.

Thanks to the direct collaboration with the HAMMER authors, it has been possible to install and integrate the usage of the library in CMSSW 10.2.3. The procedure followed is documented at the following link: `https://github.com/ocerri/BPH_RDntuplizer/blob/master/build_Hammer_notes.txt`. During the MINIAOD processing of the signal samples, the weight to change the FF scheme from ISGW2 to CLN is computed along with the 8 corresponding variations. The weights are then applied to the selected candidates when filling the histograms used in the fit. As a reference, the distributions of weights in $B \to D^{*}\mu\nu$ and of the true MC $q^2$ are shown in Figure 4.28. There, the central values corresponding to the reweighing from ISGW2 to the CLN is compared to the distributions obtained when weight for a CLN FF scheme with one of the parameter varied by $1\sigma$. From the pictured it is possible to notice how: the weights do not take extreme values and average at about 0.8; the kinematic distributions are minimally deformed by the FF scheme parameter shift.

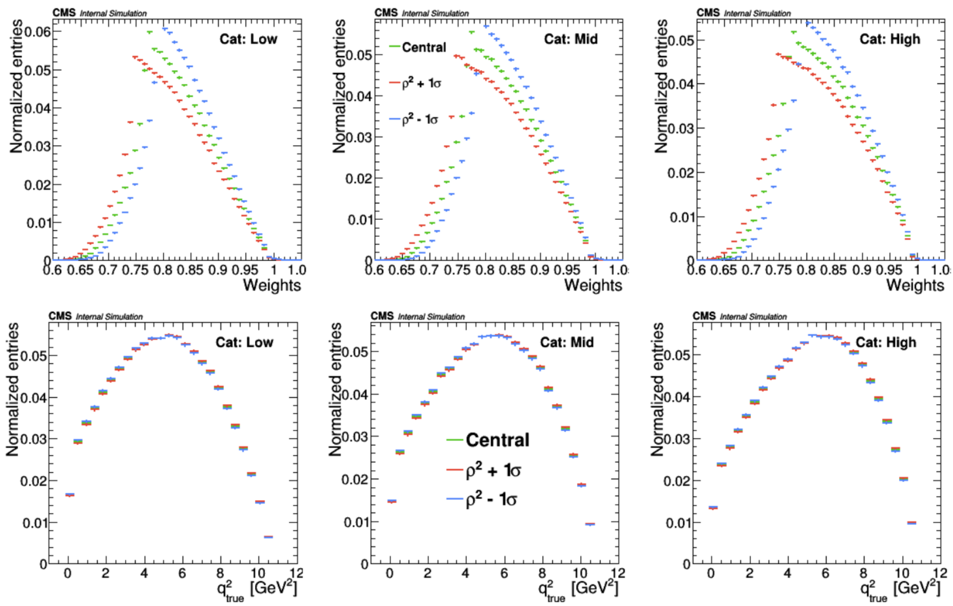


Figure 4.28: Top: Distributions of weights in $B \to D^*\mu\nu$ MC sample. For each category, central weights as well as weights when $\rho^2$ is varied by $1\sigma$ are plotted in different colors.
Bottom: Distribution of the true MC $q^2$ in $B \to D^*\mu\nu$ MC sample.

Since the difference of kinematic distributions between FF schemes is very small, two dedicated MC samples without pileup have been produced to test the reweighing procedure: one with the ISGW2 scheme and one with a HQET2 scheme.

The appropriate Hammer weights to reweigh the first one into the second have been computed and applied as a closure test. Figure 4.29 shows the result in the most sensitive variable $q^2$. Distributions of other variables ($M_{\text{miss}}$, $E^*_\mu$, etc.) are omitted since no difference between ISGW2 and HQET2 was observable. Given that efficiency are computed before applying these weights, in order to keep the proper normalization, Hammer weights are applied after being multiplied by the ratio between the total width of the original FF scheme used in production and the total width of the FF scheme targeted by the particular weight.

This re-weighting approach is also extended to the form factors of B decays involving $D^{**}$ mesons using the matrix elements still available in Hammer. In particular, for decays involving the narrow $D^{3/2+}$ resonances ($D_1$ and $D_2^*$) and for wide $D^{1/2+}$ resonances ($D_0^*$ and $D_1^*$) the BLR scheme [57] is used. A special treatment is reserved for $D(2S)$ and $D(2S)^*$ because no scheme exists in Hammer so the authors provided us with a temporary parametrization still under study.

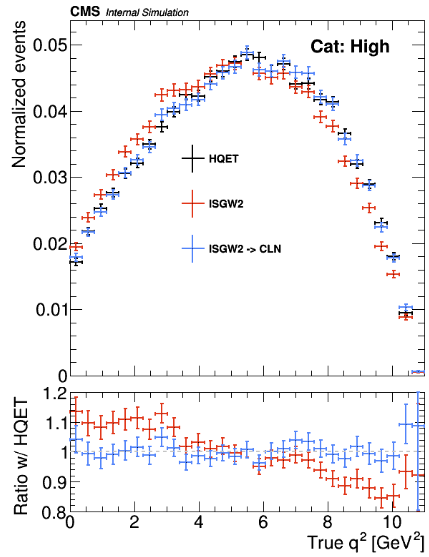


Figure 4.29: Distribution of the true MC $q^2$ for the MC produced with HQET (black), bare ISGW2 (red) and ISGW2 reweighed with the appropriate HAMMER weights to HQET (blue). The blue ratio is compatible with one, underlying the success of the closure test.

## 4.5 Signal Determination

As discussed in 4.3, events are divide into three category based on the trigger path considered. Within each category, a signal region and 5 control regions are defined based on the number and charges of the additional tracks. The signal region is further divided into 4 based on the measured value of $q^2$. In total the fit is simultaneously performed on 27 distributions, $5 \times 3 = 15$ representing the control regions and $4 \times 3 = 12$ representing the signal regions. The Combine [58] library is used to perform such a fit.

The signal region is defined by vetoing the presence of additional tracks. In this region, the two dimensional distribution of $M^2_{\text{miss}}$ vs $E^*_\mu$ is fit in four different bins of $q^2$ chosen such that: the $B \to D^*\tau\nu$ contribution (as of MC) is non-negligible only in the two highest bins; and the last bin maximize the ratio between $B \to D^*\tau\nu$ and $B \to D^*\mu\nu$ events. As a technical detail, the 2D distributions are unrolled into a 1D histogram in order to be fed to Combine. Control regions are defined by requiring the presence of N=1 and N=2 tracks, splitting the possible different charges combination Q: two control region with one track, Q=+1 and Q=-1; and three control region with two tracks given the possible sum of charges combinations Q=+2,0,-2. Practically, in order to threat both charge conjugate candidates in data consistently, the additional tracks charge sum Q is multiplied by the charge sign of the muon. In all control regions with non zero additional tracks charge sum, the distribution of the invariant mass of all the reconstructed hadrons ($M_{\text{had}}$) is fit. This mass includes the three mesons produced in the $D^*$ decay ($\pi\pi K$) and the additional track(s). In the $\{N = 2, Q = 0\}$ control region, the visible mass $M_{\text{vis}}$, computed including all the particles considered for $M_{\text{had}}$ plus the $\mu$, is fit. This choice of observables had been done in order to maximise the sensibility to the normalization of the different backgrounds. Given the additional difficulties in modeling the $\{N = 2, Q = 0\}$ region, a 2D distribution of $M_m^2 iss$ vs $q^2$ is used in this control region. It is possible to see a visual representation of the different fit categories and regions in Fig. 4.30. While the signal region and its observables are designed to maximize the sensitivity

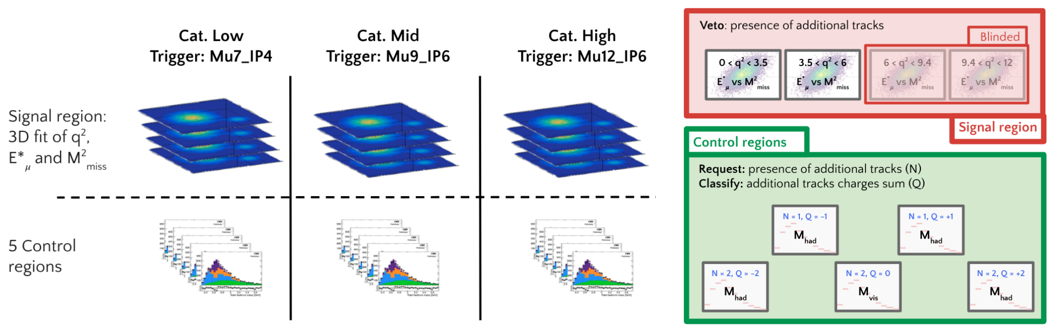


Figure 4.30: Cartoons representing the fit structure. The 3 different categories are pictured on the left and the regions structure within each category is pictured on the right.

to $\mathcal{R}(D^*)$, control regions are though to constrain backgrounds' normalization and to validate the simulation. The fit is run simultaneously in signal and control regions using the RooFit-based utility *Combine* [59] under CMMSW 10_2_13. Since Combine only take as inputs one dimensional histograms, the two dimensional distributions of $M^2_{\text{miss}}$ versus $E^*_\mu$ are practically unrolled into a single dimensional histogram. In order to retain a sufficiently fine grained binning without ending up up with a sparse histogram, during the unrolling procedure bins with less than 3 expected events are merged with a neighboring one.

The fit procedure consist in the optimization of MC histograms template based pdf to binned data. While data histograms are simply created by filling the evens in the appropriate bin, each MC histogram is scaled to reflect a normalization equal to the number of expected events $n_{exp}$. For each MC histogram, $n_{exp}$ is estimated with the following expression

$$n_{exp} = \mathcal{L} \cdot \sigma_{pp\to b\bar{b}} \cdot \varepsilon_{\text{gen}} \cdot \mathcal{B}_f \cdot \varepsilon_{\text{ntp}} \cdot \varepsilon_{\text{sel}} \cdot \varepsilon_{\text{histo}},$$

where $\mathcal{L}$ is the total luminosity measured (for the given category) in data, $\sigma_{pp\to b\bar{b}}$ is the cross section returned by Pythia (same for all the processes and categories), $\mathcal{B}_f$ is the decay chain forced branching fraction, $\varepsilon_{\text{ntp}}$ and $\varepsilon_{\text{sel}}$ are the efficiency of the selection process, and $\varepsilon_{\text{histo}} = N_{\text{histo}}/N_{\text{sel}}$ is the ratio between the total number of MC event selected for the given category and the number of events which filled the histogram.

As explained in Sec. 4.2, a total of 21 MC process are considered: 2 signal process, $B \to D^*\tau\nu$ and $B \to D^*\mu\nu$; 10 processes of the type $B \to D^{**}\mu\nu$; and 9 process of the double charmed decay type $B \to D^*H_c(\mu X)$. The script used to setup and run the fit in Combine can be found at `https://github.com/ocerri/BPH_RD_Analysis/blob/master/Combine/runCombine.py`.

**Systematic Uncertainty Nuisance Parameters**

In Combine, only the $B \to D^*\tau\nu$ is considered a *signal* process. In this way, a rate parameter $r$ is automatically created to float the total number of events coming from this process. As a result of the normalization choices made and the $\mathcal{R}(D^*) = 1$ assumption made in Sec. 4.2 when computing the branching fraction for $B \to D^*\tau\nu$, we have that by definition $r \equiv \mathcal{R}(D^*)$. Since in this analysis the uncertainty on the branching fraction of $\tau \to \mu\nu\nu$ is completely degenerate with $r$, it has not been considered among the systematics, but it is negligible given its value of $(17.39 \pm 0.04)\%$. It would indeed appear as a multiplicative factor with a relative uncertainty of 0.2%, well beyond the reach of this study. A set for nuisance parameters, for a total of over 1300, are used in the fit to model systematics and propagate uncertainties.

**Scale nuisance** This section describes the set of nuisance parameters used to model effect that appears as multiplicative factors to processes normalization. The following list of parameters, all constrained with a log normal distribution, is used in each category:

- `overallNormMu7_IP4`, `overallNormMu9_IP6` and `overallNormMu12_IP6` with an uncertainty of 100% are applied to all processes in all regions respectively in the low, mid, and high category to the model overall normalization with practically a flat distribution. The associated uncertainty can mainly be associated to the cross section estimated by Pythia, since all the MC samples are produced with the same Pythia process.

- `trkEff`, a scale parameter with gaussian constrain $1 \pm 0.021$ meant to model the tracking efficiency uncertainty which does not enter in the signal region, enters linearly in $N = 1$ control regions and quadratically in $N = 2$ control regions.

- A total of 9 scale nuisance are used to model branching fraction uncertainty which appear as overall normalization in front of whole MC samples. Example of these parameters are the inclusive muon branching fraction of charmed mesons like $D^0 \to \mu + X$ and the ratio $\mathcal{R}(D^{**})$.

**Shape nuisance** This set of nuisance parameters are used to model uncertainties through the modification of the shape (and normalization) of the provided MC distributions. As explained in [59], for each shape nuisance, two additional histogram per affected process have to be provided in all the region where the nuisance is considered. These two histograms represent the distribution of the process when the underlying uncertainty is shifted by $1\sigma$. The estimation for intermediate values is obtained with an interpolation of the bin content of the three provided histograms. The following shape nuisance are considered:

- A total of 22 `trgMuX_IPYSF_ptZ`, are used to model the uncertainty of trigger SF derived in Sec. 4.4 for each trigger in each bin of $p_T$ separately. Each of them is only affecting the category where the trigger is used.

- A total of 7 `softTrkEff`, associated with soft track efficiency uncertainty as a function of $p_T$. Practically, 7 bins are defined in the range 0.5 GeV to 2 GeV with associated uncertainties decreasing from 10% to 1%. For each bin, the weight of tracks belonging to the $p_T$ range is varied by the bin uncertainty and all the observable distributions recomputed.

- 16 shape nuisance to describe the FF scheme eigenvector variations derived from Hammer. In particular, 4 are used to describe the variations of the CLN scheme and the remaining ones for the excited $D^{**}$ processes.

- `fDststWide` models the uncertainty on the fraction of wide resonance in $B \to D^{**}\mu\nu$. It is applied to only the process involving $D^{**0} \to D^{*-}\pi^+$ because it is the only one where such effect is observable. An uncertainty of 20% is associated to the provided histograms, in agreement with $\Gamma_{14}$ for $B^+$ in [1].

- `D2420_10Width` and `D2460_1StWidth` model the uncertainty on the width of the narrow $D^{**}$ resonances decaying into one $\pi$. A reasonable uncertainty of 15% is estimated.

- `DstPiPiWidth` model the uncertainty on the width of the $D^{**}$ decaying into $\pi\pi$. A reasonable uncertainty of 15% is estimated.

- A total of 16 nuisance are introduced to further model $B^0 \to D^*\ell\nu(n)\pi$ mix compositions by propagating the uncertainty on all the sub-processes discussed in Tab. 4.2.

- A total of 35 nuisance are introduced to further model $B^0 \to D^*H_c(\mu X)$ mix compositions by propagating the uncertainty on all the sub-processes discussed in Tab. 4.6.

A set of plot showing how the shape uncertainties modify the expected distribution of total number of events can be found in the internal note in [60].

**Monte Carlo statistics** The systematic uncertainty due to the limited MC statistic is modeled with the `autoMCStats` Combine option. The threshold above which the Barlow-Beeston [61] approach is used is set to 5 events.

**Remarks About Using a 3D Likelihood Instead of an MVA**

We studied the possibility of improving the analysis sensitivity by introducing additional observables in the control region and/or processing the chosen set with a multi-variate analysis algorithm. It is worth remembering here that, in the limit of a *fine enough* binning, a full multi dimensional likelihood fit to a set of observables is expected to outperform in sensitivity a likelihood fit to every test statistics obtained from the same set (e.g. the output of a classifier). While it is possible to achieve a reasonable multi-dimensional binning with up to 3 variables (like for the case of the analysis's baseline with $q^2$, $E^*_\mu$ and $M^2_{\text{miss}}$), this is practically impossible for a higher number of observables. For the purpose of this study we developed two different MVA based on `GradientBoostingClassifier` from scikit-learn with two different set of input variables. The first one (v0) is trained with the same 3 three observables $q^2$, $E^*_\mu$ and $M^2_{\text{miss}}$ as the nominal multidimensional fit. The second one (v1) is trained with 6 input observables, adding $\eta_B$, $p_T^{\text{vis}}$ and the impact parameter of the muon w.r.t. the $D^*$ vertex ($IP_{D^*,\mu}$) to the three ones from v0. Both BDTs are trained with a one-versus-all classification loss with the $B \to D^*\tau\nu$ MC sample as a one class and all the other MC samples as the other one. What is more, during the training process the events are reweighed such that the three samples made of $B \to D^*\tau\nu$, $B \to D^*\mu\nu$ and the expected mix of all the backgrounds have the same impact on the loss. BTDs have been chosen for this study not only for their vast success in literature and training stability but also because of the capability of defining a variable ranking which could be used to gain insights about input variables relevance. Tab. 4.8 reports the relative feature importance for the input observables used in v0 and v1. It can be observed that for both models the $M^2_{\mathrm{miss}}$

Table 4.8: Input variables relative importance (out of 100) for the two BDTs. It is clear how the $M^2_{\mathrm{miss}}$ is accountable for most of the discrimination power.

| | $q^2$ | $E^*_\mu$ | $M^2_{\mathrm{miss}}$ | $p_T^{\mathrm{vis}}$ | $IP_{D^*,\mu}$ | $\eta_B$ |
|---|---|---|---|---|---|---|
| v0 | 6 | 6 | 88 | - | - | - |
| v1 | 3.4 | 0.4 | 84.7 | 0.6 | 10.9 | < 0.01 |

accounts for more than 80% of the discrimination power. In order to evaluate the fit performance when using different variables Asimov fits to the low category has been run with the different observables. The total expected uncertainty for these fits are reported in Tab. 4.9 from which it is clear that the three dimensional fit to $q^2$, $E^*_\mu$ and $M^2_{\mathrm{miss}}$ used in the analysis has the lowest expected uncertainty.

Table 4.9: Total expected uncertainty on $\mathcal{R}(D^*)$ for an Asimov fit when the different variables distributions are used. The relative uncertainty difference $\Delta\sigma/\sigma$ from the nominal fit is also reported. A value of 0.295 for $\mathcal{R}(D^*)$ has used.

| Observables distribution | Uncertainty $[10^{-3}]$ | $\Delta\sigma/\sigma$ |
|---|---|---|
| $\{q^2, E^*_\mu, M^2_{\mathrm{miss}}\}$ | 42 | 0 |
| $\{q^2, E^*_\mu\}$ | 52 | +24% |
| $\{q^2, M^2_{\mathrm{miss}}\}$ | 44 | +6% |
| {BDT_v0 score} | 47 | +13% |
| {BDT_v1 score} | 48 | +16% |

**Blinded Fit to Real Data**

Following CMS policies, this analysis has been developed and still remains in a blinded regime, meaning that the most sensitive regions of the fitted phase space are not included in the fit. For this $\mathcal{R}(D^*)$ analysis specifically, this means that the two highest bin in $q^2$ for all the categories are never considered and, in particular, are not included in the fit to real data. Nevertheless, a blinded fit to real data is performed to access weather the model obtained by the MC simulations with the considered nuisance is compatible with the data without biasing the analysis result. The postfit distribution for the control regions of the blinded fit are presented in Fig. 4.31. The figure highlights a general agreement within the uncertainty. The

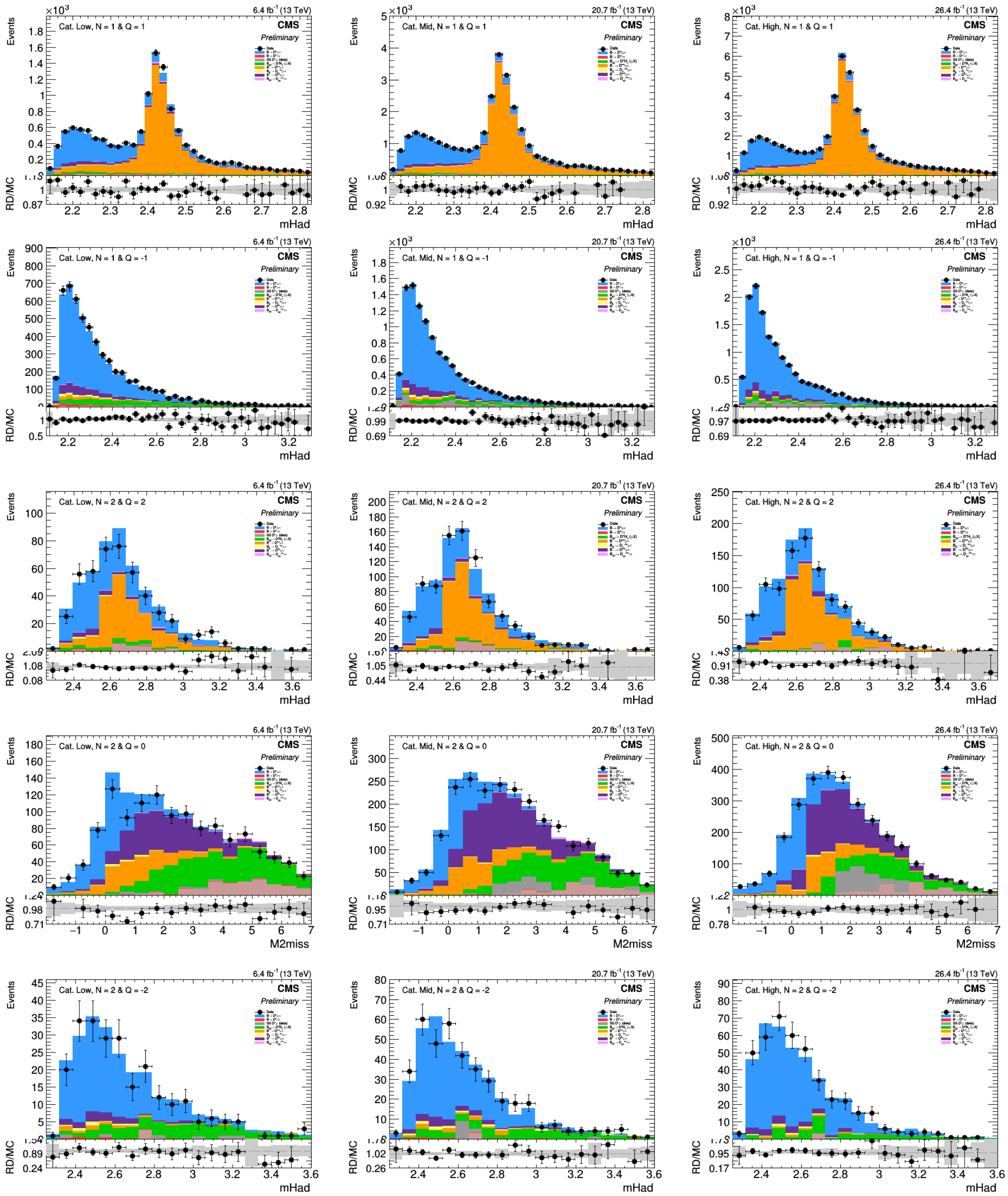


Figure 4.31: Control regions post-fit distributions for the combined fit to real data in all the three categories. Under each distributions the pulls between data and the total MC prediction expressed in data statistical uncertainty are reported. The total MC uncertainty is also reported with the pulls as a shaded area.

post fit distributions for the the signal region are reported in Figures 4.32, 4.33, and 4.34.

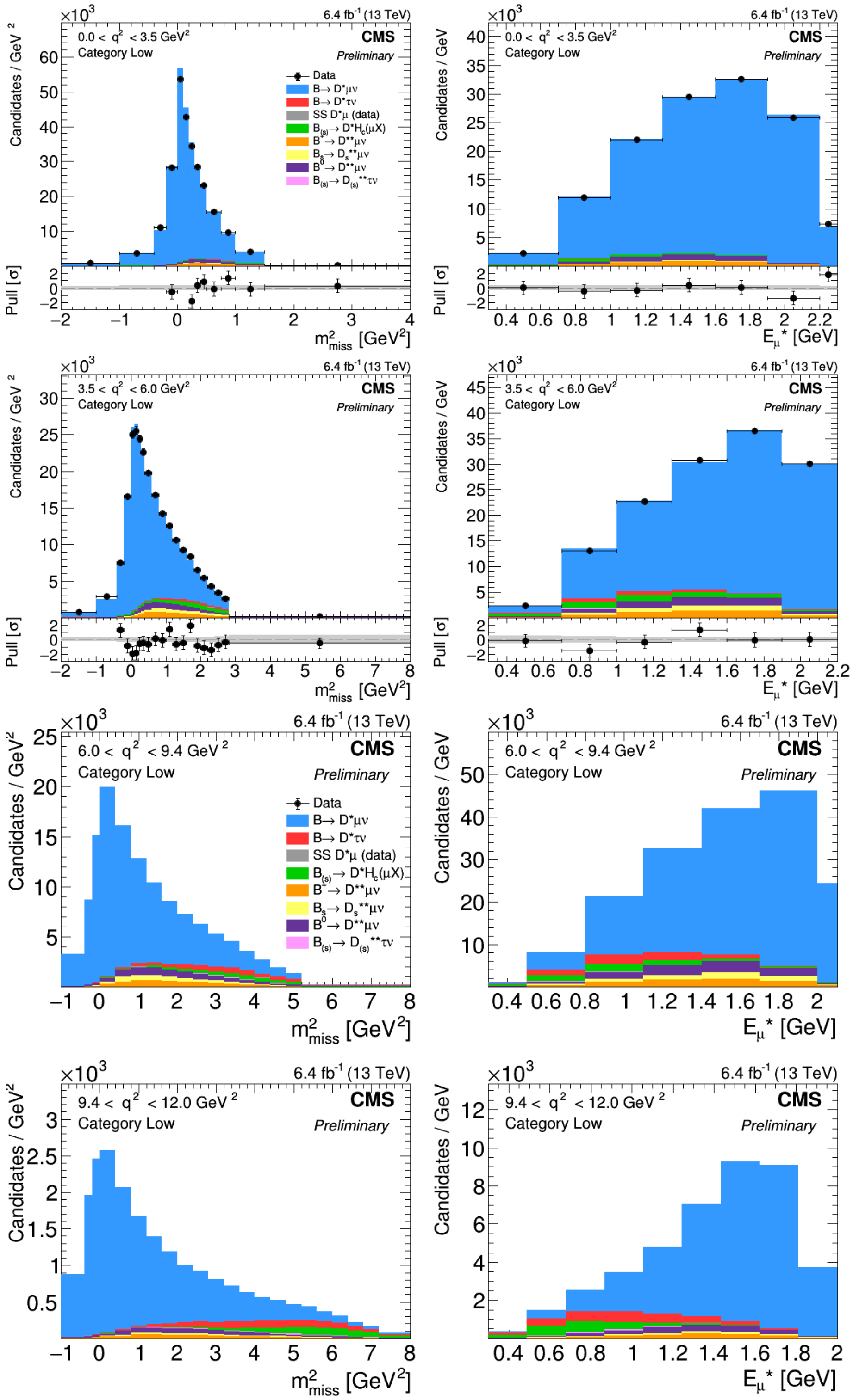


Figure 4.32: Post fit distributions for the two 1D projections of the 2D histogram of the signal region in the low category. The data histogram in the last two bins of $q^2$ is not reported as those bins are blinded.

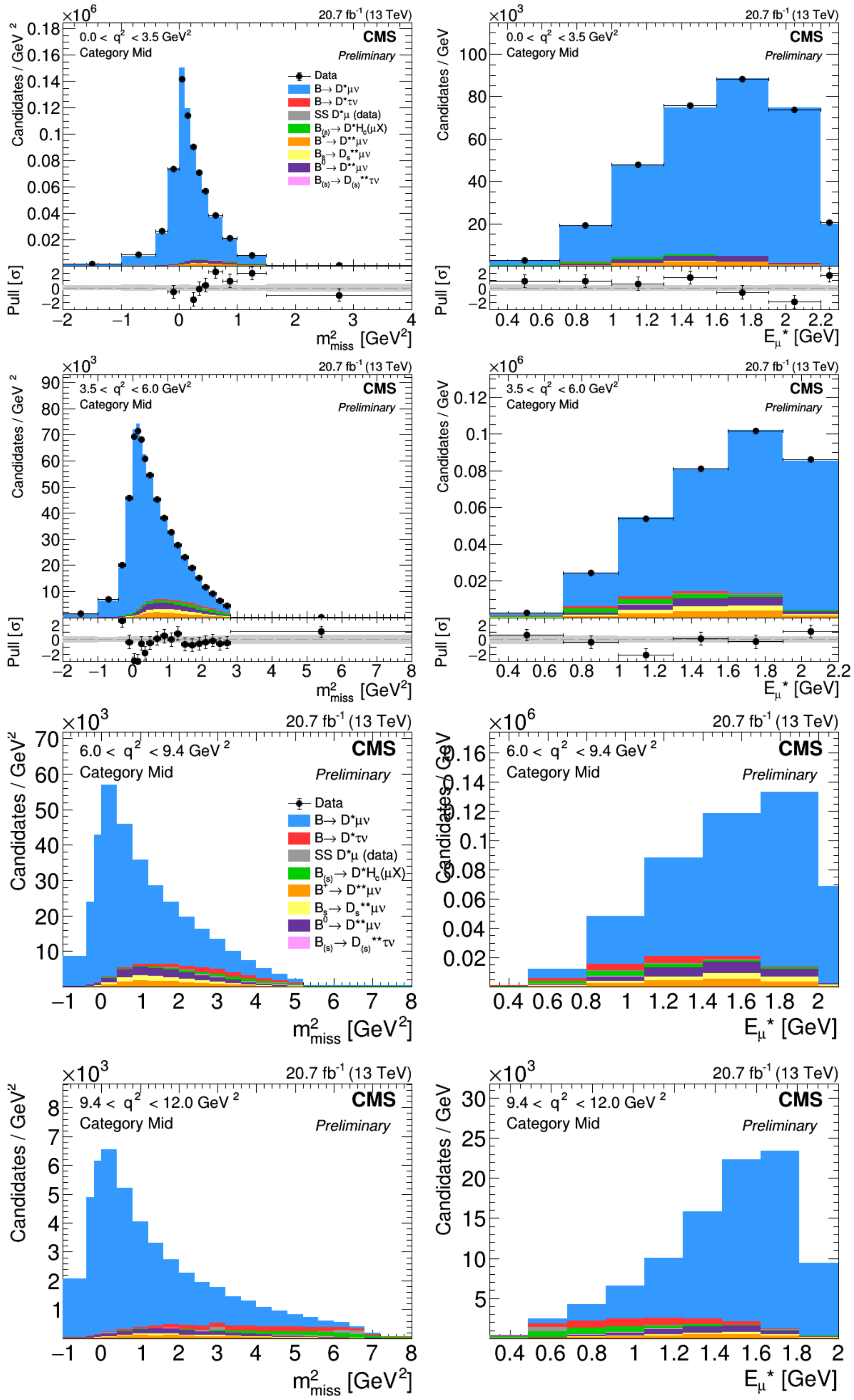


Figure 4.33: Post fit distributions for the two 1D projections of the 2D histogram of the signal region in the mid category. The data histogram in the last two bins of $q^2$ is not reported as those bins are blinded.

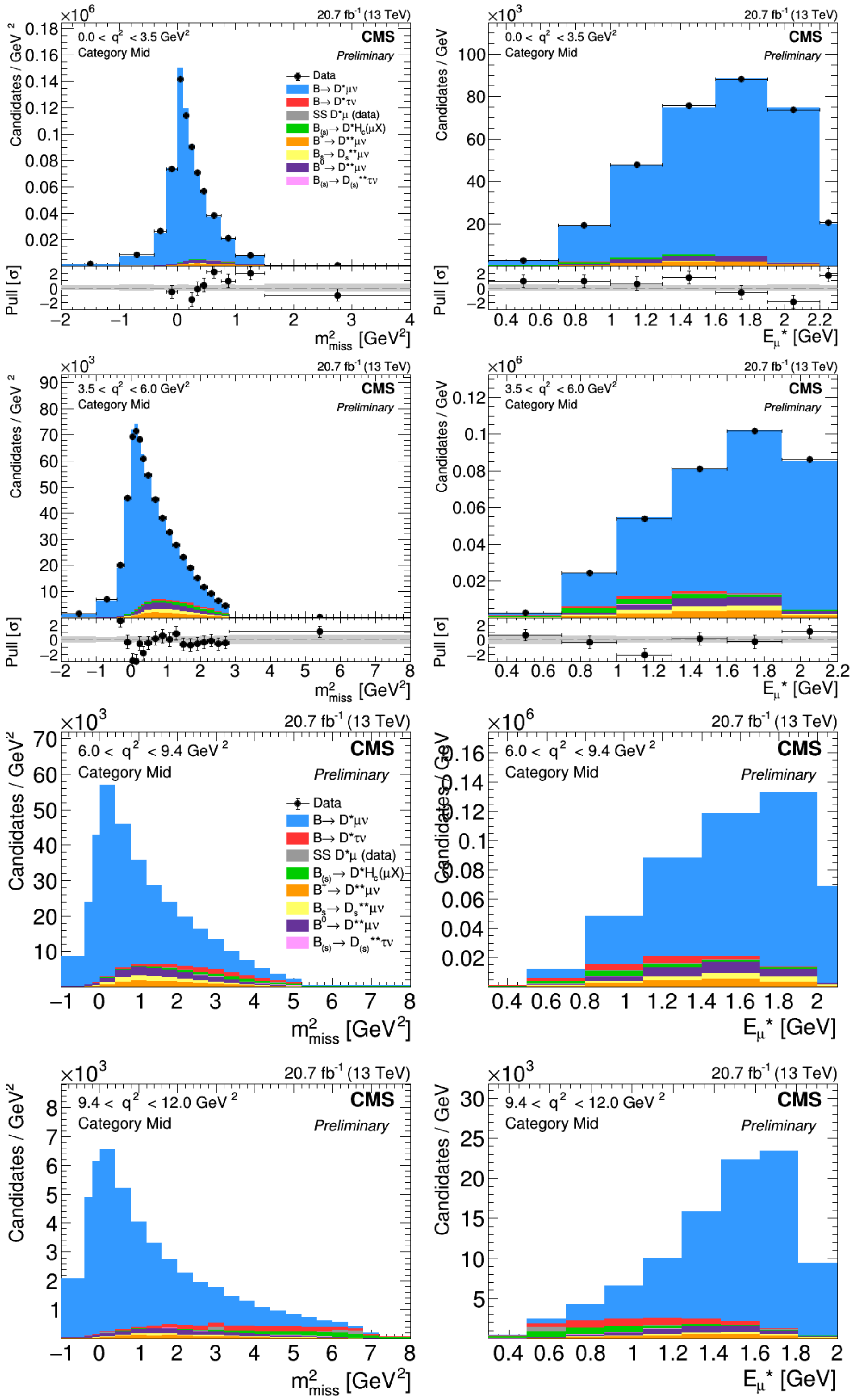

Figure 4.34: Post fit distributions for the two 1D projections of the 2D histogram of the signal region in the high category. The data histogram in the last two bins of $q^2$ is not reported as those bins are blinded.

The histograms there reported do not show the unrolled 2D histograms actually fed into Combine because those would be hard to understand. However, they report the re-rolled 1D projections which represent a more human readable format. For each category, the $M^2_{\text{miss}}$ and $E^*_\mu$ projection in the four bins of $q^2$ of the signal regions are shown. For the two highest bins of $q^2$ that are still blinded no data are shown and their content is ignored in the fit. Generally data and MC post fit distributions show an agreement within $\pm 3\sigma$ as highlighted by the pulls in the bottom of each distribution. However, major tensions are present in the lower part of the $M^2_{miss}$ spectrum which result in several sigma pulls, well beyond the scale of the plot. The likelihood scan obtained in the fit is shown in Fig. 4.35. As expected, the blinded

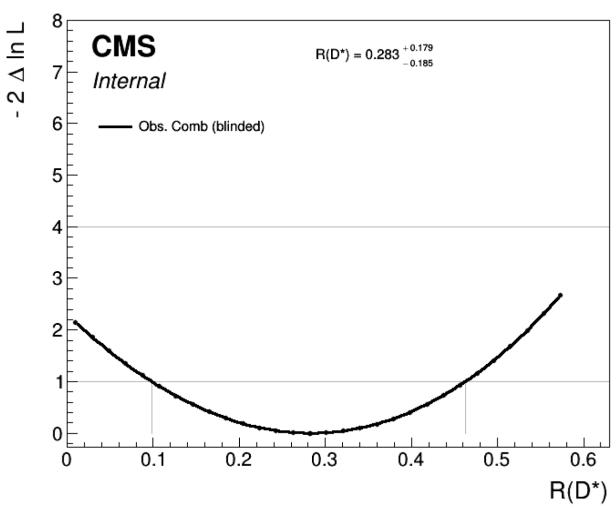


Figure 4.35: Likelihood scan for the $\mathcal{R}(D^*)$ parameter in th blinded fit to the real data.

fit results presents almost no sensitivity to the $\mathcal{R}(D^*)$ observable and is perfectly compatible with both the SM prediction and the current experimental average.

The post-fit nuisance parameters difference for this blinded fit is reported in Tab. 4.10. While most of the free parameters of the fit retain a postfit pull below the $1.2\sigma$ value, some of them report values beyond 2 or even 3 $\sigma$. In particular it is clear how the CLN FF nuisance are pulled well beyond the 95% C.L. Considering that the major impact of those nuisance is present in the signal region dominated by the $B \to D^*\mu\nu$ sample, we can associated this effect with the mismodeling at low $M^2_{miss}$ observed in the postfit distributions. Fig. 4.36 left reports the observed distribution of the post fit nuisance pulls. While the core of the distribution agrees with a Gaussian distribution, the presence of the outlayers already observed in Tab. 4.10 highlights

Table 4.10: Post fit pulls of the nuisance parameters measured in units of the prefit sigma. Only paramters with a pull above 1.2$\sigma$ are reported. Parameters with whose name ends in parenthesis, e.g. low ctrl mm mHad (1), represent the nuisance associated to the MC statistical uncertainty. B2DstCLNeig2 and B2DstCLNeig1 represent the nuisance associated with the CLN FF uncertainty. Finally, parameters starting with brB, e.g. brB DstPiMuNu 1, represent the nuisance associated with the composition of the background processes.

| Parameter | Postfit val. [$\sigma$] | Parameter | Postfit val. [$\sigma$] |
|---|---|---|---|
| B2DstCLNeig2 | $+3.33 \pm 0.81$ | low ctrl mm mHad (1) | $-2.17 \pm 0.83$ |
| brB DstPiMuNu 1 | $+2.06 \pm 0.27$ | mid ctrl mm mHad (6) | $-1.98 \pm 0.56$ |
| brB DstPiPiMuNu 4 | $+1.93 \pm 0.52$ | low ctrl pm M2miss (8) | $-1.89 \pm 0.70$ |
| brB DstPiPiMuNu 0 | $-1.78 \pm 0.74$ | high ctrl pm M2miss (7) | $-1.73 \pm 0.68$ |
| high ctrl pm M2miss (3) | $-1.56 \pm 0.58$ | B2DstCLNeig1 | $-1.50 \pm 0.86$ |
| high ctrl pm M2miss (11) | $+1.45 \pm 0.86$ | mid Unrolled q2bin1 (55) | $+1.45 \pm 0.90$ |
| high ctrl p mHad (12) | $-1.44 \pm 0.91$ | high ctrl mm mHad (6) | $-1.35 \pm 0.65$ |
| mid ctrl pm M2miss (20) | $+1.32 \pm 0.78$ | high ctrl pm M2miss (12) | $+1.31 \pm 0.83$ |
| low ctrl pm M2miss (5) | $-1.31 \pm 0.74$ | mid ctrl m mHad (9) | $+1.29 \pm 0.84$ |
| low ctrl mm mHad (11) | $-1.28 \pm 1.02$ | low ctrl m mHad (18) | $-1.28 \pm 0.90$ |
| high ctrl m mHad (4) | $-1.26 \pm 0.86$ | high Unrolled q2bin1 (36) | $-1.23 \pm 0.89$ |
| high Unrolled q2bin1 (9) | $-1.21 \pm 0.90$ | mid ctrl m mHad (29) | $-1.21 \pm 0.68$ |

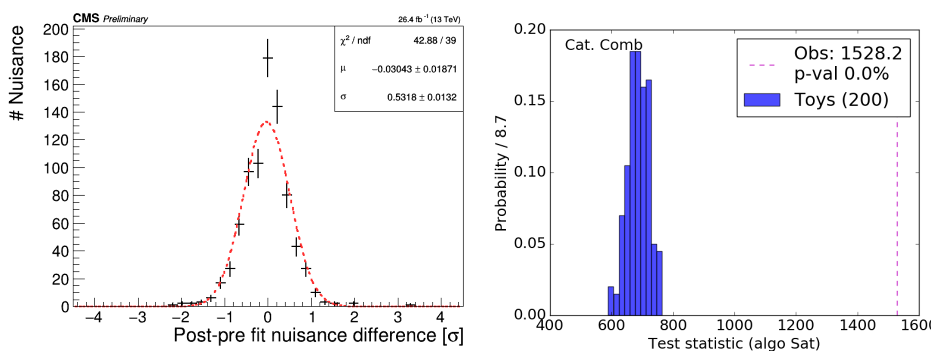


Figure 4.36: Left: Post fit distributions of the pulls of the nuisance parameters for the blinded fit.
Right: Distribution of the saturated test statistic for the toy MC (solid bled histogram) and the observed value (dashed pink).

some tensions between the data and the MC model. In order to quantify the goodness of the fit, a Saturated test is performed [62]. This test is a generalization of the $\chi^2$ test, but can be computed for an arbitrary combination of binned histograms with arbitrary constraints on the nuisance parameters. As expected from the postfit distributions and the pull values, the current fit fails the test since the observed test statistic $s_{\mathrm{obs}}$ has a value well beyond the distribution of the toy MC which has been generated to estimate the null hypothesis distribution.

Several additional fits have then been run in an effort to characterize the tensions observed. The first set of tests consist in running a fit without the signal region or, less restrictively, excluding some parts of the low $M^2_{\text{miss}}$ spectrum. As expected these fits do not presents the pulls of the CLN FF and result in a significantly reduced value of $s_{\text{obs}}$. This results underlines how part of the observed tensions are related to the need of improving the agreement between data and MC in a region almost exclusively dominated by the $B \to D^*\mu\nu$ process. The possible causes can be factorized into two groups, namely the simulation of the $B$ production and decay or the simulation of the detection and reconstruction of the particles. The simulation of the $B$ production has been explored by studying $B \to J/\psi K^*$ decays. This decay channel is reconstructed in the fully visible final state $\mu\mu K\pi$, allowing for a direct comparison of the $B$ meson kinematic without the need of approximations. No relevant difference between data and MC in this control sample is observed. The simulation of the $B$ decay has been tested by repeating the fit with different FF schemes, namely the BLPR [6] and the BGL[63], but again no relevant difference is observed. Further studies to explore the data-MC agreement in the detection and reconstruction of the particles are still undergoing. However, preliminary results point towards the need for an improved description of the soft tracks and vertex reconstruction. A second set of test fits is run masking part of the $N = 2, Q = 0$ control region. While these test show a minor general reduction of $s_{\text{obs}}$, no clear discrepancy between the background process composition is highlighted even when considering additional kinematic variables like the mass of different subsystems (e.g. the two additional tracks or the $D^*$ and one additional track) or various internal angles. Even if a minor tensions results due to the MC modeling of the $N = 2, Q = 0$ control region, the impact of the signal region is found to be far superior and hence should be solved first in order to avoid possible biases.

**Unblinded Fit to Asimov Pseudo-Data**

This section discusses the results of an unblinded fit to an Asimov dataset, i.e. where real data are replaced by pseudo-data obtained from MC setting all the parameters to their prefit expected value. The injected value of $\mathcal{R}(D^*)$ is 0.295, equal to the current experimental average. Beside being a consistency check of the fitting procedure, this fit is also useful to estimate the expected sensitivity. The uncertainty obtained in the combined fit is shown in Fig. 4.37 with a likelihood scan that also reports the breakdown of the uncertainty into statistical, systematic, and MC statistics component. The measurement is expected to have an uncertainty

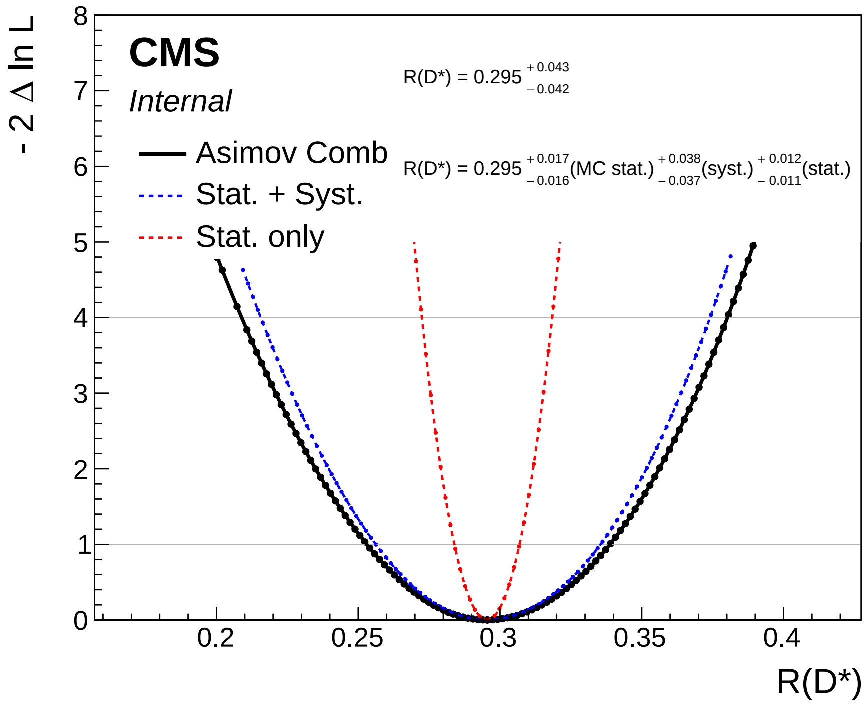


Figure 4.37: Likelihood scan for the combined Asimov fit to the three categories and uncertainty breakdown.

of about 15% and be systematically limited with all the categories having similar sensitivities. A more detailed breakdown of the expected uncertainty is presented in Tab. 4.11.

Finally, the presence of a possible bias arising in the fit on the estimated value of $\mathcal{R}(D^*)$ is evaluated. A total of 320 toy datasets are generated from the expected MC distributions dataset keeping into account both the statistics and the nuisance parameters using the Combine option `toysFrequentist`. A likelihood scan is than run on each toy. The estimated value for $\mathcal{R}(D^*)$ and its uncertainty obtained for each toy is reported in the right panel of Fig. 4.38. The injected value and the average toy result to be compatible within the uncertainty. Furthermore, normalized fit results $z$ are computed as the difference between the best fit value and the injected value divided by the estimated uncertainty. As reported in Fig. 4.38 right, the distribution of $z$ is found to be compatible with a normal Gaussian distribution $N(0, 1)$.

An additional Asimov fit with a simulated scenario with twice as many data has been also performed. As expected form the fact that the measurement is systematic

Table 4.11: Breakdown of the uncertainty for the Asimov fit to the complete set of signal and control regions.

| Uncertainty Source | Size $[10^{-2}]$ |
|---|---|
| Systematics | 4.13 |
| Finite MC sample size | 1.64 |
| Form factors $B \to D^{*}\ell\nu$ | 1.20 |
| Form factors $B \to D^{**}\ell\nu$ | 0.89 |
| $B \to D^{*}H_c$ modeling | 2.97 |
| B mesons production modeling | 0.19 |
| Detector modeling | 0.99 |
| Control to signal region transfer factors | 0.26 |
| Overall normalization | 0.05 |
| $B \to D^{**}(D^{*}\pi)\ell\nu$ modeling | 0.85 |
| $B \to D^{**}(D^{*}\pi\pi)\ell\nu$ modeling | 0.22 |
| $R(D^{(**)}_{(s)})$ | 0.78 |
| Others | 0.93 |
| Statistical | 1.15 |
| Total | 4.29 |

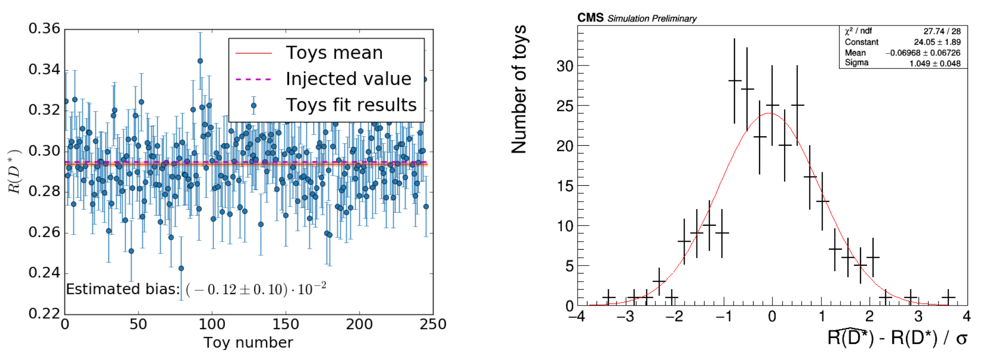


Figure 4.38: Left: Best fit value of $\mathcal{R}(D^{*})$ with its uncertainty for each toy. The injected value (dashed magenta line) and the toy results average (red line) with its uncertainty (orange band) are also shown. Right: Distribution of the normalized results with a superimposed likelihood gaussian fit.

limited, the total expected uncertainty is minimally impacted.

## 4.6 Summary and Future Prospective

This analysis note present a study to perform a $\mathcal{R}(D*)$ measurement with 2018 CMS parking data. $B$ mesons candidates are reconstructed from a muon, which is required to match the trigger object, and a $D*$ coming from the same vertex. $D*$ meson are reconstructed from three tracks compatible with the the $D^0(\to \pi K)\pi$ final state.

Additional tracks coming from the $D^*\mu$ vertex are also considered. The data are divided into three categories, one for each trigger path used for the analysis. Within each category, events are divided into 6 regions based on the number and charge sum of the additional tracks $N, Q$: one signal region where additional tracks are vetoed; 2 control regions with $N = 1$; and 3 control regions with $N = 2$. In the control region the 3-dimensional distribution of the observables $q^2$, $M^2_{miss}$ and $E^*_\mu$ are used for the fit. In the control regions the hadronic mass $m_{\mathrm{Had}} = m(\pi K\pi + n\mathrm{tracks})$ and $M^2_{miss}$ are used. All the considered observables in all the regions and categories are fit simultaneously to maximize the likelihood between the observed distributions in data and the predicted distributions from the sum of MC simulations of signal and background processes. The $\mathcal{R}(D*)$ value is a directly a free parameter of the minimization, together with more than other 100 nuisance parameters modeling systematic uncertainties.

The analysis is currently *blinded*. The high $q^2$ bins of the signal region in each category, which retains about 95% of the sensitivity, is excluded from the likelihood and there the data are masked. The blinded fit to real data is used to validate the status of the analysis, the MC simulation and the fit procedure without being exposed to experimental bias. Despite the mature state of the analysis, it is clear from the outcome of this study how the current model still presents unresolved significant tensions. Additional investigations are required to identify and mitigate the origin of these discrepancies before the analysis is able to proceed to the unblinding stage. An Asimov dataset is used to run a complete unblinded fit to assess the ultimate sensitivity of the full analysis. At the current state, the analysis is expected to have a 15% uncertainty and being limited by systematic uncertainties.

**References**


[1] Particle Data Group. "Review of Particle Physics". In: *Progress of Theoretical and Experimental Physics* 2020.8 (2020). 083C01. ISSN: 2050-3911. DOI: 10.1093/ptep/ptaa104. URL: https://doi.org/10.1093/ptep/ptaa104.

[2] S. Schael et al. "Precision electroweak measurements on the $Z$ resonance". In: *Phys. Rept.* 427 (2006), pp. 257–454. DOI: 10.1016/j.physrep.2005.12.006. arXiv: hep-ex/0509008 [hep-ex].

[3] S. Schael et al. "Electroweak Measurements in Electron-Positron Collisions at W-Boson-Pair Energies at LEP". In: *Phys. Rept.* 532 (2013), pp. 119–244. DOI: 10.1016/j.physrep.2013.07.004. arXiv: 1302.3415 [hep-ex].

[4] Georges et al. Aad. *Test of the universality of $\tau$ and $\mu$ lepton couplings in $W$-boson decays from $t\bar{t}$ events with the ATLAS detector*. Tech. rep. arXiv:2007.14040. Geneva: CERN, 2020. URL: https://cds.cern.ch/record/2725290.

[5] A. Sirlin. "Large m(W), m(Z) Behavior of the O(alpha) Corrections to Semileptonic Processes Mediated by W". In: *Nucl. Phys.* B196 (1982), pp. 83–92. DOI: 10.1016/0550-3213(82)90303-0.

[6] Florian U. Bernlochner et al. "Combined analysis of semileptonic $B$ decays to $D$ and $D^*$: $R(D^{(*)})$, $|V_{cb}|$, and new physics". In: *Phys. Rev.* D95.11 (2017). [erratum: Phys. Rev.D97,no.5,059902(2018)], p. 115008. DOI: 10.1103/PhysRevD.95.115008,10.1103/PhysRevD.97.059902. arXiv: 1703.05330 [hep-ph].

[7] Dante Bigi, Paolo Gambino, and Stefan Schacht. "$R(D^*)$, $|V_{cb}|$, and the Heavy Quark Symmetry relations between form factors". In: *JHEP* 11 (2017), p. 061. DOI: 10.1007/JHEP11(2017)061. arXiv: 1707.09509 [hep-ph].

[8] Sneha Jaiswal, Soumitra Nandi, and Sunando Kumar Patra. "Extraction of $|V_{cb}|$ from $B \to D^{(*)}\ell\nu_\ell$ and the Standard Model predictions of $R(D^{(*)})$". In: *JHEP* 12 (2017), p. 060. DOI: 10.1007/JHEP12(2017)060. arXiv: 1707.09977 [hep-ph].

[9] Yasmine Sara Amhis et al. "Averages of $b$-hadron, $c$-hadron, and $\tau$-lepton properties as of 2018". In: (2019). updated results and plots available at https://hflav.web.cern.ch/. arXiv: 1909.12524 [hep-ex].

[10] J. P. Lees et al. "Evidence for an excess of $\bar{B} \to D^{(*)}\tau^-\bar{\nu}_\tau$ decays". In: *Phys. Rev. Lett.* 109 (2012), p. 101802. DOI: 10.1103/PhysRevLett.109.101802. arXiv: 1205.5442 [hep-ex].

[11] M. Huschle et al. "Measurement of the branching ratio of $\bar{B} \to D^{(*)}\tau^-\bar{\nu}_\tau$ relative to $\bar{B} \to D^{(*)}\ell^-\bar{\nu}_\ell$ decays with hadronic tagging at Belle". In: *Phys. Rev.* D92.7 (2015), p. 072014. DOI: 10.1103/PhysRevD.92.072014. arXiv: 1507.03233 [hep-ex].

[12] S. Hirose et al. "Measurement of the $\tau$ lepton polarization and $R(D^*)$ in the decay $\bar{B} \to D^*\tau^-\bar{\nu}_\tau$". In: *Phys. Rev. Lett.* 118.21 (2017), p. 211801. DOI: 10.1103/PhysRevLett.118.211801. arXiv: 1612.00529 [hep-ex].

[13] S. Hirose et al. "Measurement of the $\tau$ lepton polarization and $R(D^*)$ in the decay $\bar{B} \to D^*\tau^-\bar{\nu}_\tau$ with one-prong hadronic $\tau$ decays at Belle". In: *Phys. Rev.* D97.1 (2018), p. 012004. DOI: 10.1103/PhysRevD.97.012004. arXiv: 1709.00129 [hep-ex].

[14] A. Abdesselam et al. "Measurement of $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$ with a semileptonic tagging method". In: (2019). arXiv: 1904.08794 [hep-ex].

[15] Roel Aaij et al. "Measurement of the ratio of branching fractions $\mathcal{B}(\bar{B}^0 \to D^{*+}\tau^-\bar{\nu}_\tau)/\mathcal{B}(\bar{B}^0 \to D^{*+}\mu^-\bar{\nu}_\mu)$". In: *Phys. Rev. Lett.* 115.11 (2015). [Erratum: Phys. Rev. Lett.115,no.15,159901(2015)], p. 111803. DOI: 10.1103/PhysRevLett.115.159901, 10.1103/PhysRevLett.115.111803. arXiv: 1506.08614 [hep-ex].

[16] R. Aaij et al. "Measurement of the ratio of the $B^0 \to D^{*-}\tau^+\nu_\tau$ and $B^0 \to D^{*-}\mu^+\nu_\mu$ branching fractions using three-prong $\tau$-lepton decays". In: *Phys. Rev. Lett.* 120.17 (2018), p. 171802. DOI: 10.1103/PhysRevLett.120.171802. arXiv: 1708.08856 [hep-ex].

[17] M. Tanaka. "Charged Higgs effects on exclusive semitauonic $B$ decays". In: *Z. Phys.* C67 (1995), pp. 321–326. DOI: 10.1007/BF01571294. arXiv: hep-ph/9411405 [hep-ph].

[18] W. Buchmüller, R. Rückl, and D. Wyler. "Leptoquarks in lepton-quark collisions". In: *Physics Letters B* 191.4 (1987), pp. 442–448. ISSN: 0370-2693. DOI: https://doi.org/10.1016/0370-2693(87)90637-X. URL: http://www.sciencedirect.com/science/article/pii/037026938790637X.

[19] Sacha Davidson, David Bailey, and Bruce A. Campbell. "Model independent constraints on leptoquarks from rare processes". In: *Zeitschrift für Physik C Particles and Fields* 61.4 (1994), pp. 613–643. ISSN: 1431-5858. DOI: 10.1007/BF01552629. URL: https://doi.org/10.1007/BF01552629.

[20] Riccardo Barbieri et al. "Anomalies in B-decays and U(2) flavor symmetry". In: *The European Physical Journal C* 76.2 (2016), p. 67. ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-016-3905-3. URL: https://doi.org/10.1140/epjc/s10052-016-3905-3.

[21] Riccardo Barbieri and Andrea Tesi. "B-decay anomalies in Pati–Salam SU(4)". In: *The European Physical Journal C* 78.3 (2018), p. 193. ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-018-5680-9. URL: https://doi.org/10.1140/epjc/s10052-018-5680-9.

[22] Admir Greljo, Gino Isidori, and David Marzocca. "On the breaking of Lepton Flavor Universality in B decays". In: *JHEP* 07 (2015), p. 142. DOI: 10.1007/JHEP07(2015)142. arXiv: 1506.01705 [hep-ph].

[23] Sofiane M. Boucenna et al. "Phenomenology of an $SU(2) \times SU(2) \times U(1)$ model with lepton-flavour non-universality". In: *JHEP* 12 (2016), p. 059. DOI: 10.1007/JHEP12(2016)059. arXiv: 1608.01349 [hep-ph].

[24] Dario Buttazzo et al. "B-physics anomalies: a guide to combined explanations". In: *Journal of High Energy Physics* 2017.11 (2017), p. 44. ISSN: 1029-8479. DOI: 10.1007/JHEP11(2017)044. URL: https://doi.org/10.1007/JHEP11(2017)044.

[25] "Recording and reconstructing 10 billion unbiased b hadron decays in CMS". In: (2019). URL: https://cds.cern.ch/record/2704495.

[26] "Data Parking and Data Scouting at the CMS Experiment". In: (2012). URL: https://cds.cern.ch/record/1480607.

[27] *BRIL Work Suite*. https://cms-service-lumi.web.cern.ch/cms-service-lumi/brilwsdoc.html. Accessed: 2020-07-06.

[28] Torbjörn Sjöstrand et al. "An Introduction to PYTHIA 8.2". In: *Comput. Phys. Commun.* 191 (2015), pp. 159–177. DOI: 10.1016/j.cpc.2015.01.024. arXiv: 1410.3012 [hep-ph].

[29] Anders Ryd et al. "EvtGen: A Monte Carlo Generator for B-Physics". In: (2005).

[30] Daryl Scora and Nathan Isgur. "Semileptonic meson decays in the quark model: An update". In: *Phys. Rev. D* 52 (5 1995), pp. 2783–2812. DOI: 10.1103/PhysRevD.52.2783. URL: https://link.aps.org/doi/10.1103/PhysRevD.52.2783.

[31] Stephan Duell et al. "HAMMER: Reweighting tool for simulated data samples". In: *PoS* ICHEP2016 (2017), p. 1074. DOI: 10.22323/1.282.1074.

[32] C. Bernard et al. "$\overline{B} \to D^* \ell \overline{\nu}$ form factor at zero recoil from three-flavor lattice QCD: A model independent determination of $|V_{cb}|$". In: *Phys. Rev. D* 79 (1 2009), p. 014506. DOI: 10.1103/PhysRevD.79.014506. URL: https://link.aps.org/doi/10.1103/PhysRevD.79.014506.

[33] Svjetlana Fajfer, Jernej F. Kamenik, and Ivan Ni šand ži ć. "$B \to D^* \tau \overline{\nu}_\tau$ sensitivity to new physics". In: *Phys. Rev. D* 85 (9 2012), p. 094025. DOI: 10.1103/PhysRevD.85.094025. URL: https://link.aps.org/doi/10.1103/PhysRevD.85.094025.

[34] Elisabetta Barberio and Zbigniew Was. "PHOTOS: A Universal Monte Carlo for QED radiative corrections. Version 2.0". In: *Comput. Phys. Commun.* 79 (1994), pp. 291–308. DOI: 10.1016/0010-4655(94)90074-4.

[35] J. P. Lees et al. "Observation of $\overline{B} \to D^{(*)} \pi^+ \pi^- \ell^- \overline{\nu}$ Decays in $e^+e^-$ Collisions at the $\Upsilon(4S)$ Resonance". In: *Phys. Rev. Lett.* 116 (4 2016), p. 041801. DOI: 10.1103/PhysRevLett.116.041801. URL: https://link.aps.org/doi/10.1103/PhysRevLett.116.041801.

[36] Jochen Dingfelder and Thomas Mannel. "Leptonic and semileptonic decays of $B$ mesons". In: *Rev. Mod. Phys.* 88 (3 2016), p. 035008. DOI: 10.1103/RevModPhys.88.035008. URL: https://link.aps.org/doi/10.1103/RevModPhys.88.035008.

[37] Florian U. Bernlochner, Zoltan Ligeti, and Sascha Turczyk. "A proposal to solve some puzzles in semileptonic $B$ decays". In: *Phys. Rev. D* 85 (9 2012), p. 094033. DOI: 10.1103/PhysRevD.85.094033. URL: https://link.aps.org/doi/10.1103/PhysRevD.85.094033.

[38] J. L. Goity and W. Roberts. "Soft pion emission in semileptonic B meson decays". In: *Phys. Rev. D* 51 (1995), pp. 3459–3477. DOI: 10.1103/PhysRevD.51.3459. arXiv: hep-ph/9406236.

[39] J. P. Lees et al. "Observation of $\overline{B} \to D^{(*)}\pi^+\pi^-\ell^-\overline{\nu}$ decays in $e^+e^-$ collisions at the $\Upsilon(4S)$ resonance". In: *Phys. Rev. Lett.* 116.4 (2016), p. 041801. DOI: 10.1103/PhysRevLett.116.041801. arXiv: 1507.08303 [hep-ex].

[40] A. M. Sirunyan et al. "Particle-flow reconstruction and global event description with the CMS detector". In: *JINST* 12.10 (2017), P10003. DOI: 10.1088/1748-0221/12/10/P10003. arXiv: 1706.04965 [physics.ins-det].

[41] Giovanni Petrucciani, Andrea Rizzi, and Carl Vuosalo. "Mini-AOD: A New Analysis Data Format for CMS". In: *J. Phys. Conf. Ser.* 664.7 (2015), p. 7. DOI: 10.1088/1742-6596/664/7/072052. arXiv: 1702.04685 [physics.ins-det].

[42] R. Fruhwirth, W. Waltenberger, and P. Vanlaer. "Adaptive vertex fitting". In: *J. Phys. G* 34 (2007), N343. DOI: 10.1088/0954-3899/34/12/N01.

[43] Serguei et al. Chatrchyan. "Description and performance of track and primary-vertex reconstruction with the CMS tracker". In: *JINST* 9 (2014). Comments: Replaced with published version. Added journal reference and DOI, P10009. 80 p. DOI: 10.1088/1748-0221/9/10/P10009. arXiv: 1405.6569. URL: https://cds.cern.ch/record/1704291.

[44] *"CMSSW Muon ID guide"*. https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideMuonId. Accessed: 2020-02-19.

[45] A. M. Sirunyan et al. "Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV". In: *JINST* 13.06 (2018), P06015. DOI: 10.1088/1748-0221/13/06/P06015. arXiv: 1804.04528 [physics.ins-det].

[46] Kirill Prokofiev and T. Speer. "A kinematic and a decay chain reconstruction library". In: *Computing in high energy physics and nuclear physics. Proceedings, Conference, CHEP'04, Interlaken, Switzerland, September 27-October 1, 2004*. 2005, pp. 411–414. URL: http://doc.cern.ch/yellowrep/2005/2005-002/p411.pdf.

[47] *"Kinematic Vertex Fit guide"*. https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideKinematicVertexFit. Accessed: 2020-02-19.

[48] *Slide show: "Soft tracks RD/MC disagreement"*. https://docs.google.com/presentation/d/169nBUAmgbaFyJo-FNHs36bSovdLleB1AK8nKsW0fLE4/edit?usp=sharing. Accessed: 2020-06-06.

[49] Serguei Chatrchyan et al. "Alignment of the CMS tracker with LHC and cosmic ray data". In: *JINST* 9 (2014), P06009. DOI: 10.1088/1748-0221/9/06/P06009. arXiv: 1403.2286 [physics.ins-det].

[50] Elisabetta Manca. "Validation of the muon momentum resolution in view of the W mass measurement with the CMS experiment". In: (2016). URL: https://cds.cern.ch/record/2233647.

[51] *"Reference muon id, isolation and trigger efficiencies for 2018 data."*. https://twiki.cern.ch/twiki/bin/viewauth/CMS/MuonReferenceEffs2018. Accessed: 2020-02-24.

[52] *ROOT TEfficiency documentation*. https://root.cern.ch/doc/master/classTEfficiency.html. Accessed: 2021-02-23.

[53] Irinel Caprini, Laurent Lellouch, and Matthias Neubert. "Dispersive bounds on the shape of anti-B —> D(*) lepton anti-neutrino form-factors". In: *Nucl. Phys.* B530 (1998), pp. 153–181. DOI: 10.1016/S0550-3213(98)00350-2. arXiv: hep-ph/9712417 [hep-ph].

[54] Florian U. Bernlochner et al. "Combined analysis of semileptonic $B$ decays to $D$ and $D^*$: $R(D^{(*)})$, $|V_{cb}|$, and new physics". In: *Phys. Rev.* D95.11 (2017). [erratum: Phys. Rev.D97,no.5,059902(2018)], p. 115008. DOI: 10.1103/PhysRevD.95.115008,10.1103/PhysRevD.97.059902. arXiv: 1703.05330 [hep-ph].

[55] C. Glenn Boyd, Benjamin Grinstein, and Richard F. Lebed. "Precision corrections to dispersive bounds on form factors". In: *Phys. Rev. D* 56 (11 1997), pp. 6895–6911. DOI: 10.1103/PhysRevD.56.6895. URL: https://link.aps.org/doi/10.1103/PhysRevD.56.6895.

[56] Yasmine Sara Amhis et al. "Averages of $b$-hadron, $c$-hadron, and $\tau$-lepton properties as of 2018". In: (2019). arXiv: 1909.12524 [hep-ex].

[57] Florian U. Bernlochner, Zoltan Ligeti, and Dean J. Robinson. "Model independent analysis of semileptonic $B$ decays to $D^{**}$ for arbitrary new physics". In: *Phys. Rev. D* 97.7 (2018), p. 075011. DOI: 10.1103/PhysRevD.97.075011. arXiv: 1711.03110 [hep-ph].

[58] *The Combine package*. https://cms-analysis.github.io/HiggsAnalysis-CombinedLimit/. Accessed: 2022-07-18.

[59] *HiggsAnalysis-CombinedLimit documentation*. http://cms-analysis.github.io/HiggsAnalysis-CombinedLimit/. Accessed: 2020-07-08.

[60] Olmo Cerri et al. “Lepton flavor universality test at CMS with R(D*) measurement in the full leptonic tau final state.” In: *CMS AN-2019/162* (2022).

[61] Roger J. Barlow and Christine Beeston. “Fitting using finite Monte Carlo samples”. In: *Comput. Phys. Commun.* 77 (1993), pp. 219–228. DOI: 10.1016/0010-4655(93)90005-W.

[62] Robert D. Cousins. “Generalization of Chisquare Goodness-ofFit Test for Binned Data Using Saturated Models , with Application to Histograms”. In: 2013.

[63] A. Bazavov et al. “Semileptonic form factors for $B \to D^{*}\ell\nu$ at nonzero recoil from 2 + 1-flavor lattice QCD”. In: (May 2021). arXiv: 2105.14019 [hep-lat].

*Chapter 5*

# CONCLUSIONS

*“Considerate la vostra semenza: fatti non foste a viver come bruti ma per seguir virtute e canoscenza."*
Consider your origins: you were not made to live as brutes, but to follow virtue and knowledge.
— Dante Alighieri, *Divina Commedia – Inferno*

This thesis documents the tough, exciting, and passionate journey I sailed during my time as a graduate student at the California Institute of Technology. During these five years of hard work, I was able to contribute to some of the most interesting and active research areas in the field of high energy particle physics. Many of the ideas presented here started off as scratch on a piece of paper or a blackboard in some CERN building, and eventually evolved into their current mature statuses after undergoing several discussions, changes, and countless hours of dedication. Considering the lack of a clear way to resolve the outstanding open questions about the Standard Model, I diversified my research focus in three ways. First, I worked on software developments to incorporate recent machine-learning techniques in our experimental quest, with particular interest in enhancing model-independent methods to look for unexpected evidence of new physics. Second, I contributed to the methodological and hardware work to upgrade our current detector leveraging on new technologies in order to open the possibility for new research paths. Finally, I exploited high statistics CMS datasets to propose a precision measurement, which could help tighten our grasp of the Standard Model.

In Chapter 2, leveraging on the expertise I acquired during my master thesis, we develop a strategy to isolate potential BSM events produced by the LHC using a variational autoencoder trained on a reference sample. This algorithm is designed to be used in the trigger system of LHC experiments to identify recurrent anomalies, which might otherwise escape observation. We also demonstrate how the algorithm can be trained directly on data with no sizable performance loss, enhanced robustness against systematic uncertainties, and significant simplifications of the training and deployment procedure. We show that events produced by not-yet-excluded BSM models with cross sections in the range of $O(10)$ to $O(100)$ pb could be isolated in a $\sim$ 30% pure sample of $\sim$ 43 events selected per day. We believe that such an application can help extend the physics reach of the current and next stages of the CERN LHC. The effectiveness of the proposed strategy on real back-testing data is also demonstrated in a follow-up work. Given the unfortunate timing of the LHC schedule, no new data were collected by CMS from the moment in which this work was concluded to the defense of this thesis, and hence, this approach has not yet been deployed on real-time trigger. However, given the positive feedback received by the CMS collaboration and the entire community, we hope that future efforts will motivate similar strategies to be employed in the LHC experiments trigger.

Chapter 3 discusses one of the main achievements of this thesis, where we develop of a method to use a time-of-flight (TOF) measurement to perform particle identification at general-purpose LHC experiments. After deriving the analytical formula for the expected performance, we estimate the particle identification potential to be significant within SM particles up to a transverse momentum of few GeV. Using a fast simulation, we propose an approach to perform a TOF measurement with minimal assumptions. Deploying a deterministic annealing to reconstruct vertices, we achieve a $\hat{z}$ resolution of about 50 (80) $\mu$m for PV (PU) tracks for TOF resolution of about 30 ps. We demonstrate how such a detector would enable four-dimensional vertex reconstruction and the identification of charged particles through its TOF measurement. Using long-lived top squark pair production as a benchmark example, we showcase significantly improved sensitivity within the searches for long-lived particles that are quantified in a factor of 5 to 10 for top squark masses above 300 GeV. Finally, we show that if the TOF measurement could be utilized in the trigger system, an additional sensitivity improvement of a factor of 2 to 5 could be realized for top squark masses below 200 GeV. This result provides good motivation for further work on the design and realization of a TOF based trigger for long-lived particles. Chapter 3 also discusses my in-depth study to measure the expected temperature difference between the cooling pipe and the SiPMs sensors in the MTD module for different design geometries. Using a mock-up module made in the Caltech Lab, we reproduced the operating characteristics of the detector and we used 8 thermistors to simultaneously measure the temperature across the system. When the power injected in the array is equivalent to expected SiPM power consumption, we measure a $\Delta T$ between 1 and 4 Kelvin degrees. As expected, a taller cooling support showed the best cooling capabilities with the array measured to be between 1 K and 2 K hotter than the cooling pipe with the results for the short design degrading of 15-20%. Consistently, the design without a separation between the sensitive elements and the electronics (C-shape) underperforms the other two designs by a few degrees. We measured a temperature increase of about 0.6 K/W of power injected across all the arrays.

Finally, in Chapter 4 we focus on the first ever $\mathcal{R}(D^*)$ measurement at CMS. Using the $b\bar{b}$ enriched 2018 CMS parking data, $B$ mesons candidates are reconstructed from a muon and a $D^*$ meson coming from the same vertex, selecting eventual tau in the full leptonic final state. After considering three categories defined by different triggers, the analysis leverages on control regions based on the presence of additional tracks to constrain the backgrounds. In the signal region, the 3-dimensional distribution of the observables $q^2$, $M^2_{miss}$, and $E^*_\mu$ is fit. All the considered observables in all the regions and categories are fit simultaneously to maximize the likelihood between the observed distributions in the data and the predicted distributions from the sum of MC simulations of signal and background processes. The $\mathcal{R}(D^*)$ value is a free parameter of the minimization, together with more than other 100 nuisance parameters modeling systematic uncertainties. While the analysis is still blinded, a complete fit is performed on a Asimov dataset to assess the ultimate sensitivity of the full analysis, which is found to be of about 15%. A collaboration-wide effort to better understand the performance of the tracker in the soft regime is still required to achieve the confidence required to unblind the analysis. Considering the amount of data gathered, the systematic uncertainty will be the limiting factor and the MC generation will continue to present a harsh challenge. The lepton flavor universality, and in particular the $\mathcal{R}(D^*)$ measurement, is one of the most interesting and relevant challenges that physics at the energy and intensity frontiers is facing. The LHC is nowadays providing a good experimental environment to explore new and more accurate approaches to this topic. The CMS experiment has the opportunity and all the tools to perform flavor measurements of unprecedented precision.

# LIST OF ILLUSTRATIONS

# LIST OF TABLES

*Semel Normalista Semper.*