National and Kapodistrian University of Athens

Thesis submitted in fulfillment of Doctor of Philosophy degree requirements

Search for Supersymmetric particles with compressed mass spectrum in proton-proton collisions at a centre-of-mass energy of 13 TeV

Author: Emmanouil Vourliotis
Supervisor: Professor Paris Sphicas

Athens, March 2022

Advisory Committee Members:

P. Sphicas, Prof. of Physics (Athens)
N. Tetradis, Prof. of Physics (Athens)
K. Vellidis, Associate Prof. of Physics (Athens)

Examination Board Members:

C. Botta, Assistant Prof. of Physics (Zürich)
D. Fasouliotis, Prof. of Physics (Athens)
K. Foudas, Prof. of Physics (Ioannina)
N. Saoulidou, Associate Prof. of Physics (Athens)

ΙΣΝ/SNF ΙΔΡΥΜΑ ΣΤΑΥΡΟΣ ΝΙΑΡΧΟΣ
STAVROS NIARCHOS FOUNDATION

The research presented was carried out within the framework of a Stavros Niarchos Foundation grant to the National and Kapodistrian University of Athens

Abstract

The main subject of this thesis is a search for Supersymmetric (SUSY) particles utilizing the experimental data of proton-proton collisions at a center-of-mass energy of 13 TeV collected by the Compact Muon Solenoid (CMS) experiment during the Run 2 (2016–2018) operation of the Large Hadron Collider (LHC) accelerator, amounting up to 137 fb^{-1} of integrated luminosity. The data analysis focuses on final states with two or three low transverse momentum (p_{T}) leptons (electrons or muons), at least one pair of which has opposite sign charge and the same flavor. The leptons are accompanied by a moderate or large amount of missing transverse energy. Such signatures can be produced in SUSY models, involving either electroweakino or top squark production, with a compressed mass spectrum, i.e. small mass differences between the lightest SUSY particles. The results of the search show agreement between the observed data and the Standard Model predictions. They are used to extract mass exclusion limits for the SUSY particles of the previously mentioned SUSY models. An important part of this thesis is also dedicated to the development of software and firmware algorithms for the Phase 2 upgrade of the CMS Level-1 trigger (L1T) system. First, a set of corrections is derived for L1T muons captured by the Phase 1 demonstrator of the Phase 2 L1T scouting system. In addition, a multivariate algorithm for the L1T Correlator subsystem is developed with the aim of identifying electromagnetic, hadronic and pile-up clusters from the future High Granularity Calorimeter of CMS. One of the L1T Correlator data organization algorithms, the "deregionizer" algorithm, is also discussed and studies on improvements to it are presented. Finally, the development of a firmware module for sorting according to their p_{T} the electromagnetic objects of the L1T Correlator, along with its corresponding software emulator, is described.

Acknowledgements

Υπάρχουν τόσα άτομα που έχουν συμβάλει στην ολοκλήρωση αυτού του ταξιδιού με πολύ διαφορετικές αλλά εξίσου σημαντικές συνεισφορές, που η πιο δίκαιη σειρά να πω "ευχαριστώ" είναι η χρονολογική.

Το ταξίδι μου στον κόσμο της Φυσικής ξεκίνησε σχεδόν από τύχη. Τότε, το μακρινό 2010, έκανα αίτηση για ένα θερινό σχολείο Φυσικής για μαθητές λυκείου στο Perimeter Institute και έβαλα ένα "στοίχημα" με τον εαυτό μου: "Αν επιλεγώ, η Φυσική θα γίνει για μένα κάτι παραπάνω από χόμπι"· όπερ και εγένετο.

Ωστόσο, το ταξίδι δε θα μπορούσε να ξεκινήσει καν αν δεν είχα τους γονείς και την οικογένειά μου να με στηρίζουν αδιάκοπα. Μου προσέφεραν εφόδια χωρίς τα οποία τίποτα απ' ό,τι ακολούθησε εκείνο το "στοίχημα" δε θα ήταν δυνατό. Δε βρίσκω τις λέξεις να τους ευχαριστήσω αρκετά. Ειδική μνεία αξίζει στην αδερφή μου: Είναι η ήσυχη δύναμη που πάντοτε καθόταν υπομονετικά δίπλα μου, όποτε της το ζητούσα, για να ακούσει τα "ακαταλαβίστικα" φυσικά και προγραμματιστικά προβλήματα που με απασχολούσαν.

Η οικογένειά μου με εφοδίασε για τη ζωή και ο Τάκης, ο δάσκαλός μου στη Φυσική στο λύκειο, μου έδωσε ώθηση στα πρώτα επιστημονικά μου βήματα. Μου μετέδωσε τον ενθουσιασμό και την περιέργειά του. Μου προσέφερε αμέτρητες ευκαιρίες να εξερευνήσω τα ενδιαφέροντά μου και να διακριθώ. Ήταν εκεί για να λύσει τις πιο τρελές μου απορίες. Ήταν ένας πραγματικός μέντορας για μένα. Είμαι χαρούμενος που τον γνώρισα και πολύ περισσότερο που τον έχω ακόμα στη ζωή μου.

Ο Τάκης δεν ήταν ο μοναδικός μέντορας αυτού του ταξιδιού. Το 2015 είχα την τύχη να ξεκινήσω να δουλεύω με τον καθηγητή Πάρη Σφήκα. Επιθυμούσα τη συνεργασία με τον κ. Σφήκα, γιατί εξ' αρχής με είχε εντυπωσιάσει η μεθοδικότητά του στις διαλέξεις και η

συγκροτημένη σκέψη του που οδηγούσε στη θεμελίωση λεπτών θεωρητικών εννοιών με τον πιο απτό τρόπο. Ακόμα μεγαλύτερη εντύπωση μού είχε κάνει το γεγονός ότι ασχολείται με την πειραματική φυσική. Έτσι, στα μάτια μου, ένωνε και ενώνει άρτια τους κόσμους της θεωρίας και του πειράματος και εκφράζει αυτό που θα ήθελα να γίνω: ένας καθολικός επιστήμονας.

Ευχαριστώ τον κ. Σφήκα που με επέλεξε αρχικά για πτυχιακή εργασία μαζί του και για όλη την ακόλουθη συνεργασία μας. Χωρίς αυτόν θα ήμουν κάτι λιγότερο από αυτό που είμαι. Όχι μόνο για τη γνώση και τις ευκαιρίες που μου έδωσε αλλά, πάνω απ' όλα, για την εμπιστοσύνη που μου έδειξε· για τις επιλογές που άνοιγε μπροστά μου και με άφηνε πάντα ελεύθερα να κάνω. Με έμαθε να είμαι ανεξάρτητος, αυτοδίδακτος και υπεύθυνος. Όχι μόνο με προετοίμασε να γίνω ένας αυτόνομος ερευνητής αλλά και ένας αυτόνομος άνθρωπος. Η συνεισφορά του εκτείνεται πολύ πέρα από αυτή τη διατριβή και θα με συνοδεύει μια ζωή.

Θα ήθελα να ευχαριστήσω τον κ. Σφήκα για ένα ακόμη πράγμα: για την εξασφάλιση της χρηματοδότησης της δουλειάς μου όλα αυτά τα χρόνια, μέσα σε δύσκολες και περίπλοκες συνθήκες. Δε θα μπορούσα να παραλείψω να εκφράσω τις ευχαριστίες μου στο Ίδρυμα Σταύρος Νιάρχος που, με την υποτροφία που προσέφερε, συνεισέφερε στο μεγαλύτερο μέρος αυτής της χρηματοδότησης.

Ο καθηγητής Κώστας Βελλίδης ήταν ένας ακόμη άνθρωπος που συνέβαλε καθοριστικά στην εξέλιξή μου κατά τη διάρκεια του διδακτορικού. Ευχαριστώ για τις ανεκτίμητες συμβουλές, τις συζητήσεις για τη Φυσική αλλά και για τη ζωή, πάντοτε με τη συνοδεία μιας κρύας μπύρας. Θα ήθελα επίσης να ευχαριστήσω και όλα τα υπόλοιπα μέλη της επταμελούς επιτροπής, τους καθηγητές Νίκο Τετράδη, Cristina Botta, Δημήτρη Φασουλιώτη, Κώστα Φουντά και Νίκη Σαουλίδου, για την ανάγνωση και την εξέταση της διατριβής μου.

Το να είναι κανείς διδακτορικός φοιτητής σημαίνει να μαθαίνει κάθε μέρα. Και η διαδικασία της μάθησης σπάνια είναι μοναχική. Έτσι και στη δική μου περίπτωση, κατά τη διάρκεια της εργασίας μου πάνω στα θέματα της διατριβής, είχα τη χαρά να δουλέψω με πολύ αξιόλογους επιστήμονες: Με τον Emilio Meschi, ο οποίος, πάντα με πολλή υπομονή και διάθεση να προσφέρει τις γνώσεις του, με βοήθησε να φέρω εις πέρας το πρώτο project που μου ανατέθηκε. Με τον Giovanni Petrucciani, ο οποίος με έμαθε να δουλεύω μεθοδικά και

συγκροτημένα. Κάθε project μαζί του ήταν μια ευκαιρία για ανακαλύψεις, γιατί ήξερε πάντα να παρέχει τις κατάλληλες οδηγίες, ώστε να δίνει το έναυσμα για προσωπική έρευνα. Χωρίς τον Giovanni, οι γνώσεις μου στον κώδικα θα ήταν σαφώς φτωχότερες. Με τον Marco Peruzzi, ο οποίος είναι πραγματικός συνάδελφος για μένα. Πάντα πρόθυμος να συζητήσει, να διδάξει και να διδαχτεί, να βρει κάτι καινούργιο και να βελτιώσει τα παλιά, ο Marco με συνόδεψε σε ό,τι και να έκανα. Του είμαι ευγνώμων για τις ειλικρινείς μας συζητήσεις, τις πολύτιμες συμβουλές του και τη στήριξή του σε κάθε περίσταση. Με την Cristina Botta, η οποία παρακολούθησε από κοντά την πορεία μου. Με το ταλέντο της στην οργάνωση και την οξυδέρκειά της, συνεισέφερε ουσιαστικά στην επιτυχία της δουλειάς που περιλαμβάνεται στη διατριβή.

Τα χρόνια του διδακτορικού θα ήταν σημαντικά πιο ανιαρά αν δεν είχα τη συντροφιά των φίλων μου στη Γενεύη. Τον Αντώνη τον γνώρισα στα προπτυχιακά μου χρόνια και τον βρήκα πάλι στο διδακτορικό. Τον ευχαριστώ για τη βοήθειά του στα πρώτα μου βήματα αλλά και τις διεισδυτικές ερωτήσεις του σε όλες μας τις συζητήσεις. Ο Γιώργος είναι μια αστείρευτη πηγή γέλιου αλλά και γνώσης για μένα. Η συνύπαρξή μας στο γραφείο ήταν απολαυστική και είμαι ευτυχής που η σχέση μας συνεχίζει παρά την απόσταση που μας χωρίζει. Με την Ιωάννα δουλέψαμε πολύ κοντά και ήμαστε το στήριγμα ο ένας του άλλου σε όλα τα θέματα της δουλειάς. Τα working meetings θα περνούσαν πολύ πιο αργά χωρίς τις συζητήσεις μας. Ο Παντελής, αν και ο πιο ήσυχος του γραφείου, είχε πάντα να προσφέρει ενδιαφέρουσες προτάσεις και ιδέες σε κάθε κουβέντα. Με τον Πάνο δουλέψαμε μαζί αλλά, πάνω απ' όλα, ζήσαμε τη ζωή. Ανακαλύψαμε μέρη, ανθρώπους, εμάς. Τού είμαι ευγνώμων για όλο το χρόνο που περάσαμε και περνάμε μαζί. Είναι κρίμα που δε γνώρισα το Γιώργο παρά μόνο αργότερα στο διδακτορικό. Κατευθείαν όμως ταιριάξαμε, δεθήκαμε, μοιραστήκαμε σκέψεις και ανησυχίες. Είμαι σίγουρος ότι θα συνεχίσουμε να μοιραζόμαστε και στο μέλλον. Ευχαριστώ τη Sarah που με βοήθησε να μάθω τόσα πολλά και διαφορετικά πράγματα. Γελάσαμε και κλάψαμε μαζί και, μέσα από όλα όσα περάσαμε, τελικά γίναμε και οι δύο καλύτεροι άνθρωποι. Ο Paul και η Rosalye, οι συγκάτοικοί μου, έκαναν την κάθε μέρα χαρούμενη και ανέμελη, ανεξαρτήτως συνθηκών. Η ζωή στη Γενεύη θα ήταν πολύ πιο δύσκολη χωρίς αυτούς. Τους ευχαριστώ για όλη την αγάπη και τη φροντίδα που μου έδειξαν.

Αν και πέρασα το μεγαλύτερο μέρος του διδακτορικού στην Ελβετία, η στήριξη των αγαπημένων μου προσώπων στην Αθήνα και σε άλλα μέρη του κόσμου με συνόδευε πάντα. Ευχαριστώ τους παιδικούς μου φίλους, Γιάννη και Μπίλλυ, που πάντα με έψαχναν και πάντα τους έβρισκα. Οι συναντήσεις μας ήταν οι σημαντικές στιγμές χαλάρωσης που επιζητούσα μέσα στην "τρελή" καθημερινότητα. Ο Βαγγέλης, ο Θωμάς και ο Νίκος ήταν οι άνθρωποι με τους οποίους μπορούσα να συζητήσω τα πάντα. Μαζί τους ήμουν ο εαυτός μου, τον οποίο αποδέχονταν και στήριζαν σε κάθε περίπτωση. Με την Κρινιώ μεγαλώσαμε μαζί με όλους τους τρόπους, εκτός από τον κυριολεκτικό. Ήταν εκεί για να συζητήσει μαζί μου από τα πιο ασήμαντα μέχρι τα πιο σημαντικά ζητήματα και είχε και έχει πάντα το τρόπο να μου αλλάζει τη διάθεση προς το καλύτερο. Η Νεφέλη ακολούθησε το ταξίδι πιο κοντά από οποιονδήποτε άλλον. Δεν υπάρχουν τα λόγια να την ευχαριστήσω για τη παρουσία της στα πάνω μου και στα κάτω μου· είναι απλά ξεχωριστή. Ο Χριστόδουλος, ως φίλος και ως συγκάτοικος, με συντρόφευσε σε μια από τις πιο σημαντικές αλλαγές στη ζωή μου. Μου λείπουν ήδη τα καθημερινά ανέκδοτα και TIL του. Ευχαριστώ τη Μαρίνα· η συντροφιά της μού έδωσε την ώθηση να φτάσω μέχρι το τέλος του ταξιδιού και μου δίνει τη δύναμη στα επόμενα βήματα.

Αυτή τη διατριβή θα ήθελα να την αφιερώσω στον παππού μου που έφυγε πριν προλάβει να τη δει ολοκληρωμένη. Ξέρω τη χαρά που έπαιρνε μέσα από μένα και ελπίζω να συνεχίζω να τον κάνω περήφανο εκεί που είναι.

Contents

Chapter 1

The Standard Model of Particle Physics

Particle Physics is the study of the fundamental building blocks of matter and the interactions between them. Over the past century, a compilation of theories have come to form what is known as the Standard Model (SM), which describes very successfully the phenomena that govern the most fundamental particles known thus far. This chapter presents a short of overview of the basic aspects of the SM.

1.1 The Fundamental Particles and Forces of the Standard Model

The elementary particles of the SM are divided in two major categories, fermions and bosons. This division is based on the intrinsic spin of the particles: Fermions have spin values to a half-integer multiple of $\hbar$, while bosons have spin equal to an integer multiple of $\hbar$.

Fermions are the fundamental constituents of matter. Their wave functions are antisymmetric with respect to the interchange of space and spin coordinates. This leads to the Pauli exclusion principle: Two or more identical fermions cannot occupy the same quantum state simultaneously. As a consequence, fermions obey the Fermi–Dirac statistics, which describes the behavior of fermion ensembles. All elementary known fermions have spin 1/2 (throughout the text natural units are used, hence $c = 1$ and $\hbar = 1$).

Fermions are further categorized as leptons, if they do not interact via the

strong interaction (summarized in 1.3), or as quarks, which carry color charge and interact also strongly. Quarks are subject to a phenomenon called "color confinement": Quarks cannot be isolated but only appear in groups, called hadrons. Hadrons that consist of two quarks are called mesons, while hadrons that consist of three quarks are called baryons[1]. Even though quarks have fractional electric charge ($-1/3$ or $2/3$), hadrons only appear with integer multiple of the elementary electric charge (i.e. the electric charge carried by the proton). There are six known flavors of quarks: up (u), down (d), strange (s), charm (c), bottom (b) and top (t). Leptons also exist in six flavors, three of them charged: the electron (e), the muon (μ) and the tau (τ); and three of them electromagnetically neutral: the e neutrino (ν_e), the μ neutrino (ν_μ) and the τ neutrino (ν_τ). For each fermion, there is a corresponding antiparticle, which is identical in every respect to the particle, except for their strong, weak and electromagnetic charges, which are opposite to those of the particle.

Fermions are also grouped in generations. Table 1.1 shows how fermions are split in generations, along with the electric charge and mass of each particle. All masses and particle properties mentioned below are taken from "The Review of Particle Physics (2020)" [1].

	Leptons			Quarks		
		Charge [e]	Mass [GeV]		Charge [e]	Mass [GeV]
1st generation	ν_e	0	$< 1.1 \times 10^{-9}$	d	$-1/3$	4.67×10^{-3}
	e	-1	0.511×10^{-3}	u	$+2/3$	2.16×10^{-3}
2nd generation	ν_μ	0	$< 0.19 \times 10^{-6}$	s	$-1/3$	0.093
	μ	-1	0.106	c	$+2/3$	1.27
3rd generation	ν_τ	0	$< 18.2 \times 10^{-6}$	b	$-1/3$	4.18
	τ	-1	1.778	t	$+2/3$	172.76

Table 1.1: The fermion generations. The electric charge and mass of each particle are included.

The carriers of the fundamental forces of Nature are bosons, which will be concisely described in the following sections. Bosons are named after the

[1]In the past decade there is mounting evidence that four-quark and five-quark bound states also exist. It is not, however, established yet whether these are "hadron molecules" or genuine strongly-bound states like mesons and baryons.

Bose–Einstein statistics that they obey, according to which an ensemble of them can be in the same quantum state, implying that bosons are not subject to the Pauli exclusion principle. The carriers of the electromagnetic, weak and strong interactions have spin equal to 1, while the carrier of gravity, the graviton, must have spin 2. The Higgs boson (H) is the only elementary particle with spin 0.

The photon is the carrier of the electromagnetic interaction. It has zero mass and zero electric charge and it can only interact with electrically charged particles. The gluon is the boson of the strong interaction: it carries color charge and interacts with other colored particles, including gluons. The latter property (boson-boson interaction) makes the strong interaction significantly more complicated than the electromagnetic one. Similar to the photon, the gluon has zero mass and electric charge. The weak interaction has three carriers, collectively referred to as the weak boson: the Z boson, which is electromagnetically neutral and has a mass of 91.2 GeV, and the $W^{\pm}$ bosons, which carry electromagnetic charge and have a mass of 80.4 GeV. All three of them are color neutral. Finally, the H boson is the quantum of the Higgs field, the spontaneous breaking of which causes other elementary particles to acquire mass (the mechanism by which this is realized is discussed in Sec. 1.4). The H boson has zero electric and color charge and a mass of 125.10 GeV. It interacts with all massive particles and also couples to itself, because of the potential that is needed to generate the symmetry breaking mechanism.

1.2 The Electromagnetic Interaction

The electromagnetic interaction naturally arises within the framework of relativistic quantum mechanics from the Dirac equation. The Dirac equation describes the free propagation of a fermion with mass m and spin 1/2 and is written as:

$$i\gamma^{\mu}\partial_{\mu}\psi - m\psi = 0 \tag{1.1}$$

where $\psi = \psi(x,t)$ is the fermion wavefunction and $\bar{\psi} = \psi^{\dagger}\gamma^{0}$. The γ^{μ} are 4×4 matrices, also called Dirac matrices, and are defined by their anticommutation relation:

$$\{\gamma^{\mu}, \gamma^{\nu}\} = 2g^{\mu\nu} \tag{1.2}$$

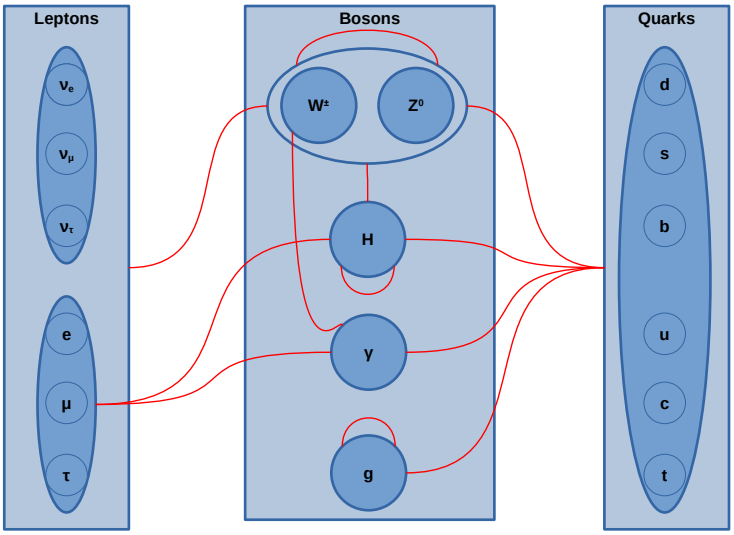


Figure 1.1: The interactions between the SM particles.

The Dirac equation can be derived from the following Lagrangian:

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu \partial_\mu - m)\psi \tag{1.3}$$

It is worth noting that this Lagrangian is invariant under a global transformation $\psi \rightarrow e^{ia}\psi$, with a being any constant. This transformation is equivalent to a phase shift of the wavefunction by the same amount in every point of the spacetime. According to Noether's theorem, this symmetry implies the existence of a conserved current, and thus also a conserved electric charge.

This is enforced by requiring the Lagrangian to be invariant under the U(1) local "gauge" transformation:

$$\psi \rightarrow e^{iea(x)}\psi, \tag{1.4}$$

where e is the electric charge and $a(x)$ can now be different in every spacetime point. In this case, the derivative is substituted by the *covariant* derivative D^μ, which changes as $D^\mu \rightarrow e^{ia(x)}D^\mu$ under the previous transformation. This requires the introduction of an extra field, A^μ, which is required to transform in a way that ensures invariance of the Lagrangian under local phase changes. This is achieved with the substitution $\partial^\mu \rightarrow D^\mu$, where D^μ is the covariant derivative is defined as:

$$D^\mu \equiv \partial^\mu - ieA^\mu \tag{1.5}$$

and A^μ is required to transform as:

$$A^\mu \to A^\mu + \frac{1}{e}\partial^\mu a \tag{1.6}$$

under Eq. 1.4.

Inserting Eq. 1.5 in Eq. 1.3, we have:

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu\partial_\mu - m)\psi + e\bar{\psi}\gamma^\mu\psi A_\mu \tag{1.7}$$

The term $e\bar{\psi}\gamma^\mu\psi A_\mu$ is interpreted as the interaction of the current $j^\mu = \bar{\psi}\gamma^\mu\psi$ with electric charge e with the electromagnetic field A_μ. To complete the description of the electromagnetic field, a kinetic term for the new field A_μ is needed. This term must contain derivatives of A_μ (so that A_μ can propagate) and be invariant under the local gauge transformation. The term that fulfills this requirement is the product $F_{\mu\nu}F^{\mu\nu}$, where $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$. The full Lagrangian of electromagnetism is thus:

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu\partial_\mu - m)\psi + e\bar{\psi}\gamma^\mu\psi A_\mu - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \tag{1.8}$$

It is worth noting that the local gauge invariance forbids the introduction of other gauge-invariant terms with A_μ, e.g. a mass term $\frac{1}{2}m_A^2 A_\mu A^\mu$, leading to the conclusion that the carrier of the electromagnetic force, the photon, is massless. Equation 1.8 is the complete Lagrangian of Quantum ElectroDynamics (QED).

The above discussion highlights the analogy between classical and quantum mechanical electrodynamics: As in the classical case the motion of a particle with charge $-e$ can be described by substituting $p^\mu \to p^\mu + eA^\mu$, it turns out that quantum mechanical electrodynamics can be introduced by the corresponding substitution $i\partial^\mu \to i\partial^\mu + eA^\mu$.

1.3 The Strong Interaction

The strong interaction can be constructed in a way similar to the electromagnetic interaction. Differences arise from the fact that the electric charge is substituted by the color charge. There are three possible color combinations, introduced to explain the existence of some baryons, e.g. the Δ^{++}, that

would otherwise contain quarks in identical quantum states, in violation of the Pauli exclusion principle. Due to this fact, the strong interaction is based on an SU(3) local gauge transformation:

$$\psi \rightarrow e^{iT_b a_b(x)}\psi, \tag{1.9}$$

where T^b are eight linearly independent, traceless matrices, called generators of the SU(3) group. The T^b satisfy the following commutation relation:

$$[T_a, T_b] = if_{abc}T_c \tag{1.10}$$

where f_{abc} are the "structure constants" of SU(3). Einstein's convention, according to which common subscripts and superscripts imply summation, is used here and the remainder of this thesis.

Starting from the Dirac Lagrangian $\mathcal{L} = \bar{\psi}_j(i\gamma^\mu\partial_\mu - m)\psi_j$, modified to take into account the three different color charges with use of the j index ($j = 1, 2, 3$), the requirement of invariance under SU(3) gauge transformations leads to:

$$\mathcal{L} = \bar{\psi}_j(i\gamma^\mu\partial_\mu - m)\psi_j + g(\bar{\psi}_j\gamma^\mu T_a\psi_j)G^a_\mu - \frac{1}{4}G^a_{\mu\nu}G^{\mu\nu}_a, \tag{1.11}$$

where g is the coupling constant, G^a_μ is the gluon field, introduced in analogy with the A_μ field in electromagnetism, to ensure the gauge interaction 1.10. The gluon tensor now also contains a new term: $G^a_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu - gf_{abc}G^b_\mu G^c_\nu$. There are eight gluon fields, on for each generator, so $a = 1, 2, ..., 8$.

As in the case of QED, mass terms for the gluon field are forbidden by gauge invariance, meaning that gluons are massless. As a result, Eq. 1.11 is the Lagrangian of Quantum ChromoDynamics (QCD). The QCD Lagrangian is a mixture of terms analogous to QED, e.g. "$\bar{\psi}\partial^\mu\psi$" and "$\bar{\psi}\psi$" (fermion propagation and mass), "$\partial_\mu G^a_\nu \partial^\mu G^\nu_a$" (boson propagation) and "$g\bar{\psi}\gamma^\mu T_a\psi G^a_\mu$" (fermion-boson interaction), and new terms which come up due to the more complicated algebra of the SU(3) group. These extra terms are in the form "$gf^{abc}\partial_\mu G^a_\nu G^\mu_b G^\nu_c$" (three boson vertex) and "$g^2 f_{abc}f^{ade}G^b_\mu G^c_\nu G^\mu_d G^\nu_e$" (four boson vertex), implying the self-coupling of gluons. These terms are possible in the case of QCD but not QED, because gluons carry color charge, whereas photons are electrically neutral.

1.4 The Electroweak Interaction

At the beginning of the 20th century, the existence of β-decays, the radioactive decays in which a nuclide transforms into one of its isobars with emission of a β-particle (energetic electron or positron), posed evidence of a p $\leftrightarrow$ n transformation. The p $\rightarrow$ n transformation is not kinematically allowed outside of the nucleus, however the n $\rightarrow$ p is, implying an interaction that is neither electromagnetic nor strong in nature. The neutron decay, n $\rightarrow$ p$e^-\bar{\nu}_e$, has a mean lifetime of $\sim$15 min, much longer compared to the typical lifetimes of the strong ($\sim 10^{-23}$ s) and electromagnetic interactions ($\sim 10^{-16}$ s), hinting at an interaction that is weaker than the other two. As a result, the interaction responsible for the neutron decay, as well as for a multitude of other decays, came to be known as the "weak interaction". When particles can also decay strongly or electromagnetically, the weak interaction is "hidden", since it is much "slower".

Around the same time period, β-particles from β-decays were observed to have a continuous energy spectrum. This suggested the presence of an additional particle in the decay, which could not be detected, because it has a very small mass and does not carry a strong or electric charge, hence participating only in the new weak interaction. This new particle was named neutrino (ν) by E. Fermi.

In the example above, the carrier of the weak interaction must be a charged particle, called $W^{\pm}$ boson, depending on its charge. Therefore, such interaction is said to include a "weak charged currents" (CC). Experiment shows that only left-handed leptons (right-handed antileptons) can be involved in the weak CC interaction.

Neutrino-electron scattering, $\bar{\nu}_\mu e^- \rightarrow \bar{\nu}_\mu e^-$, was the first observation of the existence of weak neutral currents (NC), i.e. weak interactions which include a neutral mediator, called Z boson. In contrast with the weak CC, the weak NC also involves right-handed charged leptons and quarks.

Experiments suggest that the weak CC and NC have incredibly similar structure. From a theoretical point of view, they could be described by an $SU(2)_L$ structure, where the L subscript implies coupling the weak bosons only to left-handed fermions. The weak CC interactions are associated to the invariance under $SU(2)_L$ local gauge transformations:

$$\psi \rightarrow e^{ig\mathbf{T}\cdot\mathbf{a}(x)}\psi \tag{1.12}$$

where g is the weak interaction coupling and $\mathbf{T}$ are the three generators of $\mathrm{SU}(2)_L$, which can be expressed in terms of the Pauli matrices as $\mathbf{T} = \frac{1}{2}\sigma$.

In this case, the W$\pm$ bosons can be associated to linear combinations of the $W_\mu^{(1)}$ and $W_\mu^{(2)}$ gauge fields of the $\mathrm{SU}(2)_L$ local gauge symmetry, i.e. $W_\mu^\pm = \left(W_\mu^{(1)} \mp W_\mu^{(2)}\right)/\sqrt{2}$, while the Z boson can correspond to the $W_\mu^{(3)}$ gauge field. However, the involvement of the right-handed component in the NC complicates the description. In the electroweak model by Glashow, Salam and Weinberg (GSW), this problem is solved by writing the physical states of the Z boson and the photon as linear combinations of the $W_\mu^{(3)}$ and B_μ fields:

$$\begin{aligned} Z_\mu &= -B_\mu \sin\theta_W + W_\mu^{(3)} \cos\theta_W \\ A_\mu &= +B_\mu \cos\theta_W + W_\mu^{(3)} \sin\theta_W \end{aligned} \tag{1.13}$$

where θ_W is the weak mixing angle, also called the "Weinberg angle", and B_μ is the gauge field of a new $\mathrm{U}(1)_Y$ local gauge symmetry, which replaces the U(1) local symmetry of electromagnetism. This gauge symmetry is invariant under the transformation:

$$\psi \to e^{ig'\frac{Y}{2}a(x)}\psi, \tag{1.14}$$

where g' is the hypercharge coupling and Y is a new kind of charge, called *weak hypercharge*. The resulting interaction term of this gauge symmetry is identical to the one of QED (second term in Eq. 1.8) with the substitutions $e \to g'\frac{Y}{2}$ and $A_\mu \to B_\mu$.

With this mixing of the neutral fields, the $\mathrm{SU}(2)_L$ structure can be saved: The electromagnetic current, j_μ^{em}, and the weak NC, j_μ^{NC}, can be split into two components, one of which transforms according to $\mathrm{SU}(2)_L$ structure, $j_\mu^{(3)}$, while the other is invariant under the SU(2) transformations, j_μ^Y. From Eq. 1.13 follows:

$$\begin{aligned} \frac{j_\mu^{\mathrm{NC}}}{g_{\mathrm{NC}}} &= -\frac{g'}{2} j_\mu^Y \sin\theta_W + g j_\mu^{(3)} \cos\theta_W \\ \frac{j_\mu^{em}}{e} &= +\frac{g'}{2} j_\mu^Y \cos\theta_W + g j_\mu^{(3)} \sin\theta_W \end{aligned} \tag{1.15}$$

where g_{NC} and e are the couplings involved in weak NC and electromagnetic interactions respectively.

As a result, in the GSW model, the electromagnetic interaction is incorporated with the weak interaction, by expanding the underlying symmetry group to $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$. This is the first step towards unifying the two interactions into one interaction, called electroweak (EW) interaction.

Left-handed fermions are included in $\mathrm{SU}(2)_L$ doublets, denoted as L,

$$\begin{pmatrix} \nu_\mathrm{e} \\ e \end{pmatrix}_L, \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}_L, \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}_L, \begin{pmatrix} \mathrm{u} \\ \mathrm{d}' \end{pmatrix}_L, \begin{pmatrix} \mathrm{c} \\ \mathrm{s}' \end{pmatrix}_L, \begin{pmatrix} \mathrm{t} \\ \mathrm{b}' \end{pmatrix}_L \tag{1.16}$$

with weak isospin $T^{(3)} = \pm 1/2$, while right-handed fermions are $\mathrm{SU}(2)_L$ singlets, denoted as R,

$$e_R,\ \mu_R,\ \tau_R,\ u_R,\ d_R,\ c_R,\ s_R,\ t_R,\ b_R$$

with weak isospin $T^{(3)} = 0$. The primes in Eq. 1.16 denote linear combinations of flavor eigenstates that allow the observed flavor-changing, charge-current interactions, e.g. $K^+ \to \mu^+ \nu_\mu$, which involves an effective u-s coupling. In effect, the physical s′ is a linear combination of the mass eigenstates of d, s and b quarks.

For invariance under both the $\mathrm{SU}(2)_L$ and the $\mathrm{U}(1)_Y$ local gauge transformations, particles in the same $\mathrm{SU}(2)_L$ doublet must have the same value of Y. At the same time, Eq. 1.15 suggests that the charge Q can be written as a linear combination of the hypercharge Y and third component of the weak isospin:

$$Q = \alpha T^{(3)} + \beta Y \tag{1.17}$$

For the specific case of ν_{eL} and e_L, it is required that $Y_{\nu_{eL}} = Y_{e_L}$. Combining this with the expression of Eq. 1.17 and the Q and $T^{(3)}$ values for ν_{eL} ($Q_{\nu_{eL}} = 0$, $T^{(3)}_{\nu_{eL}} = +1/2$) and e_L ($Q_{e_L} = -1$, $T^{(3)}_{e_L} = -1/2$), the parameters α and β can be determined. The solution yields:

$$Q = T^{(3)} + \frac{Y}{2} \tag{1.18}$$

Eq. 1.18 implies that:

$$j^{em}_\mu = j^{(3)}_\mu + \frac{1}{2} j^Y_\mu \tag{1.19}$$

Comparing Eq. 1.15 with Eq. 1.19, a relation between the electromagnetic coupling, e, the weak coupling, g, and the hypercharge coupling, g', can be extracted:

$$e = g \sin\theta_W = g' \cos\theta_W \tag{1.20}$$

Taking the definitions above into account and by inputting the charge Q of each fermion in Eq. 1.18, the hypercharge charge Y can be computed, as shown in Table 1.2.

Fermions	Q	$T^{(3)}$	Y
$\nu_{eL},\ \nu_{\mu L},\ \nu_{\tau L}$	0	$+1/2$	-1
$e_L,\ \mu_L,\ \tau_L$	-1	$-1/2$	-1
$e_R,\ \mu_R,\ \tau_R$	-1	0	-2
$u_L,\ c_L,\ t_L$	$+2/3$	$+1/2$	$+1/3$
$u_R,\ c_R,\ t_R$	$+2/3$	0	$+4/3$
$d'_L,\ s'_L,\ b'_L$	$-1/3$	$-1/2$	$+1/3$
$d_R,\ s_R,\ b_R$	$-1/3$	0	$-2/3$

Table 1.2: The charge Q, weak isospin $T^{(3)}$ and hypercharge Y of fermions.

By analogy to the construction of QED, the EW sector Lagrangian is written as

$$\begin{aligned}\mathcal{L} =& \bar{L}\gamma^\mu \left(i\partial_\mu - g\mathbf{T}\cdot\mathbf{W}_\mu - g'\frac{Y}{2}B_\mu \right) L \\ &+ \bar{R}\gamma^\mu \left(i\partial_\mu - g'\frac{Y}{2}B_\mu \right) R - \frac{1}{4}\mathbf{W}_{\mu\nu}\cdot\mathbf{W}^{\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu}\end{aligned} \tag{1.21}$$

where $\mathbf{W}_\mu$ is the vector of the $\mathrm{SU}(2)_L$ gauge fields ($W^i_\mu,\ i = 1,2,3$) and B_μ is the single $\mathrm{U}(1)_Y$ gauge field. $\mathbf{T}$ is the vector of the SU(2) generators. Two separate coupling constants, g and g', are included.

As previously, the gauge invariance of the Lagrangian in Eq. 1.21 does not allow mass terms for any of its vector fields. This is contradicting the experimental observation that the weak interaction has limited range, implying that W and Z bosons are massive. The Brout–Englert–Higgs mechanism was developed to reconcile the invariance to gauge symmetries, which is important for the renormalization of the theory, with the existence of massive

vector bosons. The mechanism consists in the addition to the gauge-invariant Lagrangian of an extra field, whose state of lowest energy breaks the symmetry. The Lagrangian that needs to be added to Eq. 1.21 is given by

$$\mathcal{L} = \left| \left(i\partial_\mu - g\mathbf{T}\cdot\mathbf{W}_\mu - g'\frac{Y}{2}B_\mu \right)\phi \right|^2 - \mu^2\phi^\dagger\phi - \lambda\left(\phi^\dagger\phi\right)^2, \tag{1.22}$$

where $|...|^2 = (...)^\dagger(...)$. In this case, ϕ is an SU(2) doublet of complex scalar fields

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \sqrt{\frac{1}{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix}$$

It is important to note that the Lagrangian 1.22 is invariant under the SU(2) local gauge transformation $\phi \to e^{ig'T_b a_b(x)}\phi$.

If $\mu^2 < 0$ and $\lambda > 0$, then the last two terms of Lagrangian 1.22 produce a "mexican-hat" potential, $V(\phi)$, i.e. a potential with a local maximum on the complex space at $(0,0)$ and a local minimum at $|\phi_0| = v = \sqrt{\frac{-\mu^2}{\lambda}}$. By choosing the minimum at $\phi_1 = \phi_2 = \phi_4 = 0$ and $\phi_3 = v$, so that $\phi = \sqrt{\frac{1}{2}}\begin{pmatrix} 0 \\ v \end{pmatrix}$, the new field has the appropriate quantum numbers ($T^{(3)} = -\frac{1}{2}, Y = 1$) to break both $\mathrm{SU}(2)_L$ and $\mathrm{U}(1)_Y$, while leaving $\mathrm{U}(1)_{em}$ invariant ($Q = 0$). This choice will lead to a massless photon and massive weak bosons. Expanding the first term of the Lagrangian 1.22 with the minimum expectation value chosen above yields the following terms:

$$\begin{aligned} &\left(\frac{1}{2}vg\right)^2 W^+W^- + \frac{1}{8}v^2\left(g^2\left(W_\mu^{(3)}\right)^2 - 2gg'W_\mu^{(3)}B^\mu + g'^2B_\mu^2\right) = \\ &= \left(\frac{1}{2}vg\right)^2 W^+W^- + \\ &\quad + \frac{1}{8}v^2(g^2+g'^2)\left(\frac{gW_\mu^{(3)} - g'B_\mu}{\sqrt{g^2+g'^2}}\right)^2 + \\ &\quad + 0\left(\frac{g'W_\mu^{(3)} + gB_\mu}{\sqrt{g^2+g'^2}}\right)^2 \end{aligned} \tag{1.23}$$

The first term is the mass term of the W boson, $W_\mu^\pm = (W_\mu^{(1)} \mp iW_\mu^{(2)}/\sqrt{2})$ correspond to the vector bosons responsible for the weak CC. The second

term is the mass term of the Z boson, responsible for the weak NC. Finally, the fourth term was added to show the existence of another field, which is massless, and corresponds to the photon. With these definitions, the Z boson field and the photon field are given by the expressions:

$$Z_\mu = \frac{gW_\mu^{(3)} - g'B_\mu}{\sqrt{g^2 + g'^2}},\ A_\mu = \frac{g'W_\mu^{(3)} + gB_\mu}{\sqrt{g^2 + g'^2}} \tag{1.24}$$

By combining Eq. 1.13 and Eq. 1.24, the g and g' couplings can be related with the help of θ_W:

$$\sin\theta_W = \frac{g'}{\sqrt{g^2 + g'^2}},\ \cos\theta_W = \frac{g}{\sqrt{g^2 + g'^2}},\ \tan\theta_W = \frac{g}{g'} \tag{1.25}$$

From 1.23, the mass of the W and Z bosons are:

$$M_{\mathrm{W}} = \frac{1}{2}vg,\ M_{\mathrm{Z}} = \frac{1}{2}v\sqrt{g^2 + g'^2} \tag{1.26}$$

Combining Eqs. 1.25 and 1.26 results in a relation between the W and Z boson masses: $M_W = M_Z \cos\theta_W$. It is also worth noting that putting together the Lagrangian 1.21 with the expression of the photon field 1.24 to recover the correct term for the electromagnetic current, j_μ^{em}, gives rise to the SM relation $g\sin\theta_W = g'\cos\theta_W = e$, which shows that g and g' are not independent.

Using a first order approximation for ϕ around the minimum of the potential,

$$\phi = \sqrt{\frac{1}{2}}\begin{pmatrix} 0 \\ v + h(x) \end{pmatrix}, \tag{1.27}$$

the second and third terms of Eq. 1.22 give:

$$-\lambda v^2 h^2 - \lambda v h^3 - \frac{1}{4}\lambda h^4 \tag{1.28}$$

The first term of Eq. 1.28 is the mass of the Higgs boson, $m_H = \sqrt{2\lambda v^2}$, while the second and third terms describe cubic and quartic Higgs self-interaction. The mass of the Higgs boson is not predicted by the SM, since v which is the only free dimensionful parameter in the SM Lagrangian, is left to be measured by experiments.

As mentioned previously, fermionic mass terms of the form:

$$-m_\psi\bar{\psi}\psi = -m_\psi\bar{\psi}\left(\frac{1}{2}(1-\gamma_5)+\frac{1}{2}(1+\gamma_5)\right)\psi = -m_\psi(\bar{\psi}_R\psi_L+\bar{\psi}_L\psi_R)$$

are forbidden, since they mix the left-handed doublets, ψ_L, with the right-handed singlets, ψ_R, therefore breaking the gauge invariance. However, the same Higgs field that was used for the masses of the vector bosons can give masses to the fermions. The Lagrangian

$$\mathcal{L} = -Y_i\left(\bar{L}_i\phi R_i + \bar{R}_i\phi^\dagger L_i\right) \tag{1.29}$$

is invariant under the $\mathrm{SU}(2)_L\times\mathrm{U}(1)_Y$ gauge transformations. This is achieved by introducing gauge-invariant terms that include the newly introduced ϕ field, which has exactly the required gauge group quantum numbers to couple L and R. The i indices denote the different flavors.

In the case of leptons, Eq. 1.29 successfully generates their masses. This can be shown by substituting the first order approximation of ϕ, 1.27, in the Lagrangian, e.g. for the electron. The Lagrangian 1.29 then becomes:

$$\begin{aligned}\mathcal{L} =& -Y_e\left(\begin{pmatrix}\bar{\nu}_e & \bar{e}\end{pmatrix}_L\begin{pmatrix}\phi^+\\ \phi^0\end{pmatrix}e_R + \bar{e}_R\begin{pmatrix}\phi^- & \bar{\phi}^0\end{pmatrix}\begin{pmatrix}\nu_e\\ e\end{pmatrix}_L\right) = \\ =& -\frac{Y_e}{\sqrt{2}}v(\bar{e}_Re_L+\bar{e}_Le_R) - \frac{Y_e}{\sqrt{2}}(\bar{e}_Re_L+\bar{e}_Le_R)h = \\ =& -\frac{Y_ev}{\sqrt{2}}\bar{e}e - \frac{Y_e}{\sqrt{2}}\bar{e}eh\end{aligned} \tag{1.30}$$

Starting from a gauge invariant Lagrangian, the lepton mass terms (first term of Eq. 1.30) have been constructed with $m_i = \frac{Y_iv}{\sqrt{2}}$. The interaction of the leptons with the Higgs boson is also predicted (second term in Eq. 1.30). The parameters Y_i are arbitrary; hence the actual lepton masses are left to be determined by the experiment.

The Lagrangian described in Eq. 1.29 generated masses only for the lower part of the $\mathrm{SU}(2)_L\times\mathrm{U}(1)_Y$ doublets, i.e. the charged leptons. This works for leptons, since the neutrinos are massless in the SM. A slightly modified Lagrangian is needed for the quark doublets, in which both elements in an $\mathrm{SU}(2))L$ doublet have to acquire mass in order to provide masses to both charge $+2/3$ quarks ("up-like" quarks) and charge $-1/3$ quarks ("down-like"

quarks). This is achieved by constructing a new Higgs doublet from ϕ in the following manner:

$$\phi_c = -i\tau_2\phi^* = \begin{pmatrix} -\bar{\phi}^0 \\ \phi^- \end{pmatrix} \xrightarrow[\text{breaking}]{\text{Symmetry}} \sqrt{\frac{1}{2}} \begin{pmatrix} v + h(x) \\ 0 \end{pmatrix} \tag{1.31}$$

The new Higgs doublet, ϕ_c, has opposite weak hypercharge to ϕ but transforms identically to it, therefore it can be used to modify Eq. 1.29 to generate the quark masses. By substituting 1.31 in 1.29, the latter, in the case of, e.g. u and d quarks, becomes:

$$\begin{aligned} \mathcal{L} =& - Y_d \bar{L}\phi R_d - Y_u \bar{L}\phi_c R_u + \text{h.c.} = \\ =& - Y_d \begin{pmatrix} \bar{u} & \bar{d} \end{pmatrix}_L \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} d_R - Y_u \begin{pmatrix} \bar{u} & \bar{d} \end{pmatrix}_L \begin{pmatrix} -\bar{\phi}^0 \\ \phi^- \end{pmatrix} u_R + \text{h.c.} = \\ =& - \frac{Y_d v}{\sqrt{2}} \bar{d}d - \frac{Y_u v}{\sqrt{2}} \bar{u}u - \frac{Y_d}{\sqrt{2}} \bar{d}dh - \frac{Y_u}{\sqrt{2}} \bar{u}uh \end{aligned} \tag{1.32}$$

As in the case of leptons, the first two terms are the masses of the quarks, with $m_i = \frac{Y_i v}{\sqrt{2}}$, while the third and fourth terms are the interaction of the quarks with the Higgs boson. The parameters Y_i are again arbitrary and cannot be predicted by theory. An interesting property resulting from the above discussion is that the interaction coupling of the Higgs boson to the fermions is, in all cases, proportional to the fermion masses.

The above description of the EW Lagrangian is called the *Weinberg–Salam* model.

1.5 Beyond the Standard Model

The above discussion shows that the building principles of the SM of Particle Physics can be summarized in a few points:

- **Quantum Field Theory (QFT)**, which describes the quantum mechanics of relativistic quantum fields.
- The **local gauge invariance principle**, which leads to interactions between matter particles (fermions) via gauge fields (bosons).

- The **Brout–Englert–Higgs mechanism of the electroweak symmetry**, which leads to the generation of masses for particles, which are normally forbidden by the invariance under the gauge transformations $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$.

- The **experimental data**, which guided the construction of the theory.

Putting everything together, the SM has managed to accurately describe known phenomena and predict new ones. The success of the SM is illustrated in Fig. 1.2, which shows a comparison between the SM predictions and experimental data for a multitude of processes involving the gauge bosons and top quarks, covering a cross section range of almost nine orders of magnitude.

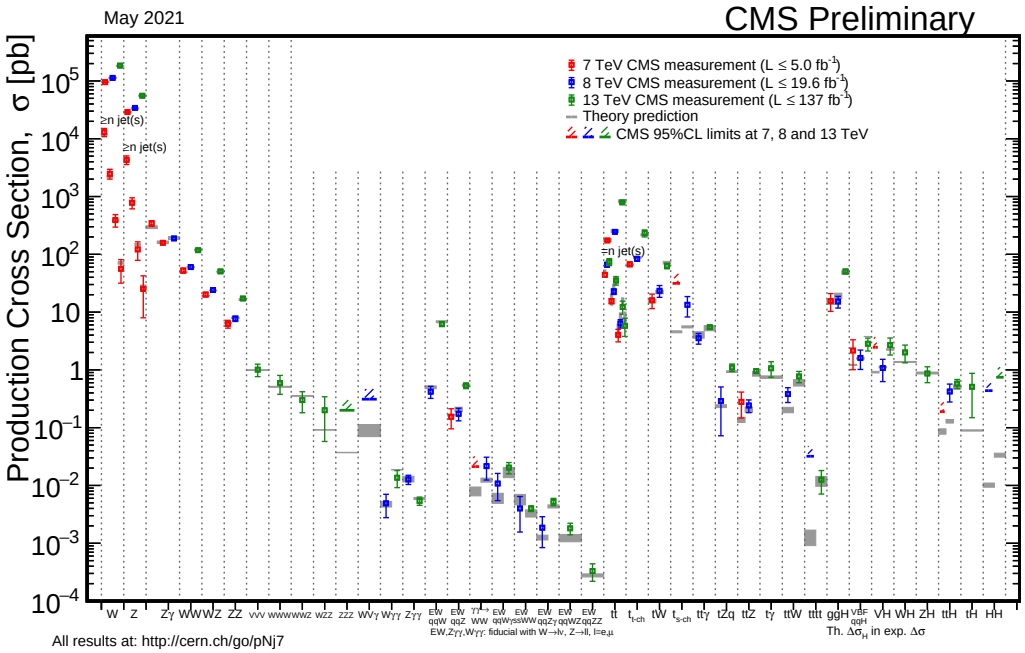


Figure 1.2: Summary of the cross section measurements of Standard Model processes by the CMS Collaboration [2].

Despite its impressive success, there are still fundamental questions that the SM leaves unanswered. These questions, that call for a theory beyond the SM, are outlined below:

- **The nature of Dark Matter**
  In the early 1930s, there were already indications that the luminous

mass of the universe was not enough to account for the observed gravitational pull in galaxies and galaxy clusters. Astronomer J. H. Oort found the velocities of stars in galaxies to be large enough for them to escape the gravitational pull of the luminous mass. Astronomer F. Zwicky applied the virial theorem on the Coma galaxy cluster, assuming only gravitational Newtonian interaction to find that the mass calculated in this way was some 50 times larger than the mass seen through light observations. Forty years later, V. Rubin et al. discovered that the velocity distribution of stars inside galaxies was not consistent with the observed luminous mass distribution, leading to the hypothesis of a significant non-luminous mass component, dubbed *dark matter* (DM). More recent estimations of galaxy cluster masses through the observation of gravitational lensing also corroborate the existence of DM. Finally, the measurement of fluctuations of the cosmic microwave background by the Wilkinson Microwave Anisotropy Probe has led to the conclusion that the total energy-matter density of the universe consists of $\sim$5% by baryonic matter, $\sim$23% by DM and $\sim$72% by *dark energy*, which is presumably the cause of the accelerating expansion of the universe. The estimated amounts of DM due to normal baryonic matter, either due to microscopic (e.g. neutrinos) or macroscopic objects (e.g. brown dwarf stars), cannot account for these observations. As a result, a new type of weakly interacting massive particle (WIMP) is needed. This is not predicted by the SM but can occur naturally in several extensions of it [3].

- **Quantum gravity**
  The SM has managed to successfully describe three out the four fundamental interactions of Nature. However, the gravitational interaction cannot be incorporated in the theory and new theoretical approaches need to be applied, so that gravity becomes consistent with the quantum mechanical description of Nature.

- **Grand unification**
  The coupling "constants" of the three SM fundamental forces change with changing energy. The variation is such that, at very high energies, the constants tend to converge to the same value: The electromagnetic coupling constant increases due to fermion loops, while the strong and weak coupling constants decrease due to the gluon and

W/Z self-interactions that lead to boson loops. This has led to conjectures that the three forces are unified into a single force by some "Grand Unification Theory" (GUT) at some very large energy scale, the so-called GUT scale. The convergence of the coupling constants, when taking into account only the SM particles, is not exact. This "imperfection" can be mitigated by theories beyond the SM that contain extra particles, modifying the running of the coupling constants. The simplest and earliest attempts at grand unification focused on enlarging the SM gauge group to, e.g., SU(5), which has SU(3), $SU(2)_L$ and $U(1)_Y$ as subgroups. While the first results on SU(5) were very promising, the non-observation of proton decay, predicted by SU(5), led to other extensions of the SM.

- **The naturalness problem**
  The SM is a renormalizable theory, i.e. the higher order quantum corrections can be absorbed into a finite set of physical variables of the theory in such a way that all observable quantities have only logarithmic dependence on the energy scale. In fact, since New Physics is expected to become dominant at very high energies, e.g. at the Planck scale, Λ_P, when the gravitational interaction also becomes important, the calculations of the SM should stay finite up to this very large, arbitrary scale Λ. In the case of fermion masses, the higher order corrections give terms that are proportional to $\ln\Lambda$, i.e. they are rather stable to even large changes of Λ. The exception to this comes from the higher order corrections to the mass of the Higgs boson, which are proportional to Λ^2. Taking into account that $\Lambda^2 \sim \Lambda_P^2 \sim 10^{38}\,\mathrm{GeV}^2$, whereas $m_H^2 \sim 10^4\,\mathrm{GeV}^2$, the cancellation of these higher order terms need to span over 30 orders of magnitude. This "unnatural" cancellation that, in the framework of the SM, can only be explained by coincidence, is called the *naturalness problem*. Theories beyond the SM provide more "natural" solutions to the scale of the Higgs boson mass.

- **The nature and properties of neutrinos**
  In the SM, there are no right-handed neutrinos and, thus, neutrinos are massless. However, experimental results of astronomic origin [4] discovered that neutrinos oscillations exist, i.e. the neutrinos can change flavor. The phenomenon has since been observed and measured accu-

rately in ground-based experiments at nuclear reactors and accelerators. This change of flavor is only possible if neutrinos have non-zero masses, which calls for the modification of the SM. Apart from that, the question of whether the neutrinos are Majorana particles, i.e. they are their own antiparticles, is yet unanswered: The SM does not predict their exact nature and the experimental attempts to distinguish neutrinos and antineutrinos have proven extremely difficult, with the most promising experiment being neutrinoless β-decay, which is allowed only if neutrinos are Majorana particles.

- **Matter-antimatter asymmetry**
  The observed universe is dominated by matter. However, there is no apparent reason why matter should dominate over antimatter; in a universe where the CP operator, where C is the charge conjugation operator and P is the parity operator, is a complete symmetry of Nature, equal amounts of matter and antimatter are expected and their annihilation would lead to a photon dominated universe. Given the observed dominance of matter over antimatter in the universe, CP symmetry must be violated. There are only elements in the SM construct that can lead to a CP violation: The CKM matrix, which describes the flavor mixing of quarks, and the PMNS matrix, which describes the mixing of neutrinos. The violation of CP in the CKM matrix has been extensively studied, whereas CP violation in the PMNS matrix will hopefully be probed in a new generation of neutrino experiments. However, even if the violation in the latter case is maximal, the total CP violation in the SM is still insufficient to account for the observed matter-antimatter asymmetry; a solution to this problem requires extra sources of CP violation, outside the framework of the SM.

- **Large number of arbitrary parameters**
  The SM (including a minimal extension that takes into account the neutrino masses) contains 26 parameters that cannot be predicted by the theory but are rather determined by observations. These parameters are:

  – The 12 fermion masses:

$$m_{\nu_1},\ m_{\nu_2},\ m_{\nu_3},\ m_e,\ m_\mu,\ m_\tau,\ m_d,\ m_s,\ m_b,\ m_u,\ m_c,\ m_t$$

– The 3 coupling constants:

$$\alpha,\ G_F,\ \alpha_S$$

– The 2 Higgs potential parameters:

$$v,\ m_H$$

– The 8 PMNS and CKM matrices mixing angles:

$$\theta_{12},\ \theta_{13},\ \theta_{23},\ \delta \text{ and } \lambda,\ A,\ \rho,\ \eta$$

– The phase of the QCD Lagrangian that could lead to CP violation in the strong interaction:

$$\theta_{CP} \simeq 0$$

The theoretical prediction of these parameters could only be given by some more fundamental theory beyond the SM, if such a theory exists.

Chapter 2

Supersymmetry

This chapter is dedicated to elaborating on one of the most attractive beyond the SM (BSM) theory, Supersymmetry (SUSY). The presentation of the theory is loosely based on two introductory overviews of SUSY, [5] and [6].

SUSY is a proposed symmetry of nature relating fermions and bosons. Supersymmetric transformations can be generated by an operator Q so that:

$$Q\,|\text{Fermion}\rangle = |\text{Boson}\rangle \qquad Q\,|\text{Boson}\rangle = |\text{Fermion}\rangle$$

As a consequence of the above relation, the operator Q carries half-integer spin, therefore it is a fermionic operator, i.e. it must be an anticommuting spinor. Its hermitian conjugate, $Q^\dagger$, is also a supersymmetry operator. These operators must fulfil the following (anti)commutation relations:

$$\begin{aligned} \{\, Q, Q^\dagger \,\} &= P^\mu \\ \{\, Q, Q \,\} = \{\, Q^\dagger, Q^\dagger \,\} &= 0 \\ [P^\mu, Q] = [P^\mu, Q^\dagger] &= 0 \end{aligned}$$

Irreducible representations of the supersymmetric algebra are called *supermultiplets*. Each supermultiplet contains both bosons and fermions which are connected, up to a spacetime translation/rotation, with the relation $f(Q, Q^\dagger)\,|\text{Fermion}\rangle = |\text{Boson}\rangle$, where f is a function of the supersymmetry generators. The Q and $Q^\dagger$ operators commute with the P^2 and gauge transformation operators, which means that particles in the same supermultiplet, also called *superpartners*, have the same mass and electric, weak and color

charges. It can be shown that each supermultiplet contains the same number of bosonic and fermionic degrees of freedom.

There are two kinds of supermultiplets that are of special interest due to the particle content of the SM: Chiral supermultiplets are the supermultiplets whose fermionic state is one Weyl fermion, which has two helicity states, hence two fermionic degrees of freedom. As a result, chiral supermultiplets contain two scalar degrees of freedom, combined into a complex scalar field. Gauge supermultiplets are the supermultiplets that contain one spin-1 vector boson. Gauge invariance forbids a massive vector boson, therefore the vector boson has to be massless (at least until the gauge symmetry is broken). A massless vector boson has two helicity states, hence two bosonic degrees of freedom. The appropriate fermionic state to balance these degrees of freedom is a massless spin-1/2 Weyl fermion, called *gaugino*. From the fact that the adjoint representation of a gauge group is its own conjugate, it follows that fermions in the gauge supermultiplets have the same gauge transformations for their left- and right-handed components. Particles in the supermultiplets defined above, which are useful for building extensions of the SM, contain superpartners with spin difference equal to 1/2.

It is worth noting here that it is convenient and also conventional that the fermions of chiral supermultiplets are defined only by their left-handed components. This is possible due to the definition of the charge conjugation operator. For a Weyl fermion

$$\psi = \begin{pmatrix} \zeta \\ \xi \end{pmatrix},$$

where ζ and ξ are the right- and left-handed components, respectively, given the definition of the γ matrices in the Dirac–Pauli representation, it is true that:

$$\begin{aligned} \psi^{(\bar{f})} &\equiv i\gamma^2 \left(\psi^{(f)}\right)^* \Rightarrow \\ \xi^{(\bar{f})} &\equiv i\sigma^2 \left(\xi^{(f)}\right)^* \end{aligned}$$

As a result, a Weyl fermion can be expressed only in terms left-handed components of itself and its antiparticle:

$$\psi^{(f)} = \begin{pmatrix} i\sigma^2 \left(\xi^{(\bar{f})}\right)^* \\ \xi^{(f)} \end{pmatrix}$$

2.1 Motivation for Supersymmetry

- **The naturalness problem**
  At one-loop level, the source of the Λ divergences mentioned in Sec. 1.5 come from Dirac fermion loops, which introduce corrections of the form:

$$\Delta m_H^2 = -\frac{|\lambda_f|^2}{8\pi^2}\Lambda^2 + \dots \tag{2.1}$$

  On the other hand, corrections that come from a complex scalar particles have the form:

$$\Delta m_H^2 = \frac{\lambda_S}{16\pi^2}\Lambda^2 + \dots \tag{2.2}$$

  The ellipses represent terms, proportional to the squared particle mass, that grow at most logarithmically with Λ. It is evident that, if $|\lambda_f|^2 = \lambda_S$, the leading terms of Eq. 2.1 and Eq. 2.2 can cancel out each other due to their opposite sign. There is an additional factor 2 missing from the scalar term, implying that two scalar fields are required to cancel a single fermionic contribution. This is exactly the correspondence among boson and fermion superpartners described in the introduction of Sec. 2. It can also be proven that this cancellation is achieved at any higher-order correction. As a result, a SUSY extension of the SM naturally provides the mechanism to explain why the mass of the Higgs boson can be at the electroweak scale without relying to unnatural levels of fine-tuning.
  The absence, experimentally, of a spin-0 particle with the mass and charges of the electron clearly shows that, for a realistic SUSY model, in which the supersymmetric partners have masses different than their SM counterparts, the cancellation of terms cannot be exact. However, as long as the characteristic scale of these masses is not much higher than the TeV scale, SUSY provides an elegant solution to the naturalness problem.

- **The nature of Dark Matter**
  A SUSY feature with important phenomenological consequences is the definition of R parity. Even though they are forbidden in the SM, baryon number (B) and lepton number (L) violations are perfectly allowed in SUSY. Such violations would conflict with observations, since

they would lead to the decay of the proton, which has been found to have a half-life longer than $10^{34}\ yrs$. Nevertheless, B and L conservation cannot be imposed explicitly, since both numbers are known to be violated in non-pertubative electroweak effects, negligible at ordinary energies but possibly important at much higher energies. A proposed solution to implicitly conserve both numbers is to introduce a new symmetry, R parity, defined as

$$R = (-1)^{3B+L+2s},$$

where s is the particle spin. It can be shown that R equals to $+1$ for SM particles, while it is equal to -1 for their superpartners. If it is multiplicatively conserved, R parity implies the conservation B and L and has important phenomenological consequences. The most important consequence is that the lightest supersymmetric particle (LSP) is stable and, if it is electrically neutral, provides an attractive dark matter candidate, hence providing a potential solution to one of the greatest unanswered question of the SM. Other consequences include the production of SUSY particles only in pairs in SM particle interactions, and thus in collider experiments, and the decay of SUSY particles only into odd numbers of LSPs.
It is worth noting that a small violation of R parity can be accommodated within the present bounds of observations and have interesting phenomenological effects. However, this possibility will not be considered in the following.

- **Grand unification**
  A SUSY extension of the SM modifies the running coupling constants of the theory in a way such that they convincingly converge to a common value at very high energies. Given the proposals of GUTs, this unification of the gauge couplings provides another encouraging motivation for SUSY.

2.2 Supersymmetric Lagrangians

This section outlines the possible Lagrangians of a general SUSY theory. Based on these general expressions, assumptions can be made that lead to useful cases of extensions to the SM.

Starting from the case of a chiral supermultiplet, this includes a left-handed Weyl fermion ξ and a complex scalar ϕ. The simplest Lagrangian containing only the kinematic terms of both fields is:

$$\mathcal{L}_{\text{chiral,free}} = \mathcal{L}_{\text{scalar}} + \mathcal{L}_{\text{fermion}} = \partial^\mu \phi^* \partial_\mu \phi + i\xi^\dagger \bar{\sigma}^\mu \partial_\mu \xi, \tag{2.3}$$

where $\bar{\sigma}^0 = \sigma^0 = \boldsymbol{I}$ and $\bar{\sigma}^i = -\sigma^i$ for $i = 1, 2, 3$, with σ^i being the 2×2 Pauli matrices.

Using the simplest form for the supersymmetric transformation of the scalar field,

$$\delta\phi = \epsilon\xi, \tag{2.4}$$

the proper transformation of the supersymmetric part of the fermionic field can be inferred. The transformation, needed to leave the Lagrangian 2.3 invariant is, up to a multiplicative factor:

$$\delta\xi_\alpha = -i(\sigma^\mu \epsilon^\dagger)_\alpha \partial_\mu \phi \tag{2.5}$$

a is the spinor index.

For the Lagrangian 2.3 to be supersymmetric, apart from its invariance to supersymmetric transformations, it needs to be shown that the supersymmetric algebra closes, i.e. that that the commutator of transformations parametrized by two different spinors ϵ_1 and ϵ_2 is also a symmetry of the theory. In the case of the scalar field, this relation is automatically fulfilled:

$$(\delta_{\epsilon_1}\delta_{\epsilon_2} - \delta_{\epsilon_2}\delta_{\epsilon_1})\phi = (\epsilon_1 \sigma^\mu \epsilon_2^\dagger - \epsilon_2 \sigma^\mu \epsilon_1^\dagger)(i\partial_\mu)\phi, \tag{2.6}$$

where $i\partial_\mu = P_\mu$ is the generator of spacetime translations.

In the case of the fermionic field, the supersymmetric algebra closes only on-shell. This is a consequence of the fact that, off-shell, the fermionic degrees of freedom do not match the bosonic ones: Off-shell, the fermionic field can be complex, hence having four degrees of freedom. To account for this, an *auxiliary*, complex scalar field F needs to be introduced with proper transformation to keep the total Lagrangian invariant but also make the algebra close even off-shell:

$$\delta F = i\epsilon^\dagger \bar{\sigma}^\mu \partial_\mu \xi \tag{2.7}$$

The Lagrangian term for this auxiliary field that needs to be added to the Lagrangian 2.3 is

$$\mathcal{L}_{\text{chiral,aux}} = F^* F, \tag{2.8}$$

which implies that there is no kinetic term for this field (since there is no derivative $\partial_\mu F$) and that its equation of motion is simply $F = F^* = 0$. The two extra scalar degrees of freedom vanish on-shell.

The most general form of a Lagrangian containing renormalizable interactions of the chiral superpartners can include only terms with mass dimension less than four:

$$\mathcal{L}_{\text{chiral,int}} = \left(-\frac{1}{2} W^{ij} \xi_i \xi_j + W^i F_i + x^{ij} F_i F_j \right) + \text{c.c.} + U, \tag{2.9}$$

where W^{ij}, W^i, x^{ij} and U are degree 1-, 2-, 0- and 4-degree polynomials in ϕ_i and ϕ^{*i}. However, Eqs. 2.4, 2.5, 2.7 imply that there can be no supersymmetric transformation to cancel out terms involving the x^{ij} and U terms. As a result, the Lagrangian 2.9 can be simplified to

$$\mathcal{L}_{\text{chiral,int}} = \left(-\frac{1}{2} W^{ij} \xi_i \xi_j + W^i F_i \right) + \text{c.c.} \tag{2.10}$$

The variation of this Lagrangian (up to terms linear in F_i that trivially cancel out) leads to:

$$\begin{aligned} \delta\mathcal{L}_{\text{chiral,int}} = &\left(-\frac{1}{2} \frac{\delta W^{ij}}{\delta \phi_k} (\epsilon \xi_k)(\xi_i \xi_j) - \frac{1}{2} \frac{\delta W^{ij}}{\delta \phi^{*k}} (\epsilon^\dagger \xi^{\dagger k})(\xi_i \xi_j) \right) + \\ &+ \left(i W^{ij} \partial_\mu \phi_j \xi_i \sigma^\mu \epsilon^\dagger + i W^i \partial_\mu \xi_i \sigma^\mu \epsilon^\dagger \right) + \text{c.c.} \end{aligned} \tag{2.11}$$

The Fierz identity[1] states that the first term can be identically zero if and only if $\frac{\delta W^{ij}}{\delta \phi_k}$ is totally symmetric under the interchange of i, j, k. However, no such identity exists for the second term, hence it must be that $\frac{\delta W^{ij}}{\delta \phi^{*k}} = 0$. This function means that W^{ij} cannot contain conjugates of ϕ^{*k}, i.e. it is a holomorphic of the complex scalar field ϕ_k. The combination of the above conclusions results in the expression

$$W^{ij} = M^{ij} + y^{ijk} \phi_k = \frac{\delta^2 W}{\delta \phi_i \delta \phi_j}, \text{ with} \tag{2.12}$$

[1]The Fierz identity [7] is an identity that allows rewriting bilinear forms of the product of two spinors as a linear combination of the products of bilinear forms of the individual spinors. The same identity can be applied to quadrilinear forms, as in this case [8].

$$W = L^i\phi_i + \frac{1}{2}M^{ij}\phi_i\phi_j + \frac{1}{6}y^{ijk}\phi_i\phi_j\phi_k, \tag{2.13}$$

where W is called the *superpotential* and L^i are parameters that only affect the scalar fields that are a gauge singlet, M^{ij} is the symmetric mass matrix for the fermion fields and y^{ijk} are the totally symmetric under the interchange of i, j, k Yukawa couplings between a scalar field ϕ_k and two fermionic fields ξ_i, ξ_j. Due to Eq. 2.12, the first two terms of the second line of Eq. 2.11 can be written as a total derivative if

$$W^i = \frac{\delta W}{\delta \phi_i} = L^i + M^{ij}\phi_j + \frac{1}{2}y^{ijk}\phi_j\phi_k \tag{2.14}$$

The total Lagrangian of a chiral supermultiplet is the sum of Eqs. 2.3, 2.8 and 2.9. However, gathering all the terms that contain the auxiliary field and writing its equation of motion, it is found that

$$F_i = -W^{*i} \tag{2.15}$$

Taking this also into account, the total chiral supermultiplet Lagrangian is:

$$\begin{aligned}\mathcal{L}_{\text{chiral}} =& \mathcal{L}_{\text{chiral,free}} + \mathcal{L}_{\text{chiral,int}} + \mathcal{L}_{\text{chiral,aux}} = \\ =& -\partial^\mu\phi^*\partial_\mu\phi + i\xi^\dagger\bar{\sigma}^\mu\partial_\mu\xi - \frac{1}{2}\left(W^{ij}\xi_i\xi_j + W^*_{ij}\xi^{\dagger i}\xi^{\dagger j}\right) - W^iW^*_i\end{aligned} \tag{2.16}$$

By substituting the expressions of W^{ij} and W^i in Eq. 2.16 and writing out the linearized equations of motions, it can be shown that the scalar and fermionic field satisfy the same wave equation, having exactly the same mass term, as stated in the introduction of this chapter.

Moving on to the gauge supermultiplets, these contain a massless gauge boson field A^a_μ and a two-component gaugino, whose left-handed field is λ^a. The index a runs over the adjoint representation of the gauge group, e.g. $a = 1, \ldots, 8$ for $\mathrm{SU}(3)_C$. In the general case, a free gauge supermultiplet Lagrangian would contain terms similar to that of the SM QCD (Sec. 1.3). This works on-shell, since both the gauge boson and the gaugino have two helicity states, hence two degrees of freedom each. However, in the off-shell case, the gauge boson acquires a third degree of freedom, related to its longitudinal polarization. At the same time, the gaugino field becomes

complex, gaining two more degrees of freedom. As a result, an extra bosonic degree of freedom needs to be added in the Lagrangian, similar to the chiral Lagrangian case. In analogy with introduction of the F field, the D field is introduced in this case, without kinetic term, so that it can be eliminated on-shell by its equation of motion. With these considerations, the Lagrangian can be written:

$$\mathcal{L}_{\text{gauge,free}} = -\frac{1}{4}F^a_{\mu\nu}F^{\mu\nu a} + i\lambda^{\dagger a}\bar{\sigma}^\mu\nabla_\mu\lambda^a + \frac{1}{2}D^aD^a, \text{ with} \tag{2.17}$$

$$F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + gf^{abc}A^b_\mu A^c_\nu \tag{2.18}$$

$$\nabla_\mu\lambda^a = \partial_\mu\lambda^a + gf^{abc}A^b_\mu\lambda^c, \tag{2.19}$$

where g is the gauge coupling and f^{abc} are the totally antisymmetric structure constants of the gauge group. $f^{abc} = 0$ in the case of an Abelian group.

Using the supersymmetry transformations

$$\delta A^a_\mu = -\frac{1}{\sqrt{2}}\left(\epsilon^\dagger\bar{\sigma}_\mu\lambda^a + \lambda^{\dagger a}\bar{\sigma}_\mu\epsilon\right), \tag{2.20}$$

$$\delta\lambda^a_\alpha = \frac{i}{2\sqrt{2}}(\sigma^\mu\bar{\sigma}^\nu\epsilon)_\alpha F^a_{\mu\nu} + \frac{1}{\sqrt{2}}\epsilon_\alpha D^a, \tag{2.21}$$

$$\delta D^a = \frac{i}{\sqrt{2}}\left(-\epsilon^\dagger\bar{\sigma}^\mu\nabla_\mu\lambda^a + \nabla_\mu\lambda^{\dagger a}\bar{\sigma}^\mu\epsilon\right), \tag{2.22}$$

it can be shown that the Lagrangian 2.17 is invariant and that, for each field $X = F^a_{\mu\nu},\ \lambda^a,\ \lambda^{\dagger a},\ D^a$, the supersymmetric algebra closes:

$$(\delta_{\epsilon_1}\delta_{\epsilon_2} - \delta_{\epsilon_2}\delta_{\epsilon_1})X = (\epsilon_1\sigma^\mu\epsilon_2^\dagger - \epsilon_2\sigma^\mu\epsilon_1^\dagger)(-i\nabla_\mu)X \tag{2.23}$$

The final step towards the full supersymmetric Lagrangian is the inclusion of gauge interactions. By considering only the terms with mass dimension less than four, which maintain the renormalizability of the Lagrangian, the gauge-interactions part of the Lagrangian is:

$$\mathcal{L}_{\text{gauge,int}} = -\sqrt{2}g(\phi^*T^a\xi)\lambda^a - \sqrt{2}g\lambda^{\dagger a}(\xi^\dagger T^a\phi) + g(\phi^*T^a\phi)D^a \tag{2.24}$$

With these terms, the full, general chiral + gauge supersymmetric Lagrangian is:

$$\mathcal{L} = \mathcal{L}_{\text{chiral}} + \mathcal{L}_{\text{gauge,free}} + \mathcal{L}_{\text{gauge,int}} \tag{2.25}$$

However, to satisfy invariance with respect to supersymmetric transformations, the supersymmetric transformation of the auxiliary field F needs to be slightly modified:

$$\delta F_i = -i\epsilon^\dagger \bar{\sigma}^\mu \partial_\mu \xi_i + \sqrt{2} g (T^a \phi_i) \epsilon^\dagger \lambda^{\dagger a} \tag{2.26}$$

In addition to that, as in the case of the SM, gauge invariance demands that the regular derivatives in Eq. 2.16 be replaced by the covariant derivatives:

$$\nabla_\mu \phi_i = \partial_\mu \phi_i - i g A^a_\mu (T^a \phi)_i, \tag{2.27}$$

$$\nabla_\mu \xi_i = \partial_\mu \xi_i - i g A^a_\mu (T^a \xi)_i, \tag{2.28}$$

where the hermitian matrices T^a are the gauge group generators satisfying $[T^a T^b] = i f^{abc} T^c$.

With the complete Lagrangian written out, the equation of motion for the auxiliary field field D^a gives:

$$D^a = -g(\phi^* T^a \phi) \tag{2.29}$$

Following this result, it is shown that the auxiliary fields F_i and D^a can be expressed as functions of the scalar fields. Thus, the complete scalar potential is:

$$\begin{aligned} V(\phi, \phi^*) =& F^{*i} F_i + \frac{1}{2} \sum_a D^a D^a = W_i^* W^i + \frac{1}{2} \sum_a g_a^2 (\phi^* T^a \phi)^2 = \\ =& M^*_{ik} M^{kj} \phi^{*i} \phi_j + \frac{1}{2} M^{in} y^*_{jkn} \phi_i \phi^{*j} \phi^{*k} + \frac{1}{2} M^*_{in} y^{jkn} \phi^{*i} \phi_j \phi_k + \\ &+ \frac{1}{4} y^{ijn} y^*_{kln} \phi_i \phi_j \phi^{*k} \phi^{*l} + \ldots + \frac{1}{2} \sum_a g_a^2 (\phi^* T^a \phi)^2, \end{aligned} \tag{2.30}$$

where the ellipses imply terms containing L^i, omitted here for simplicity, since they do not prove useful in what follows.

A unique property of SUSY theories is that the scalar potential is completely defined by the other interactions in the theory. More specifically, the first term of Eq. 2.30, also called *F-term*, is fixed by the Yukawa couplings and fermion mass terms, whereas the second term, also called *D-term*, is fixed by the gauge interactions.

As mentioned previously, supersymmetry must be broken in any realistic SUSY model. This comes from the fact that in unbroken SUSY the masses of superparticles and SM particles must be the same. If this were the case, it then seems impossible that all these particles have been missed by all experiments to date, e.g. the scalar superpartners of the electron that should have a mass of 0.511 MeV. As a result, it is expected that SUSY is an exact symmetry that is spontaneously broken, similar to electroweak symmetry, which is also broken. This SUSY symmetry breaking should be soft, which means that the soft-breaking Lagrangian must be composed only by terms with positive mass dimension. The general soft-breaking Lagrangian has the form:

$$\begin{aligned}\mathcal{L}_{\text{soft}} = &-\left(\frac{1}{2}M_a\lambda^a\lambda^a + \frac{1}{6}a^{ijk}\phi_i\phi_j\phi_k + \frac{1}{2}b^{ij}\phi_i\phi_j + t^i\phi_i\right) + \text{c.c.} + \\ &- (m^2)^i_j\phi^{j*}\phi_i + \\ &- \frac{1}{2}c_i^{jk}\phi^{*i}\phi_j\phi_k + \text{c.c.} + \\ &- M^a_{\text{Dirac}}\lambda^a\xi_a + \text{c.c.}\end{aligned} \tag{2.31}$$

The above terms break supersymmetry since they involve only scalars and gauginos without their corresponding superpartners. In this way, they describe masses for the SUSY particles, even if their SM superpartners are massless or relatively light. It can be proven that the first two lines of Eq. 2.31 contain terms that are free of quadratic divergences to corrections to the scalar masses, at all orders of perturbation theory. They consist of gaugino masses M_a, scalar squared-mass terms $(m^2)^i_j$, cubed (a^{ijk}), squared (b^{ij}) and "tadpole" (t^i) scalar couplings. However, the t^i are applicable only in cases where ϕ_i is a gauge singlet. The term in the third line is free of quadratic divergences only if the supermultiplets do not contain gauge singlets but it is generally neglected, since it is difficult to build model with small c_i^{jk} factors. Finally, the term in the last line only applies to theories in which the chiral supermultiplets are singlets.

2.3 The Minimal Supersymmetric Standard Model

The general conclusions regarding the structure of a complete SUSY Lagrangian, drawn in the previous section, are now specified for the case of the Minimal Supersymmetric Standard Model (MSSM). The MSSM is the extension of the SM that contains the smallest set of new particles and interactions that are consistent with phenomenology. Because of this definition, the gauge groups and transformations of the MSSM are already known; they are the same as the SM ones. Half of the particle content is also already known, since it consists of all SM particles. The other half of the particle content emerges from properly populating the chiral and gauge supermultiplets with the SM particles and identifying the missing elements.

Due to the nature of the electroweak interaction, which separates the left- from the right-handed components of fermions, and given the definition of supermultiplets (see the introduction of this chapter), it is clear that fermions must be included in chiral supermultiplets. As a consequence, each fermion component has a corresponding spin-0 superpartner. Since these new particles are scalar, they are denoted by prefixing the letter "s" is prefixed to the name of their SM superpartner, e.g. *sfermions*, *squarks*, *sleptons*, *selectron*, etc. Superpartners are denoted by the SM particle name with a tilde above it. In the specific case of sfermions, a subscript is also needed to indicate the to which fermion component (L or R) the sparticle corresponds to. For example, the scalar superpartner to the left-handed electron component is denoted by $\widetilde{e}_L$, while the scalar superpartner to the right-handed electron component is denoted by $\widetilde{e}_R$. This distinction is important, since sfermions carry over the coupling properties from their SM counterparts; as an example, the $\widetilde{e}_L$ couples to the W boson, while the $\widetilde{e}_R$ does not.

The SM Higgs boson, being a scalar particle, must be included in a chiral supermultiplet. However, there is a complication that is introduced when the SM is extended to respect supersymmetry. The way that the quarks acquire masses in the SM is described by Eq. 1.32: The d-type quarks acquire their masses by means of the scalar field ϕ, while the u-type quarks acquire their masses by means of the scalar field ϕ_c, which is constructed with the help of the complex conjugate of ϕ. This is not possible in SUSY models, since the superpotential, which includes the Yukawa couplings, cannot involve complex conjugates of the fields. As a result, two separate Higgs bosons, H_d and H_u,

are needed so that both types of quarks can acquire masses. The first Higgs supermultiplet includes H_u, with weak isospin components $\begin{pmatrix} H_u^+, & H_u^0 \end{pmatrix}$, and its fermionic superpartner

$$\widetilde{H}_u = \begin{pmatrix} \widetilde{H}_u^+ \\ \widetilde{H}_u^0 \end{pmatrix}.$$

Similarly, the other Higgs supermultiplet includes H_d, with weak isospin components $\begin{pmatrix} H_d^0, & H_d^- \end{pmatrix}$, and its fermionic superpartner

$$\widetilde{H}_d = \begin{pmatrix} \widetilde{H}_d^0 \\ \widetilde{H}_d^- \end{pmatrix}$$

Collectively, the superpartners of the Higgs fields are called *higgsinos*.

The SM vector bosons, being spin-1 particles, are contained in gauge supermultiplets. The superpartner of the gluon is the *gluino*, $\widetilde{g}$, while the superpartners of the electroweak bosons are the *winos*, $\widetilde{W}^\pm$ and $\widetilde{W}^0$, and the *bino*, $\widetilde{B}^0$. Table 2.1 summarizes the particle content of the MSSM, categorized in chiral and gauge supermultiplets, as outlined above.

The next missing piece for defining the MSSM is the specification of the superpotential. The starting point is Eq. 2.13: The first term can exist only in theories that contain a scalar field that is a gauge singlet, which is not the case for the MSSM, hence $L^i = 0$. The second term gives mass to scalar bosons, H_u and H_d in the case of the MSSM, while the third term contains the Yukawa couplings and is constructed in such a way so as to give masses to the MSSM particles similarly to the SM. With these considerations, the MSSM superpotential is:

$$W_{\text{MSSM}} = \mu H_u H_d + \bar{u}\boldsymbol{y_u} Q H_u - d\boldsymbol{y_d} Q H_d - \bar{e}\boldsymbol{y_e} L H_d, \tag{2.32}$$

where the indices have been suppressed for simplicity.

A first observation is that, due to the holomorphism of the superpotential, Higgs boson mass terms of the form $H_u^* H_u$ or $H_d^* H_d$ are forbidden. Because of this property of the superpotential, Eq. 2.32 illustrates that the two Higgs bosons are needed in order to give masses to all types of quarks: H_u for the "up-like" quarks and H_d for the "down-like" quarks. The terms in Eq. 2.32 are constructed in accordance with this requirement.

A second observation concerns interactions in the MSSM. It was shown in Sec. 2.2 that the y^{ijk} are totally symmetric under the interchange of i, j,

Chiral supermultiplets

Name & Symbol		Spin-0	Spin-1/2	$SU(3)_C$, $SU(2)_L$, $U(1)_Y$
squarks, quarks (3 families)	Q	$(\widetilde{u}_L \quad \widetilde{d}_L)$	$(u_L \quad d_L)$	$(3,\ 2,\ \frac{1}{6})$
	$\bar{u}$	$\widetilde{u}_R^*$	$u_R^\dagger$	$(\bar{3},\ 2,\ -\frac{2}{3})$
	$\bar{d}$	$\widetilde{d}_R^*$	$d_R^\dagger$	$(\bar{3},\ 2,\ \frac{1}{3})$
sleptons, leptons (3 families)	L	$(\widetilde{\nu}_e \quad \widetilde{e}_L)$	$(\nu_e \quad e_L)$	$(1,\ 2,\ -\frac{1}{2})$
	$\bar{e}$	$\widetilde{e}_R^*$	$e_R^\dagger$	$(1,\ 1,\ 1)$
Higgs, higgsinos	H_u	$(H_u^+ \quad H_u^0)$	$(\widetilde{H}_u^+ \quad \widetilde{H}_u^0)$	$(1,\ 2,\ \frac{1}{2})$
	H_d	$(H_d^0 \quad H_d^-)$	$(\widetilde{H}_d^0 \quad \widetilde{H}_d^-)$	$(1,\ 2,\ -\frac{1}{2})$

Gauge supermultiplets

Name & Symbol	Spin-1/2	Spin-1	$SU(3)_C$, $SU(2)_L$, $U(1)_Y$
gluino, gluon	$\widetilde{g}$	g	$(8,\ 1,\ 0)$
winos, W bosons	$\widetilde{W}^\pm \quad \widetilde{W}^0$	$W^\pm \quad W^0$	$(1,\ 3,\ 0)$
bino, B boson	$\widetilde{B}^0$	B^0	$(1,\ 1,\ 0)$

Table 2.1: Chiral and gauge supermultiplets of the MSSM.

k. This implies not only q-q-H and ℓ-ℓ-H interactions but also q-$\widetilde{q}$-$\widetilde{H}$ and ℓ-$\widetilde{\ell}$-$\widetilde{H}$ interactions. It also implies that these interactions are all of the same strength, e.g. y_t in the case of the t and its superpartners. That stays true (up to small radiative corrections) even after soft supersymmetry breaking.

The first term of 2.32 forms the supersymmetric Higgs mass terms, which are non-negative with a minimum at $H_u = H_d = 0$. As a result, these terms cannot accommodate the electroweak symmetry breaking, for which the soft supersymmetry breaking terms are needed. The rest of the terms couple left- and right-handed squarks and sleptons to the Higgs bosons with coupling strength proportional to μy. These terms prove to be important for the mixing of the squarks and sleptons. By using the superpotential 2.32 for the construction of the scalar potential 2.30, more interactions can be identified: The fourth term of Eq. 2.30 indicates the existence of the quartic interactions, of the form $\widetilde{q}$-$\widetilde{q}$-$\widetilde{q}$-$\widetilde{q}$, $\widetilde{q}$-$\widetilde{q}$-$\widetilde{\ell}$-$\widetilde{\ell}$ and $\widetilde{q}$-$\widetilde{q}$-H-H (and the same for $q \to \ell$), all of the same strength, y^2.

The Yukawa couplings are usually small when compared to the gauge

couplings, which tend to be more important for the phenomenology of the theory. Such terms of gauge-coupling strength are the ones presented in Eq. 2.24. The first two imply gaugino couplings to q-$\widetilde{q}$, ℓ-$\widetilde{\ell}$, H-$\widetilde{H}$ pairs. The last term shows the existence of more scalar quartic terms, which, in this case, have strength proportional to the gauge couplings g.

The MSSM soft supersymmetry breaking terms can be easily written out given Eq. 2.31. By taking into account that the MSSM does not contain gauge singlets and that the c_i^{jk} terms are usually neglected, the MSSM supersymmetry soft part of the Lagrangian is:

$$\begin{aligned}\mathcal{L}_{\text{soft,MSSM}} = &-\frac{1}{2}\left(M_1\widetilde{B}\widetilde{B} + M_2\widetilde{W}\widetilde{W} + M_3\widetilde{g}\widetilde{g} + \text{c.c.}\right) + \\ &-\left(\widetilde{\bar{u}}\boldsymbol{a_u}\widetilde{Q}H_u - \widetilde{\bar{d}}\boldsymbol{a_d}\widetilde{Q}H_d - \widetilde{\bar{e}}\boldsymbol{a_e}\widetilde{L}H_d + \text{c.c.}\right) + \\ &-\widetilde{Q}^\dagger\boldsymbol{m_Q^2}\widetilde{Q} - \widetilde{L}^\dagger\boldsymbol{m_L^2}\widetilde{L} - \widetilde{\bar{u}}\boldsymbol{m_{\bar{u}}^2}\widetilde{\bar{u}}^\dagger - \widetilde{\bar{d}}\boldsymbol{m_{\bar{d}}^2}\widetilde{\bar{d}}^\dagger - \widetilde{\bar{e}}\boldsymbol{m_{\bar{e}}^2}\widetilde{\bar{e}}^\dagger + \\ &- m_{H_u}^2 H_u^* H_u - m_{H_d}^2 H_d^* H_d - (bH_uH_d + \text{c.c.}),\end{aligned} \tag{2.33}$$

where the gauge indices have been suppressed.

The first line of Eq. 2.33 includes the bino (M_1), wino (M_2) and gluino (M_3) mass terms. The second line consists of cubed couplings of scalar fields, while the third line includes squark and slepton mass terms. Finally, the last line contains the soft supersymmetry breaking terms to the Higgs potential. The scale of the parameters involved is of the order of m_{soft}, which should not be much higher than $\sim 1\,\text{TeV}$, if SUSY is to provide a convincing solution to the naturalness problem.

The number of parameters introduced for the MSSM soft supersymmetry breaking terms is very large: 105 parameters, not observed in the SM, are introduced. However, many of these terms imply flavor mixing and CP violation processes that are strongly constrained by experiments. These restrictions lead to simplifications of the parameters of Eq. 2.33:

$$\boldsymbol{m_Q^2} = m_Q^2\mathbf{1},\ \boldsymbol{m_{\bar{u}}^2} = m_{\bar{u}}^2\mathbf{1},\ \boldsymbol{m_{\bar{d}}^2} = m_{\bar{d}}^2\mathbf{1},\ \boldsymbol{m_L^2} = m_L^2\mathbf{1},\ \boldsymbol{m_{\bar{e}}^2} = m_{\bar{e}}^2\mathbf{1}, \tag{2.34}$$

$$\boldsymbol{a_u} = A_{u0}\boldsymbol{y_u},\ \boldsymbol{a_d} = A_{d0}\boldsymbol{y_d},\ \boldsymbol{a_e} = A_{e0}\boldsymbol{y_e}, \tag{2.35}$$

$$\Im(M_1) = \Im(M_2) = \Im(M_3) = \Im(A_{u0}) = \Im(A_{d0}) = \Im(A_{e0}) = 0 \tag{2.36}$$

The set of Eqs. 2.34, 2.35, 2.36 are collectively called the *soft supersymmetry breaking universality* conditions. By taking into account reasonable

assumptions driven by experimental results, these conditions bring down the new MSSM parameters from 105 to 15: 3 real gaugino masses, 5 real squark and slepton mass parameters, 3 real scalar cubic parameters and 4 Higgs mass parameters.

Electroweak symmetry breaking in the MSSM is slightly more complicated than that in the SM, due to the existence of two Higgs doublets. Considering the constraints of the theory and after the proper gauge transformations, the MSSM Higgs scalar potential becomes

$$\begin{aligned} V_{\text{Higgs,MSSM}} =&(\mu^2 + m_{H_u}^2)|H_u^0|^2 + (\mu^2 + m_{H_d}^2)|H_d^0|^2 + \\ &- (bH_u^0 H_d^0 + \text{c.c.}) + \frac{1}{8}(g^2 + g'^2)(|H_u^0|^2 + |H_d^0|^2)^2 \end{aligned} \tag{2.37}$$

The requirement that the potential 2.37 be bounded from below leads to the following condition:

$$2b < 2|\mu|^2 + m_{H_u}^2 + m_{H_d}^2 \tag{2.38}$$

Similarly, for the minimum value of the potential 2.37 to be non-zero trivially, it must be that:

$$b^2 > (|\mu|^2 + m_{H_u}^2)(|\mu|^2 + m_{H_d}^2) \tag{2.39}$$

It is worth noting that both of these requirements cannot be fulfilled if $m_{H_u}^2 = m_{H_d}^2$. Eqs. 2.38 and 2.39 are the conditions for the Higgs fields to get non-zero vacuum expectations values (VEVs). By defining

$$v_u = \left\langle H_u^0 \right\rangle, \qquad v_d = \left\langle H_d^0 \right\rangle, \tag{2.40}$$

$$\tan\beta = \frac{v_u}{v_d}, \tag{2.41}$$

where v_u and v_d are taken by convention to be positive and real, and hence $0 < \beta < \pi/2$, the known relation of the VEVs to the known Z boson mass can be written out:

$$v_u^2 + v_d^2 = v^2 = \frac{2m_{\text{Z}}^2}{g^2 + g'^2} \tag{2.42}$$

As mentioned previously, the Higgs fields in the MSSM consist of two independent complex scalar fields, therefore 8 distinct degrees of freedom.

After electroweak symmetry breaking, three of these degrees of freedom become the longitudinal modes of Z and $W^{\pm}$ bosons. As a result, there are 5 remaining Higgs scalar fields: two CP-even neutral ones, h^0 and H^0, where $m_{h^0} < m_{H^0}$, one CP-odd neutral one, A^0 and two charged ones, $H^{\pm}$, where $H^- = H^{+*}$. Provided that the VEVs of Eq. 2.40 minimize the Higgs potential at tree-level, the diagonalization of the Higgs mass matrix in Eq. 2.37 gives the following eigenvalues:

$$\begin{aligned} m^2_{A^0} &= 2|\mu|^2 + m^2_{H_u} + m^2_{H_d} \\ m^2_{h^0,H^0} &= \frac{1}{2}\left(m^2_{A^0} + m^2_{\mathrm{Z}} \mp \sqrt{(m^2_{A^0} - m^2_{\mathrm{Z}})^2 + 4m^2_{\mathrm{Z}} m^2_{A^0} \sin^2(2\beta)} \right) \\ m_{H^{\pm}} &= m^2_{A^0} + m^2_{\mathrm{W}} \end{aligned} \tag{2.43}$$

As a final note regarding electroweak symmetry breaking in the MSSM: it is interesting that the masses of quarks and leptons depend not only on the Yukawa coupling(s) but also on $\tan\beta$, e.g. for the third generation of quarks, $m_{\mathrm{t}} = y_{\mathrm{t}} v \sin\beta$ and $m_{\mathrm{b}} = y_{\mathrm{b}} v \cos\beta$.

Moving on to the masses of the neutral gaugino (neutralinos) and the charged gaugino combinations (charginos) of the MSSM, these are the mass eigenstates of the combination of gauge eigenstates: There are four neutralinos, being the combinations of neutral higgsinos, the bino and the neutral wino, and two charginos, being the combinations of the charged higgsinos and winos. According to convention, neutralinos and charginos are denoted by indices in ascending order of mass, e.g. $m_{\widetilde{N}_1} < m_{\widetilde{N}_2}$.

The masses of the neutralinos are given by the diagonalization of the neutralino mass matrix, which can result in fairly complicated expressions. However, the mass expressions are simplified in the case in which the electroweak symmetry breaking scale is much smaller than the neutralino mass matrix terms, i.e.

$$m_{\mathrm{Z}} \ll |\mu \pm M_1|, |\mu \pm M_2|. \tag{2.44}$$

In this limit, the mass eigenstates consist, to a large extent, by a specific gauge eigenstate. Some interesting cases are worth mentioning:

- $M_1 < M_2 \ll |\mu|$
  $\widetilde{N}_1$ is mostly bino with $m_{\widetilde{N}_1} \approx M_1$, $\widetilde{N}_2$ is mostly wino with $m_{\widetilde{N}_1} \approx M_2$ and $\widetilde{N}_{3,4}$ are mostly higgsino with $m_{\widetilde{N}_{3,4}} \approx |\mu|$

- $M_2 < M_1 \ll |\mu|$
  $\widetilde{N}_1$ is mostly wino with $m_{\widetilde{N}_1} \approx M_2$, $\widetilde{N}_2$ is mostly bino with $m_{\widetilde{N}_1} \approx M_1$ and $\widetilde{N}_{3,4}$ are mostly higgsino with $m_{\widetilde{N}_{3,4}} \approx |\mu|$

- $|\mu| < M_1, M_2$
  $\widetilde{N}_{1,2}$ are mostly higgsino with $m_{\widetilde{N}_{1,2}} \approx |\mu|$, with $m_{\widetilde{N}_2} - m_{\widetilde{N}_1} \sim \frac{m_W^2}{\min(M_1, M_2)}$ and $\widetilde{N}_{3,4}$ are mostly bino, mostly wino or a mix of both, depending on the M_1 and M_2 mass ordering.

The masses of the charginos are more easily obtained, since they involve the diagonalization of a 2×2 mass matrix. Looking into the interesting limit of Eq. 2.44, the chargino masses are $m_{\widetilde{C}_1} \approx M_2$ and $m_{\widetilde{C}_2} \approx |\mu|$ if $M_2 < |\mu|$ (and with the interchange $1 \leftrightarrow 2$ if $|\mu| < M_2$). It is worth noting that, in this limit, charginos are nearly mass degenerate with some of the neutralinos, although the result is not exact.

The gluino is a unique particle in the MSSM, because it is the only one that is a color octet. As a consequence, it cannot mix with any other particle. In most models, a rough relation

$$M_3 : M_2 : M_1 \approx 6 : 2 : 1 \tag{2.45}$$

is predicted near the 1 TeV scale [5]. This implies that the gluino is generally considerably heavier than the neutralinos and charginos.

Regarding the squarks and sleptons, the determination of their masses involves the diagonalization of three 6×6 matrices of up-type quarks, down-type quarks and sleptons and one 3×3 matrix for the sneutrinos. Even though the complexity is alleviated by the flavor-related assumptions that lead to the soft supersymmetry breaking universality conditions, the complete expressions are still hard to interpret. However, some general conclusions can be drawn:

- The first- and second-generation squarks are generally heavier than the sleptons.

- The left-handed squarks of the first and second generation are generally heavier than the right-handed ones. The same holds for the sleptons of the first two generations.

- The lightest charged lepton is probably the stau $\widetilde{\tau}_1$, the lightest of the $\widetilde{\tau}_L$ and $\widetilde{\tau}_R$ combinations.
- The lightest squarks are probably the lightest stop, $\widetilde{t}_1$ and the lightest sbottom, $\widetilde{b}_1$.

It can be noticed in the conclusions listed above that there is an important distinction between the third-generation sfermions and all the other sfermions. This is a result of the large Yukawa couplings and trilinear soft couplings of the third generation. While the first- and second-generation sfermions form heavy, nearly mass-degenerate pairs, the third-generation can have significant mixing and substantially lower masses.

All of the theoretical considerations above provide for a very rich phenomenology within the MSSM framework. In all of the cases presented below, it is assumed that the R parity is conserved and that the LSP is the $\widetilde{N}_1$. As a result, it is expected that any decay chain will lead to a pair of $\widetilde{N}_1$, which escape the experimental apparatuses undetected, due to their weak coupling with SM particles.

Starting with the production of SUSY particles, Feynman diagrams of different possibilities are shown in Figs. 2.1–2.2. More specifically, Fig. 2.1 illustrates the production of squark-antisquark, gluino-squark and gluino pairs, which is of QCD strength. Given that the Large Hadron Collider (LHC) is a proton-proton (p-p) collider, the cross sections of these processes are very large. This makes the (anti)squark and gluino pair production dominant, as shown in Fig. 2.3. Fig. 2.2 illustrates the production of neutralinos, charginos and sleptons from quark-antiquark interactions, all of which are of electroweak strength. As a result, these processes have cross sections more than one order of magnitude (depending on the sparticle mass) lower than the QCD-mediated production modes (Fig. 2.3).

Sparticle decays are largely dictated by the gauge eigenstate content and the handedness of the lightest supersymmetric particles. Multiple decay modes will be presented below, starting from the dominant and going to less kinematically plausible ones. To begin with squarks, the decay $\widetilde{q} \to q\widetilde{g}$ is usually the dominant one, if kinematically allowed, given that the $\widetilde{q}-q-\widetilde{g}$ vertex has QCD strength. The next possibility is the squark decay to neutralinos or charginos. In the case where the $\widetilde{N}_1$ is mostly bino and the $\widetilde{N}_2$ and $\widetilde{C}_1$ are mostly winos, the direct decay $\widetilde{q} \to q\widetilde{N}_1$ is preferred for right-handed squarks and the decays $\widetilde{q} \to q\widetilde{N}_2$ and $\widetilde{q} \to q'\widetilde{C}_1$ are preferred for left-handed

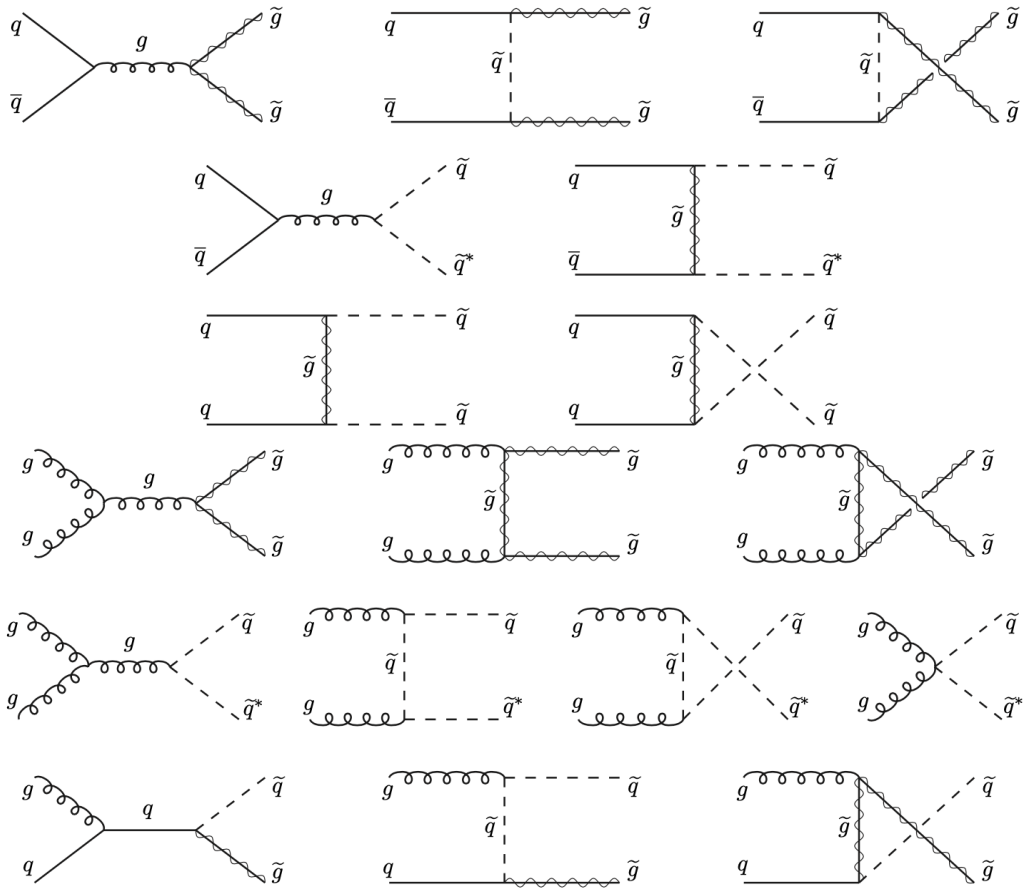


Figure 2.1: Feynman diagrams of production of gluinos and squarks from quark-antiquark annihilation (upper three rows) and from gluon-gluon or gluon-quark fusion (lower three rows) [5].

squarks, based on the relative coupling strengths involved in each case. In the aforementioned decay processes, the $\widetilde{g}$, $\widetilde{N}_2$ and $\widetilde{C}_1$ are expected to decay to more supersymmetric and SM particles (these decays are discussed in the following paragraphs), leading to long and complex chains, called *cascade decays*.

Third-generation squark decays deserve a dedicated discussion. Apart from the fact that they lead to final states with b and t quarks, which generally have more characteristic signatures in experimental detectors, their decays to neutralinos and charginos become important in cases where the latter

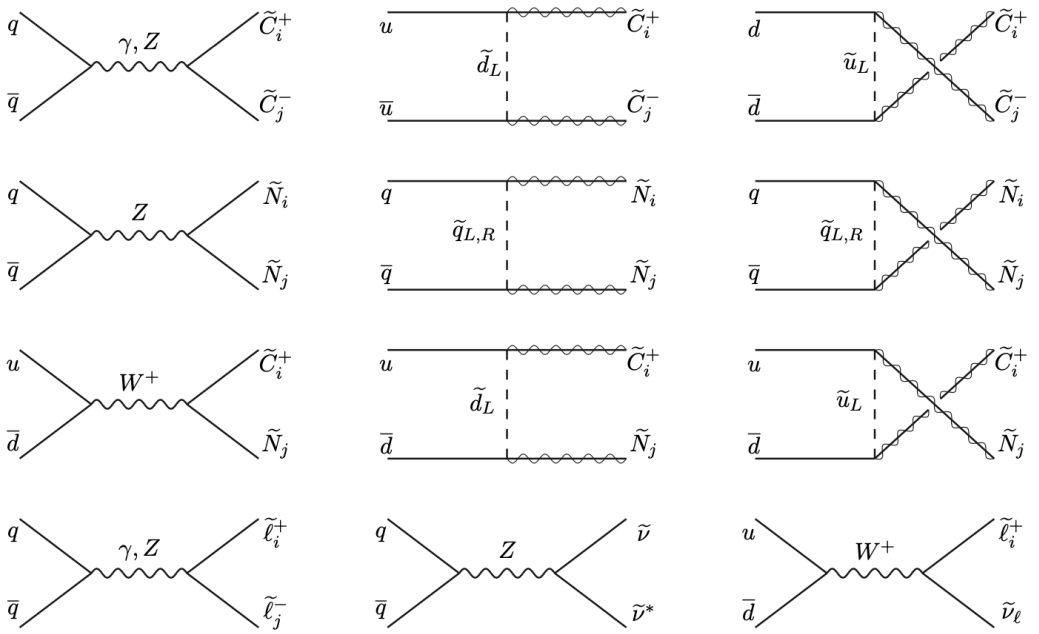


Figure 2.2: Feynman diagrams of production of neutralinos and charginos from quark and antiquark annihilation [5].

contain a large higgsino component, owing to their large Yukawa couplings. Focusing on alternative decays of $\widetilde{t}$, if the $\widetilde{t}_1 \to \mathrm{t}\widetilde{g}$, $\widetilde{t}_1 \to \mathrm{t}\widetilde{N}_2$ and $\widetilde{t}_1 \to \mathrm{b}\widetilde{C}_1$ decay modes are kinematically forbidden, the three-body decay $\widetilde{t}_1 \to \mathrm{bW}\widetilde{N}_1$ becomes the most important one. If $m_{\widetilde{t}_1} < m_\mathrm{b} + m_\mathrm{W} + m_{\widetilde{N}_1}$, even this decay mode is forbidden, leaving only the flavor-suppressed $\widetilde{t}_1 \to \mathrm{c}\widetilde{N}_1$ and the four-body $\widetilde{t}_1 \to \mathrm{bff}'\widetilde{N}_1$ decay options available. It is worth noting that, if $m_{\widetilde{t}_1} - m_{\widetilde{N}_1} \lesssim 20\,\mathrm{GeV}$, the $\widetilde{t}_1$ can become long-lived, i.e. have enough lifetime to decay at a nonnegligible distance from the interaction point of collider physics experiments [10]. However, in all of the following discussions, it is assumed that this is not the case.

As a result of their couplings in the MSSM, gluinos can only decay via on- or off-shell squarks, depending on the mass spectrum of SUSY particles. If $m_{\widetilde{g}} > m_q + m_{\widetilde{q}}$, the decay to on-shell squarks, $\widetilde{g} \to q\widetilde{q}$, is preferred, because of the QCD strength of its vertex. Since third-generation squarks are probably the lightest, the decays to $\widetilde{t}$ or $\widetilde{b}$ are generally dominant. In the case where the above processes are not kinematically accessible, the decay to virtual squarks, i.e. $\widetilde{g} \to qq\widetilde{N}_i$ or $\widetilde{g} \to qq'\widetilde{C}_i$, becomes important. Similarly to

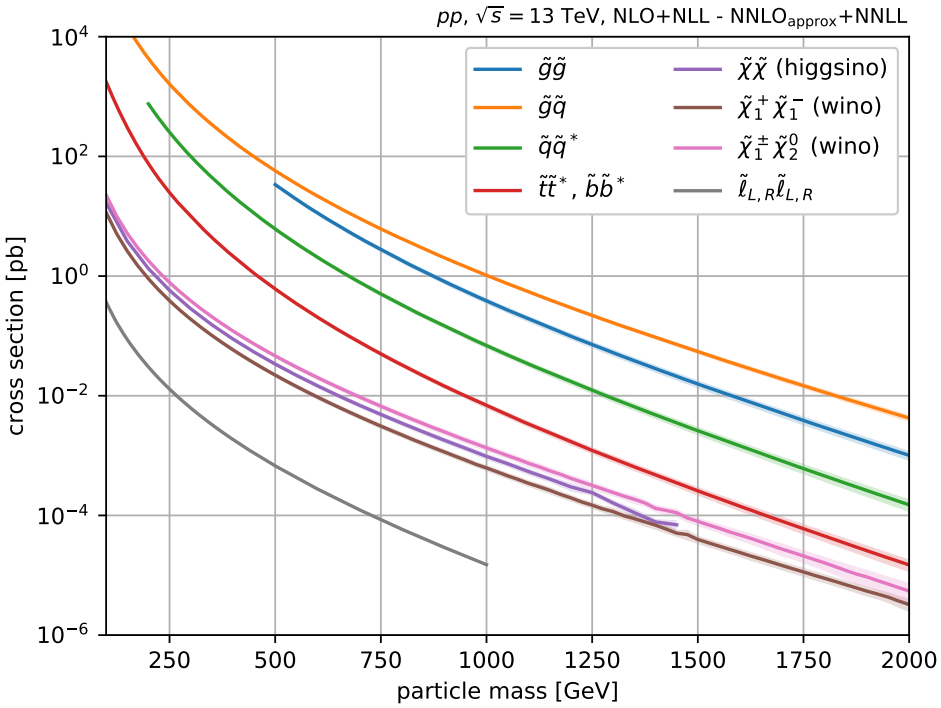


Figure 2.3: Summary of cross section values for the pair production of SUSY particles [9].

squarks, cascade decays are very common for gluinos.

Regarding neutralinos and charginos, their two-body decays are dominant if the mass spectrum allows them:

$$\widetilde{N}_i \to \mathrm{Z}/h^0/H^0/A^0\widetilde{N}_j \ (i>j), \quad \widetilde{N}_i \to \mathrm{W}^\pm/H^\pm\widetilde{C}_j^\mp, \quad \widetilde{N}_i \to \ell\widetilde{\ell}/\nu\widetilde{\nu}/q\widetilde{q},$$
$$\widetilde{C}_i^\pm \to \mathrm{W}^\pm/H^\pm\widetilde{N}_j, \quad \widetilde{C}_2^\pm \to \mathrm{Z}/h^0/H^0/A^0\widetilde{C}_1^\pm, \quad \widetilde{C}_i^\pm \to \ell^\pm\widetilde{\nu}/\nu\widetilde{\ell}^\pm/q\widetilde{q}',$$

where the decays to lighter particles, Z, W, h^0, $\widetilde{\ell}$ and $\widetilde{\nu}$ (given that squarks are generally heavier than sleptons), are more plausible. This picture slightly changes if there is a significant higgsino component in neutralinos and charginos; in that case their decay to third-generation squarks also becomes significant, enhanced by the large top and bottom Yukawa couplings.

If their two-body decay modes are closed, neutralinos and charginos can decay via three-body processes:

$$\widetilde{N}_i \to \mathrm{ff}\widetilde{N}_j \ (i > j), \quad \widetilde{N}_i \to \mathrm{ff}'\widetilde{C}_j,$$
$$\widetilde{C}_i \to \mathrm{ff}'\widetilde{N}_j, \qquad \widetilde{C}_2^{\pm} \to \mathrm{ff}\widetilde{C}_1^{\pm}.$$

The above processes, in the case of $\widetilde{N}_2$ or $\widetilde{C}_1$ decaying to $\widetilde{N}_1$, are exceptionally important for phenomenology. If the fermions involved in these decays are quarks, then this experimentally leads to final states with jets (usually b-jets in the case of intermediate off-shell Higgs boson). In the case where these fermions are leptons, the final state is experimentally very clean, since electrons and muons can be reconstructed more easily compared to jets. However, when the intermediate off-shell sparticles are sleptons, these leptons are usually staus, resulting in tau leptons in the final state, whose identification poses experimental challenges.

Sleptons decay to neutralinos and charginos via the processes:

$$\widetilde{\ell}^{\pm} \to \ell^{\pm}\widetilde{N}_i, \quad \widetilde{\ell}^{\pm} \to \nu\widetilde{C}_i^{\pm},$$
$$\widetilde{\nu} \to \nu\widetilde{N}_i, \qquad \widetilde{\ell} \to \ell^{\pm}\widetilde{C}_i^{\mp},$$

where the values of the index i are subject to constraints imposed by the mass spectrum of SUSY particles. As in the case of squarks, the branching fractions of left- and right-handed slepton decays depend on the admixture of gauginos in $\widetilde{N}_i$ and $\widetilde{C}_i$. For example, if the $\widetilde{N}_1$ is mostly bino and the $\widetilde{N}_2$ and $\widetilde{C}_1$ are mostly wino, then the direct decays to the former are preferred by right-handed sleptons, $\widetilde{\ell}_R^{\pm} \to \ell^{\pm}\widetilde{N}_1$ ($\mathrm{U}(1)_Y$ interaction strength, $\sim g'$), while left-handed sleptons preferably decay to the latter ($\mathrm{SU}(2)_L$ interaction strength, $\sim g$).

Fig. 2.4 shows some examples of neutralino, chargino and gluino decays. The branching fractions for these final states can be found be in [11, 12].

2.4 Overview of Searches for Supersymmetry at the LHC

The MSSM, along with its phenomenological features described in the previous section, has been utilized in detailed and in more simplified models [13–15] to experimentally search for SUSY. This section provides an overview of

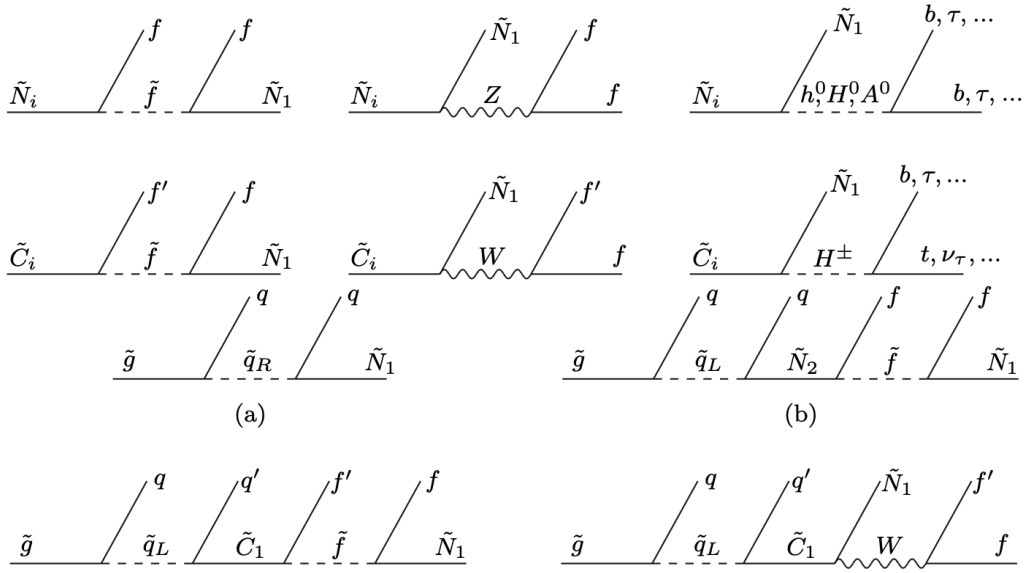


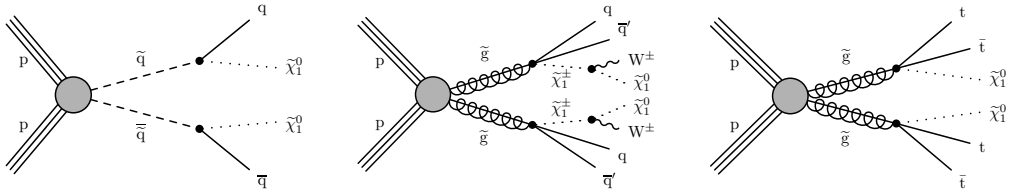
Figure 2.4: Feynman diagrams of decays of neutralinos and charginos (two upper rows) and of gluinos (two lower rows) to the LSP [5].

the most recent searches at the LHC, primarily showing results by the Compact Muon Solenoid (CMS) experiment[2]. Results by the ATLAS experiment are very similar.

The decay of colored sparticles usually leads to final states with moderate (Fig. 2.5a) to high hadronic activity, especially in the case of cascade decays (Fig. 2.5b). The possible existence of top quarks in the decay chains (Fig. 2.5c) results in more characteristic final states which can contain numerous bottom quark jets, which are generally more easily identified due to the long B hadron lifetime. In many cases, there is a significant amount of missing energy due to the presence of the pair of lightest SUSY particle in the final state, which escape detection. The total longitudinal momentum in the case of p-p collisions is not known, because the two momenta of the interacting partons, that actually participate in the interaction, cannot be measured parallel to the beam and are not equal. However, it is known that the sum of the momenta perpendicular to the beam is zero. As a result, any deviation from zero in the sum of the momenta of visible particles in the plane transverse to the beam is assigned to a new variable, the missing

[2]In what follows, the $\widetilde{N}_1$ is denoted as $\widetilde{\chi}_1^0$, the $\widetilde{N}_2$ as $\widetilde{\chi}_2^0$ and the $\widetilde{C}_1$ as $\widetilde{\chi}_1^\pm$.

transverse momentum, denoted $p_{\mathrm{T}}^{\mathrm{miss}}$, which quantifies the total transverse momentum of all invisible particles in an event.



(a) $pp \to \tilde{q}\bar{\tilde{q}}$, $\tilde{q} \to q\tilde{\chi}_1^0$

(b) $pp \to \tilde{g}\tilde{g}$, $\tilde{g} \to q\bar{q}'\tilde{\chi}_1^\pm$, $\tilde{\chi}_1^\pm \to \mathrm{W}^\pm\tilde{\chi}_1^0$

(c) $pp \to \tilde{g}\tilde{g}$, $\tilde{g} \to \mathrm{t}\bar{\mathrm{t}}\tilde{\chi}_1^0$

Figure 2.5: Diagrams involving the production of squarks and gluinos.

The dominant SM background processes in such cases are multijet events, usually originating from QCD processes affected by detector jet energy mismeasurements which cause high amounts of $p_{\mathrm{T}}^{\mathrm{miss}}$, W+jets or $\mathrm{t}\bar{\mathrm{t}}$ production, with $\mathrm{W} \to \ell\nu$, where the charged lepton is missed or absorbed into a jet, and Z+jets, with $\mathrm{Z} \to \nu\bar{\nu}$. Apart from the $p_{\mathrm{T}}^{\mathrm{miss}}$ variable, which proves to be a primary handle to separate the SUSY signal from the SM background, numerous additional variables are utilized in order to increase the sensitivity of searches for strongly produced SUSY particles with hadronic final states. Simple requirements, such a high number of jets (N_{jet}), a high number of jets identified as coming from a b quark ($N_{\text{b-jet}}$), or a large scalar sum of transverse momenta (p_{T}) of the jets in the event, denoted as H_{T}, turn out to be quite effective in separating SM processes from SUSY particle production and decay. With time, more sophisticated variables have been constructed to fully exploit the differences between signal and background processes. Examples of such variables are:

- The effective mass [16], $m_{\mathrm{eff}} = \sum_{j\in\mathrm{jets}} p_{\mathrm{T}}^{\mathrm{j}} + \sum_{\ell\in\mathrm{leptons}} p_{\mathrm{T}}^{\ell} + p_{\mathrm{T}}^{\mathrm{miss}}$, which tends to peak at the mass of SUSY particles for the signal, while being concentrated to low values for the background.

- The $\alpha_T = \frac{p_{\mathrm{T}}^{j2}}{m_{jj}}$ variable [17], which takes large values for SUSY events with invisible particles while it is constrained to be $\lesssim$0.5 for (QCD) background events.

- The stranverse mass, m_{T2} [18].
- The contransverse mass, m_{CT} [19].
- The "razor" variables [20].

Except for the hadronic final states, the production of squark and gluinos can also lead to leptonic final states in numerous cases. For example, the process in Fig. 2.5b can give up to two leptons, originating from the possible leptonic decay of the W bosons, while the process in Fig. 2.5c can result in up to four leptons, if the W bosons from the top quarks decay leptonically. The requirement of leptons in the final state has the advantage of reducing the background contribution of SM processes that are difficult to estimate, such as multijet QCD events, rendering processes like $\mathrm{t\bar{t}}$ or Drell–Yan (DY) production, the most important backgrounds. These backgrounds are in general easier to estimate, though this comes at the cost of lower branching fractions (of the leptonic decay).

As a consequence of the change of the relative contribution of SM processes, slightly different handles are used for separating between signal and background. The transverse mass, $M_T(\ell, p_{\mathrm{T}}^{\mathrm{miss}}) = \sqrt{2p_{\mathrm{T}}^{\ell} p_{\mathrm{T}}^{\mathrm{miss}} (1 - \cos\Delta\phi)}$, is used to control the $\mathrm{t\bar{t}}$ and W+jets SM process, due to the endpoint it exhibits at the mass of the W boson, while the on-shell Z boson from $\mathrm{Z} \to \ell\bar{\ell}$ SM processes are usually rejected by excluding a mass window around the M_{Z}.

The results of searches for SUSY particles through squark and gluino production have been interpreted in the framework of simplified models. In such models, most of the SUSY particles are decoupled from LHC accessible physics. The phenomenology is determined by a few specific SUSY particles, whose masses are up to the ∼TeV scale and free to vary. Fig. 2.6 shows the exclusion limits on SUSY particle masses at 95% confidence level (C.L.) in terms of a few characteristic simplified models involving the production of colored sparticles. For the model in Fig. 2.5a, the full hadronic search using the M_{T2} variable excludes squarks and the $\widetilde{\chi}_1^0$ for masses up to ∼1800 GeV and ∼1000 GeV respectively. The full hadronic search exploiting the $p_{\mathrm{T}}^{\mathrm{miss}}$ variable and the full hadronic stop search achieve the maximum performance in the models of Fig. 2.5b and Fig. 2.5c respectively, excluding the gluino with masses up to $\gtrsim$2000 GeV and the $\widetilde{\chi}_1^0$ with masses up to $\gtrsim$1200 GeV.

The above discussion on squarks mainly focuses on the first and second generation squarks. As explained in Sec. 2.3, third-generation squarks have

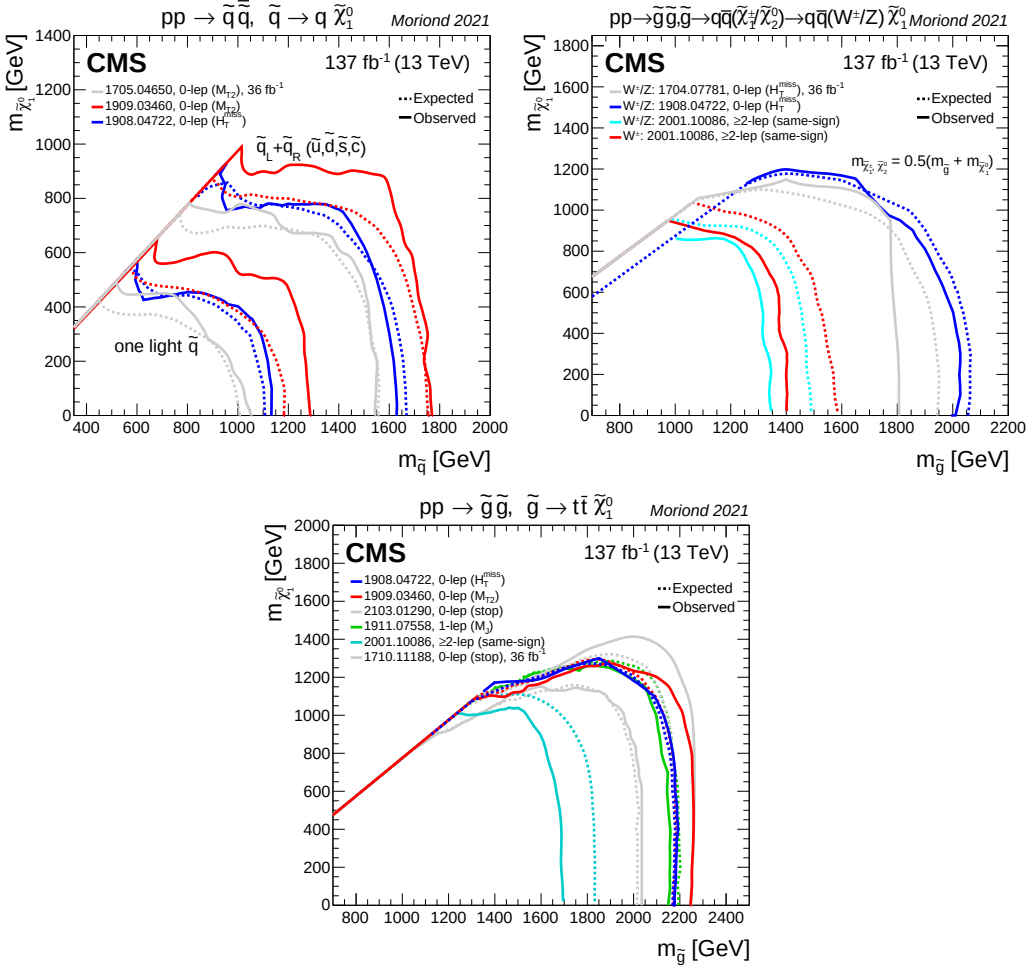
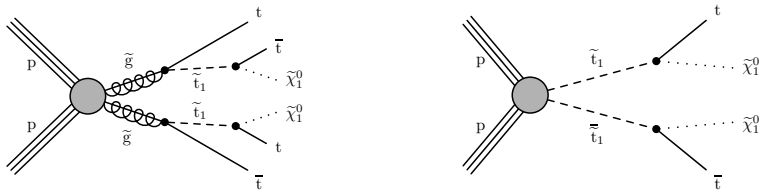


Figure 2.6: Results by CMS Collaboration for the lower mass limits, at 95% C.L., on squark pair production with $\tilde{q} \to q\widetilde{\chi}_1^0$ (left) and on gluino pair production with $\tilde{g} \to q\bar{q}'\widetilde{\chi}_1^{\pm}$, $\widetilde{\chi}_1^{\pm} \to \mathrm{V}^{\pm}\widetilde{\chi}_1^0$ (middle) and with $\tilde{g} \to \mathrm{t}\bar{\mathrm{t}}\widetilde{\chi}_1^0$ (right) [21]. The results are based on the simplified models shown in Fig. 2.5.

distinctive properties and, as a result, they are treated differently in searches as well. Taking as an example the top squark[3], it can be produced either through gluino pair production (2.7a) or directly (2.7b).

The final state and, consequently, the search strategy is highly dependent on the mass of the top squark and the mass difference with the SUSY particle

[3]In the following, the term "stop" is used interchangeably to denote the top squark.



(a) $pp \to \tilde{g}\tilde{g}$, $\tilde{g} \to t\tilde{t}_1$, $\tilde{t}_1 \to \bar{t}\tilde{\chi}_1^0$

(b) $pp \to \tilde{t}_1\bar{\tilde{t}}_1$, $\tilde{t}_1 \to t\tilde{\chi}_1^0$

Figure 2.7: Diagrams involving the (direct or indirect) production of top squarks.

it decays to. If the mass difference is large enough so that an on-shell top quark can be formed, while $\tilde{t} \to b\tilde{\chi}_1^{\pm}$ is kinematically forbidden, then $\tilde{t} \to t\tilde{\chi}_1^0$ is the preferred decay mode. In this case, the existence of at least two top quarks in the final state in combination with significant amount of p_T^{miss}, originating from the LSPs, is the main handle to discriminate the SUSY signal from the SM backgrounds. To this end, highly sophisticated taggers are used to identify hadronic decays of the top quark [22]. In the case of semileptonic t decays, i.e. a leptonic decay of the daughter W boson, efficient identification of the b quarks [23], combined with the existence of isolated leptons, are used. The latter strategy applies also to the case when the $\tilde{t} \to b\tilde{\chi}_1^{\pm}$ is kinematically allowed. For the cases where the SUSY parameter configuration is such that the above decay modes are forbidden, the top squark can decay via $\tilde{t} \to c\tilde{\chi}_1^0$ or $\tilde{t} \to bf\bar{f}'\tilde{\chi}_1^0$. These decays pose experimental challenges, due to the significantly difficult identification of the c quark and the low p_T of the final fermions.

Fig. 2.8 shows the exclusion limit at 95% C.L. set by the CMS experiment in terms of the $pp \to \tilde{t}_1\bar{\tilde{t}}_1$, $\tilde{t}_1 \to t\tilde{\chi}_1^0$ simplified model. The results of different searches, depending on the number of leptons in the final state, are shown separately. The full hadronic final state has the highest sensitivity over the most range of parameter space. The exception to this is the case where the lightest top squark is almost degenerate in mass with the lightest neutralino, for which the semi-leptonic final state, i.e. the final state with one lepton, is more important. Nevertheless, the searches are complementary and all contribute to the determination of the combined limit. As can be seen in Fig. 2.8, the combined channels have higher sensitivity, and thus higher mass

limits, than the hadronic channel alone. For a massless lightest neutralino, top squark masses are excluded up to 1300 GeV, while the maximum excluded mass for the lightest neutralino in this configuration reaches up to 700 GeV.

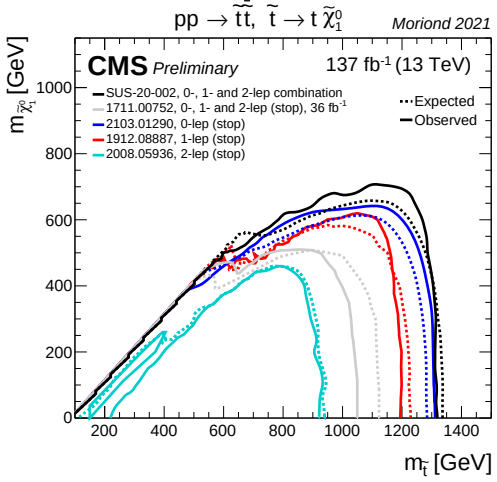
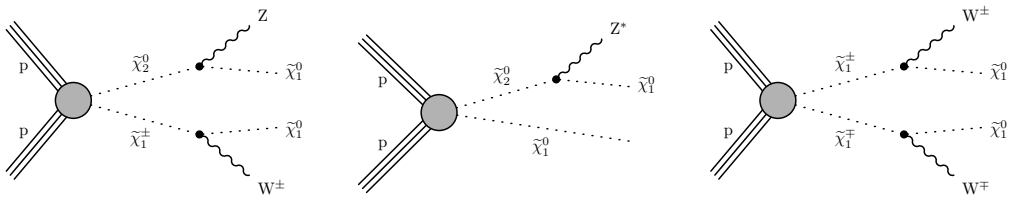


Figure 2.8: Results by CMS Collaboration for the lower mass limits, at 95% C.L., on top squark pair production with $\tilde{t} \to t\tilde{\chi}_1^0$ [21]. The results are based on the simplified model shown in Fig. 2.7, right.

Even though the production cross section of neutralinos and charginos, also called electroweakinos (EWKinos), is much lower than that of squarks and gluinos (Fig. 2.3), the EWK sector opens up the possibility to probe lower sparticle masses. The most interesting cases have been mentioned at the end of Sec. 2.3. A common feature of EWKino pair production is the existence of EW bosons in the final state. Even though hadronic decays of W, Z or H bosons are very SM-like, the requirement for leptonic decays offers satisfactory discrimination between SUSY signal and SM backgrounds. Fig. 2.9 shows different production modes and different decay modes in descending order of cross section from left to right.

The search strategy for final states originating from EWKino production depends on the mass difference between SUSY particles, which determines the mass of the final EW bosons. If the mass difference is larger than or near the mass of the Z or H bosons, then on-shell bosons are required in the final state, combined with moderate $p_{\mathrm{T}}^{\mathrm{miss}}$ and potentially jets. In the case of the Z bosons, high-p_{T} leptons are required to have a dilepton mass under a window around M_{Z}, while, in the case of the H boson, the decay products,



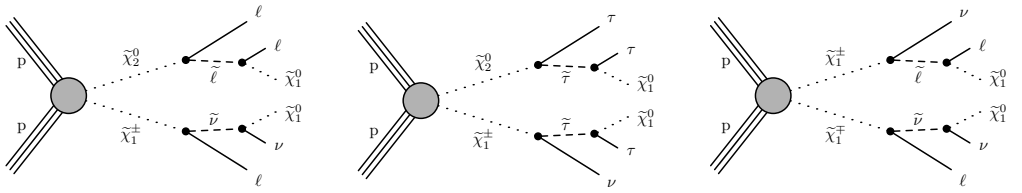
(a) $\widetilde{\chi}_2^0\widetilde{\chi}_1^\pm$ production (b) $\widetilde{\chi}_2^0\widetilde{\chi}_1^0$ production (c) $\widetilde{\chi}_1^\pm\widetilde{\chi}_1^\pm$ production

Figure 2.9: Diagrams of EWKino production, involving the decay modes $\widetilde{\chi}_2^0 \to \mathrm{Z}\widetilde{\chi}_1^0$ and $\widetilde{\chi}_1^\pm \to \mathrm{W}^\pm\widetilde{\chi}_1^0$.

e.g. $\mathrm{b\bar{b}}$, are tagged using different techniques as coming from an H boson decay. WZ(H) and ZH final states with all bosons decaying leptonically give rise to trilepton and four-lepton signatures. When one lepton in the final state is missed, e.g. not identifying one of the leptons coming from the Z boson decay of the process in Fig. 2.9a, the requirement of same-sign leptons provides a relatively SM-free signature. The dominant backgrounds in this case are misreconstructed events, e.g. events in which one of the leptons has been identified with charge of the wrong sign or originates from a heavy-flavor quark decay. Finally, there is the possibility that the mass difference between the lightest SUSY particles are far below the mass of the Z boson. In such scenarios, the $p_\mathrm{T}^\mathrm{miss}$ of the event tends to be low and the final state objects tend to have low p_T, hence posing experimental challenges in detecting, reconstructing and identifying the event objects. A detailed strategy for selecting such events and search for SUSY in this parameter space is described in Ch. 4.

The decay of EWKinos through sleptons is also possible, as shown in the diagrams of Fig. 2.10. Such scenarios can lead to similar final states as above, e.g. with a dilepton (Fig. 2.10a) or a trilepton signature (Fig. 2.10c), or final states with τ leptons (Fig. 2.10b). Final states with τ leptons can become important, since they are the dominant decay mode for the staus, which are considered to be the lightest sleptons due to their large mixing.

Based on the multiple final states of EWKino pair production, limits can be set on the masses of the $\widetilde{\chi}_2^0$ (taken to be mass degenerate with the $\widetilde{\chi}_1^\pm$) and $\widetilde{\chi}_1^0$ particles. The results of searches by the CMS Collaboration, interpreted in terms of simplified models with $\widetilde{\chi}_2^0\widetilde{\chi}_1^\pm$ pair production, are shown in Fig. 2.11. The maximum exclusion is achieved in terms of a simplified model with



(a) $\widetilde{\chi}_2^0\widetilde{\chi}_1^\pm$ production with flavor-democratic decays $\widetilde{\chi}_2^0 \to \ell\tilde{\ell} \to \ell\ell\widetilde{\chi}_1^0$, $\widetilde{\chi}_1^\pm \to \ell\tilde{\nu} \to \ell\nu\widetilde{\chi}_1^0$

(b) $\widetilde{\chi}_2^0\widetilde{\chi}_1^\pm$ production with τ-dominated decays $\widetilde{\chi}_2^0 \to \tau\widetilde{\tau} \to \tau\tau\widetilde{\chi}_1^0$, $\widetilde{\chi}_1^\pm \to \nu\widetilde{\tau} \to \nu\tau\widetilde{\chi}_1^0$

(c) $\widetilde{\chi}_1^\pm\widetilde{\chi}_1^\pm$ production with flavor-democratic decays $\widetilde{\chi}_1^\pm \to \ell\tilde{\nu} \to \ell\nu\widetilde{\chi}_1^0$

Figure 2.10: Diagrams of EWKino production, decaying via sleptons.

slepton-mediated, flavor-democratic decays (i.e. decays to all lepton flavors with equal probability) and reaches up $\widetilde{\chi}_2^0$ masses of 1300 GeV for very small masses of the $\widetilde{\chi}_1^0$.

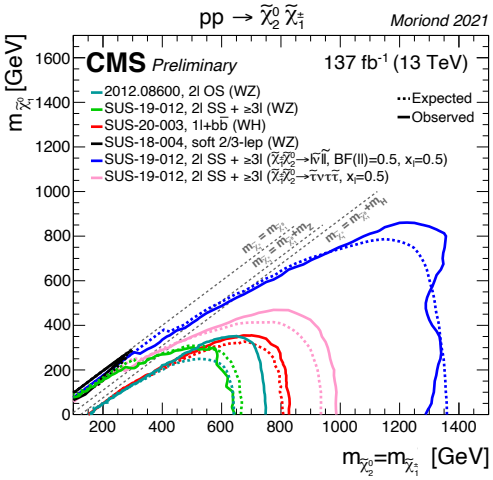
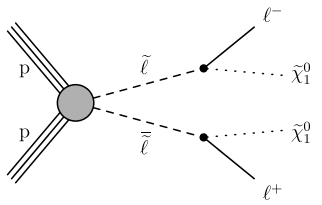


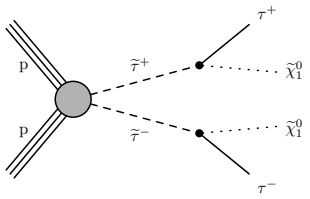
Figure 2.11: Results by CMS Collaboration for the lower mass limits, at 95% C.L., on $\widetilde{\chi}_2^0\widetilde{\chi}_1^\pm$ pair production with decay to multiple final states, as described in the plot legend [21].

Slepton pair production has even lower cross section at the LHC (Fig. 2.3). However, the signatures of slepton production are often similar to those of EWKino production, as can be seen from Fig. 2.12. As a result, searches with leptonic final states can also be interpreted in terms of simplified models of

slepton production. The exclusion limit for such a CMS search is shown in Fig. 2.13. The exclusion for first- and second-generation sleptons for a very light $\widetilde{\chi}_1^0$ reaches up to 700 GeV. Because of the special nature of third generation sfermions, described at the end of Sec. 2.3, searches targeting stau production (Fig. 2.12b) are specially designed. Stau masses of up 150 GeV are excluded for the case of $\widetilde{\tau}_R$ and $\widetilde{\tau}_L$ pair production with $m_{\widetilde{\tau}_R} = m_{\widetilde{\tau}_L}$ [24].



(a) First- and second-generation slepton pair production with $\widetilde{\ell} \to \ell\widetilde{\chi}_1^0$



(b) Stau pair production with $\widetilde{\tau} \to \tau\widetilde{\chi}_1^0$

Figure 2.12: Diagrams of direct slepton production.

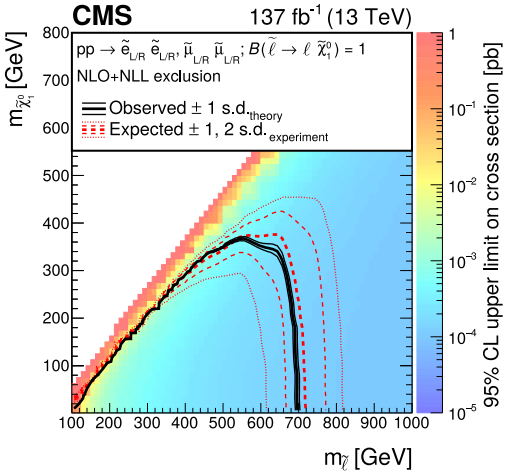


Figure 2.13: Results by CMS Collaboration for the lower mass limits, at 95% C.L., on first and second generation pair production with sleptons decaying to their corresponding superpartner and the $\widetilde{\chi}_1^0$ [25]. The results are based on the simplified model shown in Fig. 2.12a.

Chapter 3

The LHC Machine and the CMS Experiment

The European Organization for Nuclear Research, also called CERN which stands for *Conseil Européen pour la Recherche Nucléaire*, is an international research center established in 1954, in Geneva, Switzerland, with the aim of uniting European scientists towards pushing the boundaries of scientific and technological progress. The research was initially focused on nuclear physics, however the advancements in the knowledge of phenomena deeper than the nucleus have shifted the scientific program towards particle and high energy physics. Even though it started out as intergovernmental organization of 12 European countries, CERN has now expanded to 23 member states. It employs about 2600 staff members and provides infrastructures for around 12500 users of over 100 nationalities, representing more than 500 institutes universities from all around the world. To attain the best results in terms of physics requires technological achievements that can have applications much farther than fundamental science and can benefit the society as a whole. The ultimate example of this is the invention of the World Wide Web (WWW) in 1989 by Tim Berners-Lee, a British scientist working at CERN. The original conception of project and its development was realized to cover the need to automatically and promptly share information across scientists all over the world but it is now used in the everyday of most people in the world.

CERN operates the largest particle physics laboratory of the world, with its main facility being the LHC, the largest proton collider ever built. The LHC consists of two circular beam-pipes with a circumference of 26.7 km, situated in a tunnel 50 to 175m under the surface. Inside the beam-pipes,

hadrons (protons or ions) travel near the speed of light in opposite directions. Superconducting magnets are used to keep the particles along the circular track of the beam-pipes, while a number of accelerating structures boost the energy of the particles along their trajectory. A detailed description of the LHC machine is given in Sec. 3.1. The beam-pipes cross in four interaction points (IPs), enabling two proton beams to collide. In each of those IP, particle detectors are installed. The four major detectors are the ALICE detector [26], the ATLAS detector [27], the CMS detector [28] and the LHCb detector [29]. The CMS detector, whose recorded collision data have been used for the conduction of this thesis, is described in greater detail in Sec. 3.2. The most remarkable recent achievement realized with the collision data produced by the LHC is the discovery of the Higgs boson by the ATLAS and CMS collaborations in 2012 [30, 31], which led to the award of the Nobel prize to François Englert and Peter Higgs, the theoretical physicists who predicted its existence.

3.1 The Large Hadron Collider

The concept of the LHC dates back to the early 1980s, when its predecessor, the Large Electron-Proton (LEP) collider was being designed and built. The official approval for the construction of the LHC came about a decade later, in 1994. Between 1996 and 1998, the four major LHC experiments were also officially approved and the construction work commenced. The first beam circulated in the LHC on September 10th 2008. The physics operation of LHC is intermittent, with three eras of operational runs separated by shutdown periods for maintenance and upgrades. Run 1 was inaugurated with the first collisions on November 23rd 2009 and lasted until 2013. Run 2 of the LHC took place from 2015 to 2018, while Run 3 is scheduled from 2022 to 2024. The periods 2013–2015 and 2018–2022 constitute the LHC *Long Shutdown* (LS) 1 and LS2 respectively. The operational runs described above are collectively referred to as *Phase 1* of the LHC. Phase 2 of the LHC, taking place from 2027 to 2037, will involve major upgrades of the machine and the experiments that are planned to be implemented during LS3 (2024–2027). The upgraded LHC machine of Phase 2 is called *High Luminosity LHC* (HL-LHC) and it will be described in Sec. 5.1.

The LHC has been installed in the same tunnel as the LEP and uses much of the previous infrastructure. The LHC is the last ring in a large accelerator

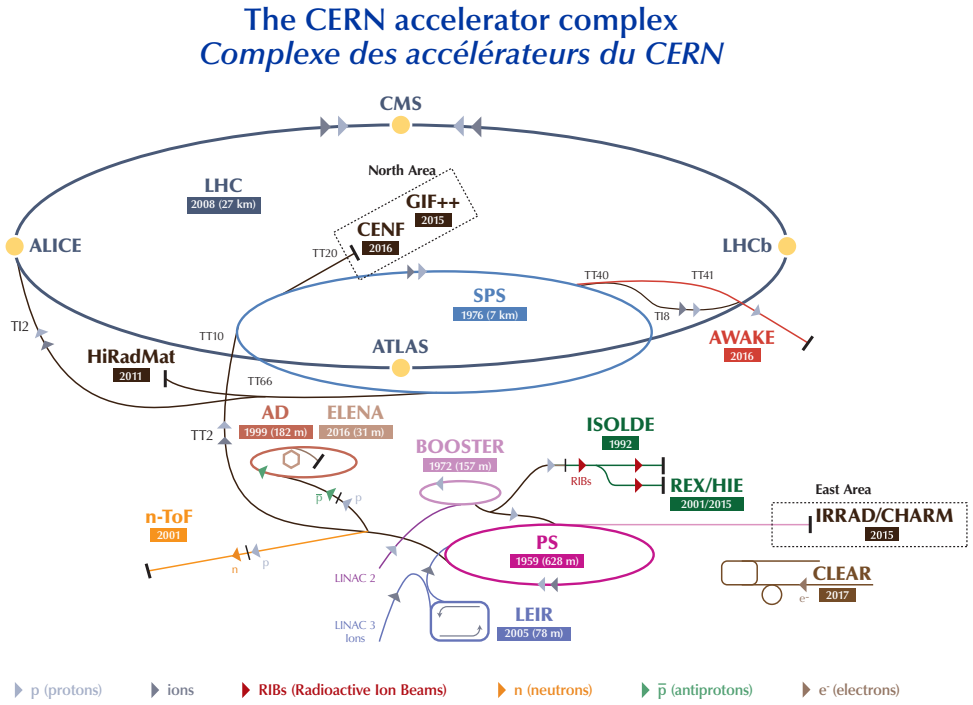


Figure 3.1: Illustration of the accelerator complex at CERN. The position of the four major LHC experiments are also shown [32].

complex, shown in Fig. 3.1. The journey of the protons that later come to constitute the beams of the LHC starts from a bottle of hydrogen gas, where strong electric field is used to strip the atoms from their electrons to yield protons. Instead of having a continuous beam, the extracted protons are then grouped into up to 2808 bunches, separated by a 25 ns interval. Each bunch contains 115 billion protons with an energy of 750 keV. Following this procedure, the proton bunches are sent to the first accelerator in the chain, the *Linear Accelerator* (LINAC 2), which raises their energy up to 50 MeV. After that, a series of circular accelerators is employed for further boosting the energy of the protons: The *Proton Synchrotron Booster* (PSB)

accelerates the protons to 1.4 GeV, the *Proton Synchrotron* (PS) pushes the beam to 25 GeV and the *Super Proton Synchrotron* (SPS) boosts them to an energy of 450 GeV.

The final step of the acceleration process that brings the protons to the nominal energy of operation, 6.5 TeV, is their injection in the LHC ring. The injection procedure takes approximately 4 minutes, while raising the proton energy to its maximal value requires around 20 minutes. The acceleration is achieved by the use of eight radiofrequency (RF) cavities per beam. To keep the beams in their approximately circular trajectory, 1232 dipole magnets are used. The required magnetic field for this purpose is 8.3 T, which creates the need for superconducting magnets, made out of a niobium-titanium alloy (NbTi). These magnets need to be cooled down to a temperature of 1.9 K by pumping superfluid Helium to the magnet system. 392 quadrapole magnets are also employed in order to maintain the beams focused. When the maximum energy is reached and the two opposite-travelling beams are stabilized, they can circulate in the LHC for more than 10 hours while intersecting at the four IPs, where the four major detectors collect the data produced by their collisions.

There are numerous parameters that characterize the LHC accelerator. One of the most important ones is the energy of the collisions. The advantage of having beams travelling in opposite directions means that the total energy of the collisions, also called *center-of-mass energy*, $\sqrt{s}$, is the sum of beam energies, as opposed to a fixed target experiment, where the total available energy is just $\sqrt{E_{beam}}$. During Run 1, the centre-of-mass energy was set at 7 and 8 TeV, while $\sqrt{s} = 13$ TeV during Run 2. To achieve such high energies the choice of particle is also crucial: Protons are used in the LHC because they are stable and they are preferred over electrons, due to their much larger rest mass, which minimizes the energy loss due to synchrotron radiation. The nominal design energy value for the LHC is $\sqrt{s} = 14$ TeV. Heavy ions (HI), more specifically Pb nuclei, are also accelerated at the LHC, leading to Pb-Pb or p-Pb collisions.

Another LHC parameter of pivotal significance is the instantaneous luminosity, L_{inst}, which determines the rate of collisions and, consequently, the number of events of physics processes. Supposing a process with cross section σ, the number of events of this specific process per unit time is:

$$\frac{\partial N}{\partial t} = L_{\text{inst}} \times \sigma \tag{3.1}$$

Eq. 3.1 makes clear that rare processes, i.e. those with very low cross section, require high luminosity to be observed. The units of L_{inst} are $m^{-2}s^{-1}$. Under the assumption of identical two beams colliding while travelling in opposite directions, the instantaneous luminosity is given by:

$$L_{\text{inst}} = \frac{N_p^2 n_b f \gamma_r}{4\pi\epsilon_n\beta^*}F,\ F = 1/\sqrt{1+\frac{\theta_c\sigma_z}{2\sigma}} \tag{3.2}$$

where N_p is the number of protons per bunch, n_b is the number of bunches, f is the revolution frequency of the bunches and γ_r is the Lorentz factor. The transverse emittance, ϵ, is a measure of the dispersity of the beam in space and momentum. A low emittance particle beam contains particles that are confined to a small distance and have nearly the same momentum, hence the likelihood of particle interactions will be greater resulting in higher luminosity. The transverse normalized emittance, ϵ_n also takes into account the change of the emittance as a function of the beam momentum, by considering the particle velocity in a small angle around the beam. It has units of length×angle. The amplitude function, β^*, characterizes the the focus of the beam at the IP. It is often referred as the distance from the focus point that the beam width is twice as wide as the focus point. It has units of length. Finally, F is a geometric factor that takes into account the crossing angle at the IP (θ_c) and the transverse (σ) and longitudinal (σ_z) size of the beam. Table 3.1 summarizes the nominal numerical values of the parameters discussed above for the LHC.

Parameter	Symbol	Nominal Value
Center-of-mass energy	$\sqrt{s}$	14 TeV
Instantaneous luminosity	L_{inst}	$10^{34}cm^{-2}s^{-1}$
Number of protons per bunch	N_p	1.15×10^{11}
Number of bunches per beam	n_b	2808
Revolution frequency	f	11245 Hz
Lorentz factor	γ_r	7461
Transverse normalized emittance	ϵ_n	3.75 μm rad
Beta-star function (at the CMS IP)	β^*	0.55 m
Geometric factor at min. β^*	F_0	0.836

Table 3.1: The nominal values of crucial LHC parameters in p-p collisions [33].

The total amount of data collected by the experiments is characterized by the integrated luminosity, $L = \int L_{\text{inst}} dt$. This is commonly measured in units of inverse femtobarn, $1\,\text{fb}^{-1} = 10^{-39}\ \text{cm}^{-2}$. The total integrated luminosity recorded by the CMS experiment is shown in the upper plot of Fig. 3.2 as a function of time.

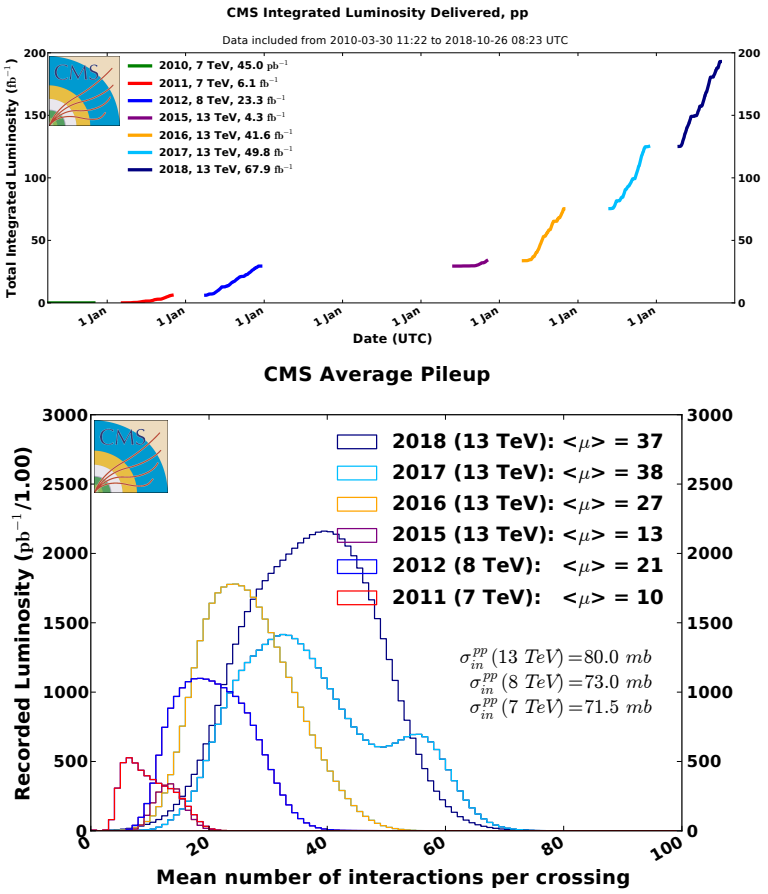


Figure 3.2: Upper: Cumulative luminosity versus day delivered to CMS during stable beams for 2010–2012 and 2015–2018 (p-p data only) at nominal center-of-mass energy. Each year starts at the endpoint of the previous year. Lower: Distribution of the average number of interactions per crossing (pileup) for pp collisions for 2011–2012 and 2015–2018 The overall mean values and the minimum bias cross sections are also shown. [34].

The high instantaneous luminosity of the LHC creates the inescapable

problem of *pileup* (PU), i.e. the number of simultaneous p-p interactions per bunch crossing. High PU conditions pose challenges for reconstructing and identifying physics objects, which leads to a degradation of their efficiency and resolution. The evolving average PU during the LHC operation, as recorded by the CMS experiment, is shown in the lower plot of Fig. 3.2.

3.2 The Compact Muon Solenoid Experiment

The Compact Muon Solenoid (CMS) experiment is one of the four major experiments installed at the IPs of the LHC. More specifically, the CMS apparatus is situated at IP5, near Cessy, France. It is one of the two general purpose experiments of the LHC (the other being the ATLAS experiment), designed to probe a broad range of physics phenomena, with focus on the study of the Higgs boson and the search of physics beyond the SM (BSM) at the TeV scale.

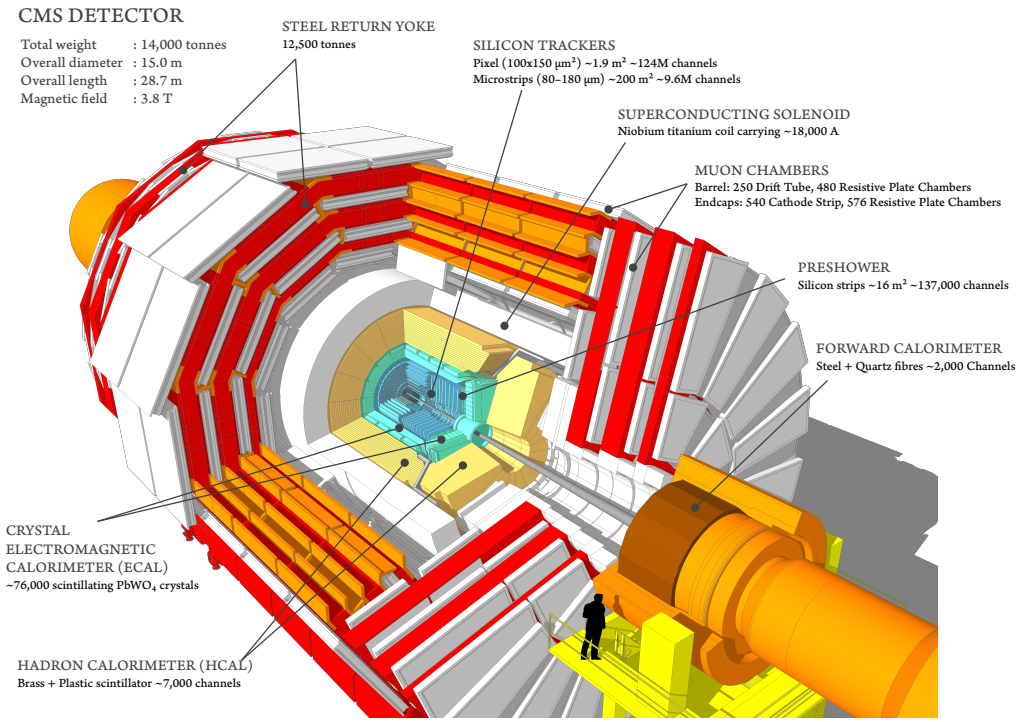


Figure 3.3: Cutaway diagrams of CMS detector [35].

A schematic view of the CMS detector is shown in Fig. 3.3. The apparatus has been designed in accordance to the concept of cylindrical layers around the beam axis. Its central feature consists of a solenoidal superconducting magnet that produces the magnetic field required to bent the paths of charged particles, so that their momentum can be measured. The solenoidal magnet is described in greater detail in Sec. 3.2.1. The rest of the subdetectors that constitute the CMS apparatus, from the beam axis towards the outer parts of the detector are: the silicon tracker (Sec. 3.2.2), the electromagnetic calorimeter (Sec. 3.2.3), the hadronic calorimeter (Sec. 3.2.4) and the muon chambers (Sec. 3.2.5). As a whole, the CMS detector is 21 m long, 15 m wide and 15 m high. Its total weight is 14000 tonnes.

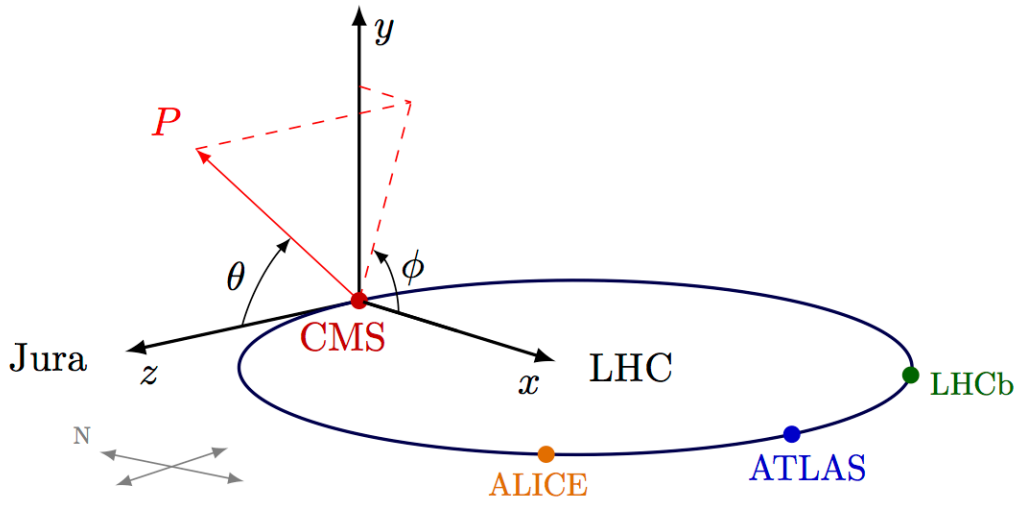


Figure 3.4: The coordinate system of the CMS detector [36].

Before delving into the description of each separate CMS subdetector, it is worth mentioning a few conventions regarding the CMS coordinate system. The origin of the coordinate system is the IP inside the CMS detector. The x-axis points towards the center of the LHC ring, the y-axis points vertically upwards and the z-axis points is aligned with the beam axis with each direction determined by the relation $\hat{z} = \hat{x} \times \hat{y}$. In terms of spherical coordinates, the distance from the origin is defined as $r = \sqrt{x^2 + y^2}$, the azimuthal angle ϕ is measured on the x-y plane, starting from the x-axis, while the polar angle θ is measured on the y-z plane, starting from the z-axis. Fig. 3.4 shows an illustration of the coordinate system described above.

As already mentioned in Sec. 2.4, the z-component momentum of the partons that participate in the hard scattering is not known. However, their momentum is zero in the transverse plane. As a result, the transverse component, defined in the x-y plane of variables, such the momentum or the mass, are used. The pseudorapidity, $\eta = -\ln\tan(\theta/2)$ (Fig. 3.5, is also used for the same reason: The difference of pseudorapidity is invariant to Lorentz boosts along the z-axis for massless particles. Following this, the angular distance is defined as $\Delta R^2 = \Delta\phi^2 + \Delta\eta^2$.

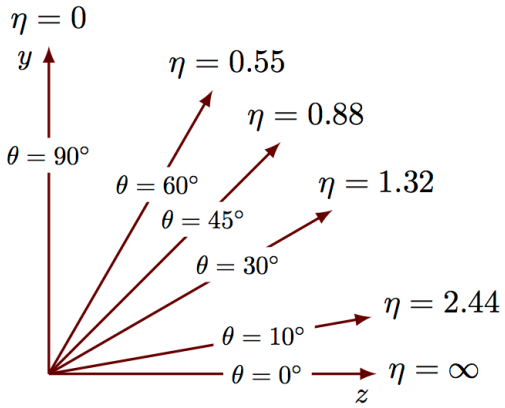


Figure 3.5: The relation between the polar angle θ and the pseudorapidity η [36].

3.2.1 Solenoidal Magnet

The 13m-long solenoid superconducting NbTi magnet [37] is the main component of the CMS detector. With a diameter of 6 m and a length of 12 m, it surrounds both the tracker and the calorimeters. This configuration achieves two goals: It maintains an almost ideally homogeneous magnetic field within the whole volume of which the CMS tracker and calorimeters can be found, while ensuring the minimum amount of material in front of these subdetectors. The magnitude of the magnetic field reaches up to 3.8 T at the center of the detector, which requires a current of 18160 A. To sustain such a high current, the superconducting magnet is kept to temperatures down to 4.7 K with the use of a liquid helium cryostat. An illustration of the magnitude of the magnetic field and its field lines is shown in Fig. 3.6.

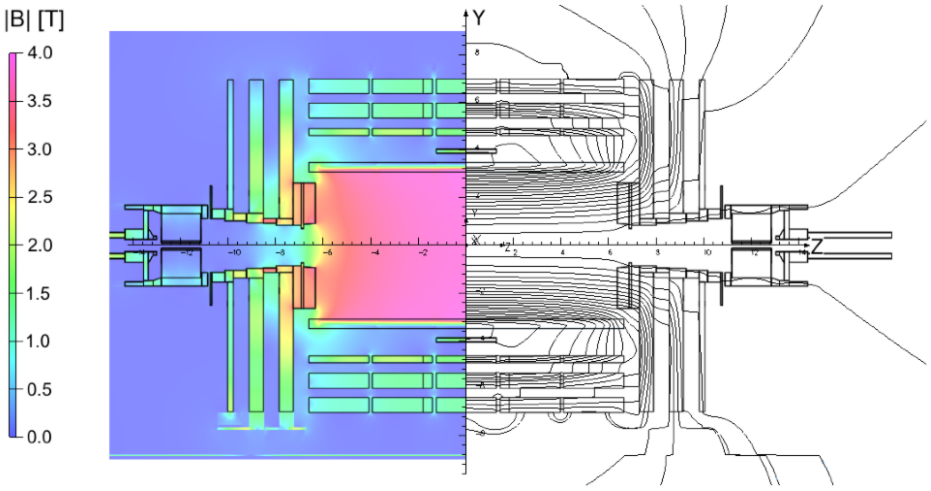


Figure 3.6: Map of the $|B|$ field (left) and field lines (right) predicted for a longitudinal section of the CMS detector [38].

Interleaved with the muon detectors, a three-layer iron structure acts as a return yoke for the magnet. It reaches out to a diameter of up to 14 m and also functions as a filter so that only muons or weakly interacting particles, such as the neutrinos, can escape to the muon chambers. This iron structure provides structural supports to the whole detector.

3.2.2 Silicon Tracker

The CMS tracker [39, 40] is the first subdetector encompassing the beam pipe. Its main goal is the tracking of the path of charged particles and the measurement of their momentum and charge. Due to its proximity to the beam pipe, it is specially constructed to be able to withstand enormous particle fluxes. The material that fulfils this requirement while also maintaining precise position measurements and fast readout even in harsh conditions is the silicon. In order to achieve a compromise between excellent detector performance and the minimization of the amount of inactive material, two configurations of silicon modules are used: The inner pixel tracker is the part closest to the IP and provides increased spatial resolution at the region of the detector where the occupancy is the highest. The outer strip tracker is optimally designed to maintain extremely high efficiency while keeping the material budget low. The CMS tracker as a whole extends from the beam

pipe up to $r = 1.1$ m, covering up to $|\eta| = 2.5$. An illustration of the subdetector in an r-z view is shown in Fig. 3.7.

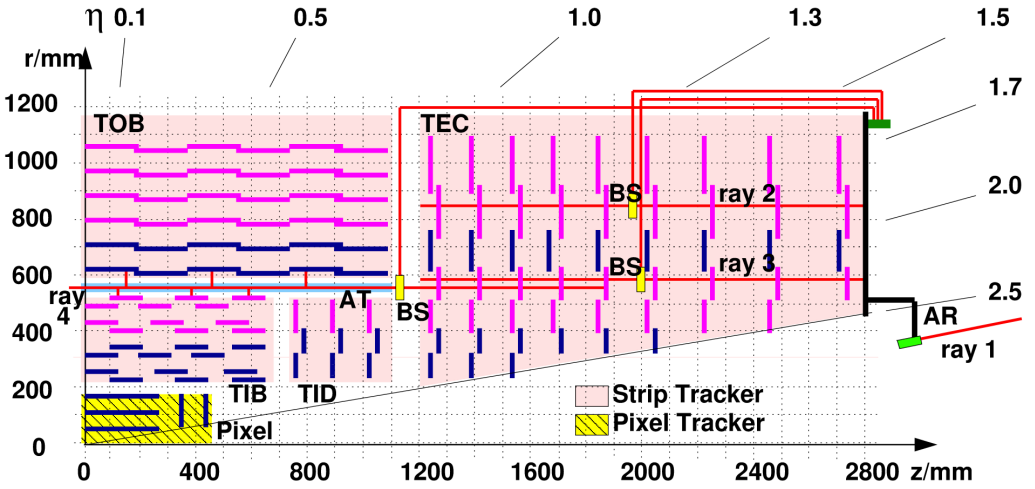


Figure 3.7: A quarter of the CMS silicon tracker in the r-z plane. The purple lines represent single module, while dark blue lines indicate double-sided modules. [41].

The increasing luminosity conditions of Run 2 created the need for an upgraded pixel detector [42]. The upgrade took place during the LHC shutdown between the 2016 and the 2017 data-taking period. The upgraded pixel detector consists of four barrel layers (BPIX) at a distance of 3.0, 6.8, 10.2 and 16 cm respectively and three endcap disks (FPIX) from $|z| = 29.1$ cm to $|z| = 51.6$ cm. The geometric layout of the upgraded pixel detector and a comparison with the previous design is shown in Fig. 3.8. In total, the pixel detector is comprised by 125×10^6 silicon pixels with dimensions 100×150 μm.

The outer strip tracker is composed of four different regions with different geometries and module arrangements to ensure full coverage. The tracker inner barrel (TIB) has four layers starting at a radius of 20 up to 55 cm, while the tracker outer barrel (TOB) consists of six layers, extending up to a radius of 1.1 m. Three disks on each endcap at longitudinal distance of $58 < |z| < 124$ cm constitute the tracker inner disks (TID), while the region $124 < |z| < 282$ cm is covered by the nine disks of the tracker endcap (TEC) on each side. Even though a large part of the outer tracker consists of single modules, double-sided modules, which enable measurements of z in the barrel

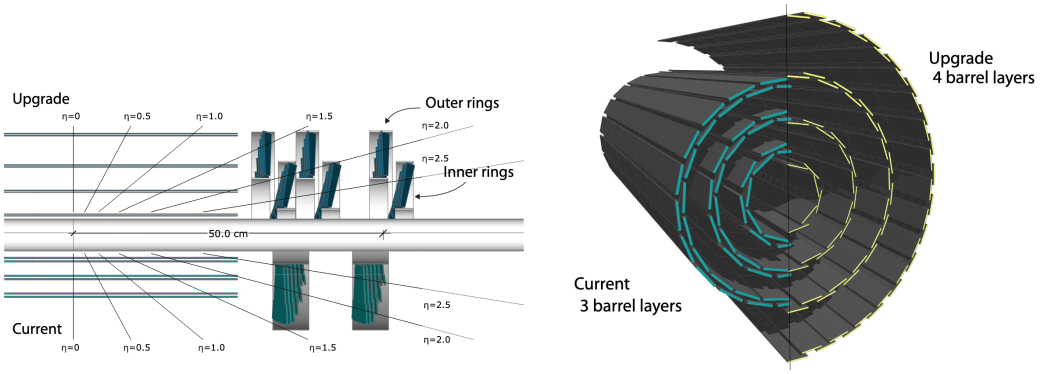


Figure 3.8: Illustration of the layout of the pixel detector before (labeled as "current") and after the 2017 upgrade. Left: View in the r-z plane. Right: Transverse-oblique view [42].

and of r in the endcap, are also utilized. The outer tracker is made out of 9.3×10^6 silicon strips of varying size and geometry.

3.2.3 Electromagnetic Calorimeter

The electromagnetic calorimeter (ECAL) [43] is the next subdetector, right outside the silicon tracker. It is designed to measure the energy of incident electrons and photons with high accuracy. The measurement is performed by means of converting the electron/photon energy to an electromagnetic shower that produces scintillating light upon the interaction with the calorimeter material. The CMS ECAL is homogeneous, as it is constructed out of crystals of lead tungsten, $PbWO_4$, which acts both as absorber and active material. Due to its very high density, it has a short radiation length, which makes it ideal for the compact design required so that the whole ECAL fits within the CMS magnet. It also produces rapid and spatial well-defined photon bursts that allow for precise measurements. However, the light yield is relatively low (30 photons/MeV, which necessitates the usage of light amplification in the photodetectors.

The CMS ECAL is hermetic, providing coverage for the region of $|\eta| < 3.0$ (Fig 3.9). To ensure the absence of acceptance gaps, the crystals have a small tilt of 3° with respect to the IP. The CMS ECAL is split into two parts: The barrel part (EB), comprised by 61200 crystals, extends up to $|\eta| < 1.479$, while the endcap parts (EE) cover the region of $1.479 < |\eta| < 3.0$ with 7324

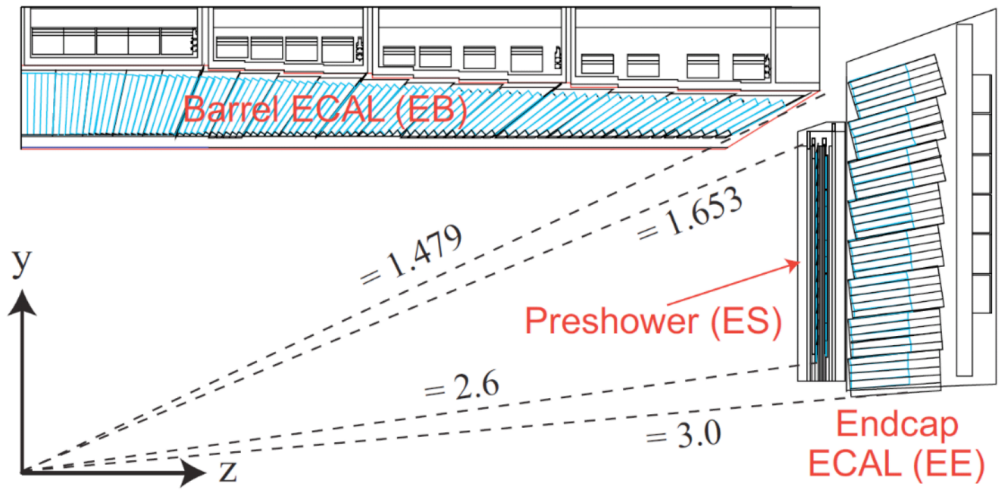


Figure 3.9: Geometric view of one quarter of the CMS ECAL in the r-z plane [44].

crystals each.

In front of the two EE, a preshower detector (ES) is installed. It is a sampling calorimeter which allows for extra spatial precision, utilized especially in distinguishing high-energy photons from lower-energy, collimated photon pairs. The ES subtends the region of $1.653 < |\eta| < 2.6$.

The energy resolution is one of the most important characteristics of the a calorimeter. In the case of the CMS barrel ECAL, the energy resolution is parametrized as:

$$\left(\frac{\sigma_E}{E}\right)^2 = \left(\frac{2.8\%}{\sqrt{E}}\right)^2 + \left(\frac{12\%}{E}\right)^2 + (0.3\%)^2, \tag{3.3}$$

where the first term accounts for the stochastic fluctuations in event-per-event basis, the second term originates from noise-related effects, e.g. electronics or PU, and the third term includes physical imperfections of the calorimeter, such as non-uniformities. As benchmark points, $(\sigma_E/E) = 1\%$ is quoted for $|\eta| = 1$ for a photon of a few tens of GeV and $(\sigma_E/E) = 2.5\%$ in quoted for a similar photon in the endcaps.

3.2.4 Hadronic Calorimeter

The CMS hadronic calorimeter (HCAL) [45] is a sampling calorimeter whose main purpose is to contain and measure the energy of hadrons. Since it is the only subdetector that can measure the properties of neutral hadrons, it is instrumental to the reconstruction of physics objects. The CMS HCAL consists of alternating layers of brass absorbers and plastic scintillator tiles. Since the nuclear radiation length is longer than the electromagnetic one, the HCAL is larger and more massive than the ECAL and part of it is placed outside the magnet coil in the barrel region. Similarly to the pixel detector, the HCAL was also upgraded during one of the Run 2 LHC shutdowns: At the end of 2017, the photodetectors of the calorimeter were replaced and an improved electronic readout system was installed. Timing and longitudinal depth information was also introduced, expanding the background rejection capabilities of the subdetector.

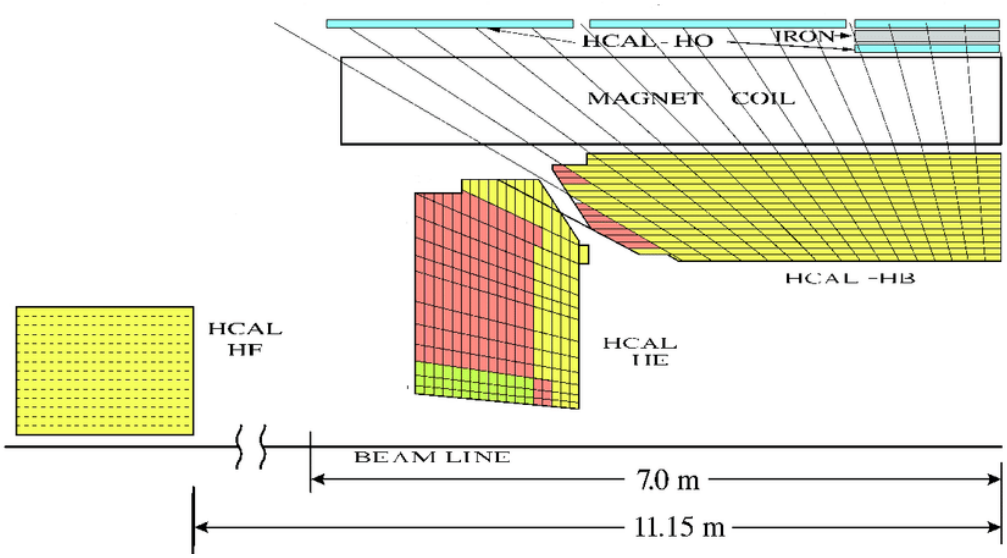


Figure 3.10: Geometric view of one quarter of the CMS HCAL in the r-z plane [46].

The HCAL is divided in four separate calorimeters: The barrel hadronic calorimeter (HB), $|\eta| < 1.4$, the endcap hadronic calorimeter (HE), $1.3 < |\eta| < 3.0$, the outer hadronic calorimeter (HO), which sits outside the magnet coil covering $|\eta| < 1.26$, and the forward hadronic calorimeter (HF), which is

installed at $z = 11.15$ m from the IP to provide acceptance for $2.9 < |\eta| < 5.2$. Because of its position, the HF receives enormous amount of particle flux and this is the reason why it is constructed from steel (used as absorber) and quartz fibres (used as sensitive material). The complete CMS HCAL is illustrated in Fig. 3.10.

The energy resolution of the HCAL is parametrized as:

$$\left(\frac{\sigma_E}{E}\right)^2 = \left(\frac{84.7\%}{\sqrt{E}}\right)^2 + (7.4\%)^2 \tag{3.4}$$

This modest resolution of the HCAL is compensated offline by algorithmic techniques (described in Sec. 3.4.3) that combine the full detector information to improve the resolution of individual subdetectors.

3.2.5 Muon Chambers

Due to their much higher mass, muons are less subject to radiative effects and can penetrate several meters of iron without interacting. As a result, the CMS muon system [47] is situated at the very edge of the detector, where muons are the only particles likely to register a signal. It sits outside the magnet coil and is interleaved with the iron return yoke, inside the 2 T return magnetic field. The CMS muon system consists of four concentric cylinders (muon stations) around the beam line (barrel, $|\eta| < 1.2$, and overlap region, $0.9 < |\eta| < 1.2$) and four disks at the end of the barrel (overlap and endcap region, $1.2 < |\eta| < 2.4$) and utilizes multiple detector technologies: 250 *drift tube chambers* (DTs) comprise the barrel muon subdetector, while 540 *cathode strip chambers* (CSCs) form the endcap muon detection layers. A redundant trigger system consists of 610 *resistive plate chambers* (RPCs) for the region $|\eta| < 1.9$. Finally, 10 prototype *gas electron multiplier* (GEM) chambers were installed as an upgrade to the muon system at the end of 2017 in the region $1.6 < |\eta| < 2.2$ [48]. Because of the many subdetector layers and the different advantages of each subdetector type that complement each other, the CMS muon system is robust and has a great capability in rejecting background noise. An illustration of the CMS muon subdetector as a whole is shown in Fig. 3.11.

The DT chambers have dimensions 2×2.4 m^2 and each is made up by up to 60 tubes. Each tube is 4 cm wide and contains a stretched wire within a mixture of Ar (85%) and CO_2 (15%), inside an electric field which is constant

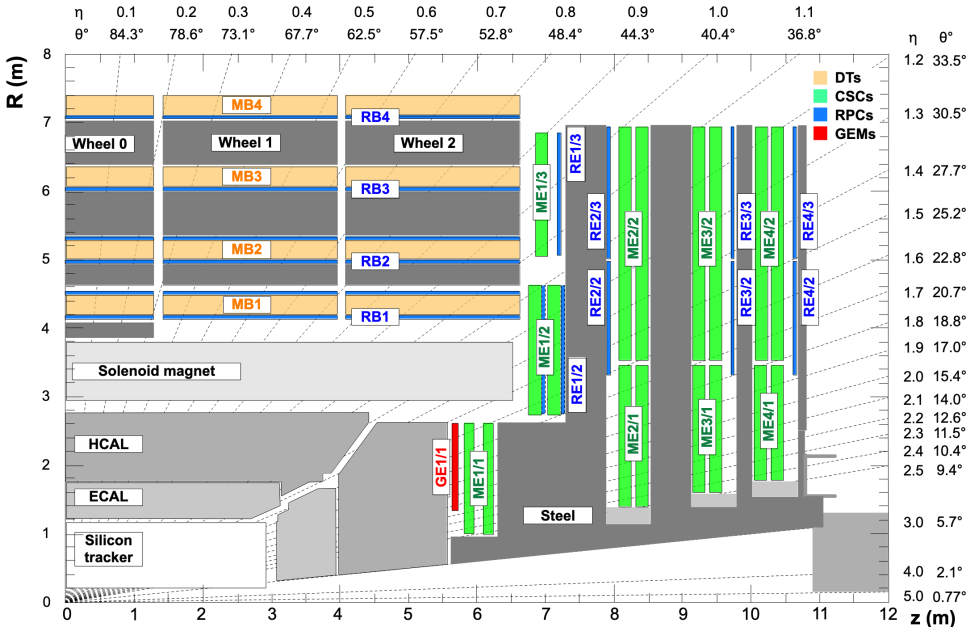


Figure 3.11: Illustration of the CMS detector in the r-z plane, depicting the locations of the various muon stations and the return yoke disks (dark areas). The drift tube stations (DTs) are labeled MB ("Muon Barrel"), the cathode strip chambers (CSCs) are labeled ME ("Muon Endcap") and the resistive plate chambers(RPCs) are labeled RB and RE for the the barrel and endcaps of CMS respectively. The prototype GEM chambers installed in the endcap at the end of 2017 are labelled as GE [48].

in time. When a muon passes through the tube, it strips gas atoms off some of their electrons, which subsequently travel along the electric field lines to the central wire. By measuring the time needed for the electrons to reach the wire, the position and angle of the incident muon can be inferred. The global position resolution of the DT chambers is $80 - 120$ μm [49].

The CSCs are mounted on the detector endcap, where the magnetic field is non-uniform and the particle flux is significantly higher. They consist of positively-charged wires (anodes) crossed with negatively-charged strips (cathodes) within a gas mixture of Ar-CO_2-CF_4 (30%-50%-20%). The electrons and ions created by the passing of a muon through a CSC move towards the anodes and the cathodes respectively, allowing for the extraction of two position coordinates. Apart from a fine segmentation that leads to a spatial

resolution of $40 - 150$ μm, the closed spaced wires of the CSCs makes them fast detectors, suitable for triggering [49].

The RPCs are formed by two parallel highly resistive, plastic plates, 2 mm thick. These are separated by a 95.5% $C_2H_2F_4$ and 4.5% i-C_4H_10 gas mixture. The operation of the RPCs is based on the phenomenon of electron avalanche: The electrons knocked off by the incident muon move through the electric field and hit other gas atoms, releasing even more electrons. The avalanche is registered by a set of detecting metallic strips after a small but precise time delay. The hit pattern on the strips can also provide an estimate of the muon momentum. Even though their spatial resolution is coarse ($0.8 - 1.2$ cm), the main advantage of RPCs is their excellent timing resolution, < 3 ns, used primarily for trigger decisions [49].

The current CMS GEM detector consists of ten 1–2 m^2 sized chambers filled with Ar-CO_2 gas mixture. Each chamber includes three layers of perforated copper-cladded polyimide foil. Voltage is placed on these layers, creating a large electric field in the foil holes. Incident muons ionize the gas and the electrons produced through this procedure pass through the holes to create an electron avalanche, which is then measured by the detector. The GEM detector provides additional redundancy and measurement points for the CMS muon system. An additional batch of 144 chambers have been installed during 2019–2020 in the first disk of both endcaps and even more chambers will be installed during the pre-Phase 2 LS, in 2024–2026.

3.3 The Trigger System of the CMS Experiment

The bunches of the LHC collide every 25 ns, leading to a rate of collisions of 40 MHz. This, combined with the fact that the typical size of a fully reconstructed event is ~ 1 MB, leads to truly enormous amounts of data that no current technology is able to read out and store. As a result, a rate reduction system is needed to select potentially interesting processes, as these are prescribed by the CMS physics program. The CMS trigger system achieves this rate reduction in two stages of increasing algorithm sophistication and complexity, described below:

- The *Level-1 trigger* (L1T) [50] is implemented on custom hardware processors running approximate event reconstruction algorithms with

the aim to reduce the event rate from 40 MHz to 100 kHz. The CMS L1T is described in Sec. 3.3.1.

- The *High-level trigger* (HLT) [51] utilizes a CPU farm to refine the event reconstruction and selection to achieve a further rate reduction down to $\sim$1 kHz. The CMS HLT is described in Sec. 3.3.2.

3.3.1 The Level-1 Trigger System

The CMS L1T system has to select potentially interesting events with a latency of 4 μs. As a consequence, it receives the raw data from only the muon and the calorimeter systems, since the iterative procedure of track finding requires much more time and is impossible to perform in the current L1T system. The architecture of Run 2 L1T system is shown in Fig. 3.12.

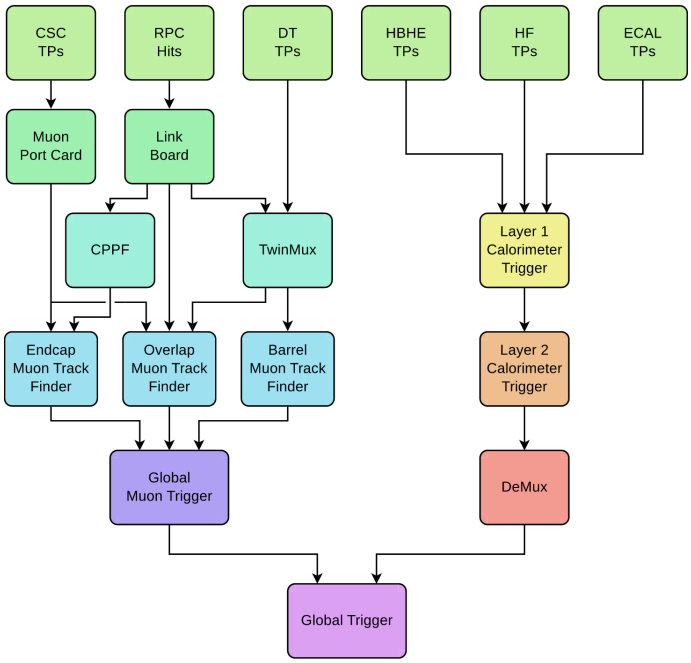


Figure 3.12: Diagram of the CMS Level-1 trigger system during Run 2 [50].

The coarse granularity and precision information from the muon and calorimeter detectors is combined to form low resolution physics objects, which are called *L1 candidates*. These can be L1 muons, L1 e/γ, L1 hadronic taus, L1 jets, L1 $p_{\mathrm{T}}^{\mathrm{miss}}$, L1 H_T, which are described in greater detail in the

next paragraphs. Their most basic kinematic observables, such as p_{T}, E_T or η, along with some more complicated ones, such as ΔR or invariant mass, are used for the event selection through the *L1T menu*, which is a list of algorithms, known as *seeds*. In the case of combining the selection criteria of different type L1 candidates, the seed is called *cross-seed* and has the advantage of lower observable thresholds, which increases the acceptance of analysis-targeted signals.

For the creation of L1 muons, the redundancy of partially overlapping subdetector systems is exploited. Three track finders are used, following the geometry of the muon subdetector. The *barrel muon track finder* (BMTF) receives data from the region $|\eta| < 0.83$. Muon hits constitute the trigger primitives (TPs), i.e. the input to the L1 trigger generated by the subdetector electronics, from the DTs and the RPCs. DT and RPC TPs in the same station are first sent to the TwinMux layer, which combines them into *superprimitives*, benefiting from the precise spatial resolution of DTs and the excellent timing resolution of the RPCs. The BMTF combines superprimitives to create the barrel L1 muon candidates.

The *overlap muon track finder* creates L1 muons in the region $0.83 < |\eta| < 1.24$. It utilizes unmerged DT TPs and RPCs hits from the TwinMux and some CSC TPs, which form 18 layers used to build tracks. For each track, a pattern recognition algorithm is initiated. It compares muon track patterns from simulated events with the track patterns formed in the detector layers, starting from a single hit as a reference point, preferably one from inner layers and with good ϕ resolution. Possible ambiguities are resolved by taking into account the ϕ spread probability density function in each layer, with respect to the reference hit. As a result, The reconstruction algorithm implemented in the OMTF can be considered as a naive Bayes classifier.

The L1 muons in the region $1.24 < |\eta| < 2.4$ are reconstructed by the *endcap muon track finder* (EMTF), which combines CSCs TPs with the hits from the RPCs mounted at the CMS endcap. A pattern recognition technique, similar to the one used in the OMTF, is utilized. The track p_{T} estimation at the endcap is complicated mostly by the irregular magnetic field strength and direction. As a solution to this complex problem, a boosted decision tree (BDT) regression is applied to compute the muon track p_{T} in the EMTF.

Finally, up to 108 muon candidates are received by the *global muon trigger* (GMT) by the three muon track finders. The GMT identifies and removes any possible duplicates and, subsequently, sorts the muons. The sorting is

performed initially performed based on the muon p_{T} and quality, keeping the 4 highest ranking L1 muons from each OMTF+EMTF endcap and the 8 highest ranking L1 muons from the barrel. A second sorting stage selects the eight highest ranking L1 muons out of these and the selected L1 muons are then forwarded to the *micro global trigger* (μGT).

The L1 e/γ, hadronic taus as well as the L1 jets and their sums are all reconstructed using information from the CMS calorimeters. The TPs of the L1 calorimeter trigger are the *calorimeter trigger towers* (TTs) which comprise of a 5×5 group of ECAL crystals and the corresponding HCAL tower behind them. The barrel TT size is $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$, while, in the endcap, where the subdetector configuration is more complicated, the TT size is variable, up to $\Delta\eta \times \Delta\phi = 0.17 \times 0.17$. The TPs from the ECAL and the HCAL are collected by the calorimeter trigger layer 1. There, the TPs are calibrated and sorted, before they are sent to the calorimeter trigger layer 2 in a time-multiplexed way, so that each layer 2 processing node has the information of the full event. The calorimeter trigger layer 2 is where all the L1 physics objects, except for the muons, are reconstructed. Finally, the physics objects are properly reordered and reformatted in the demultiplexer (DeMux) board before they are sent to the μGT.

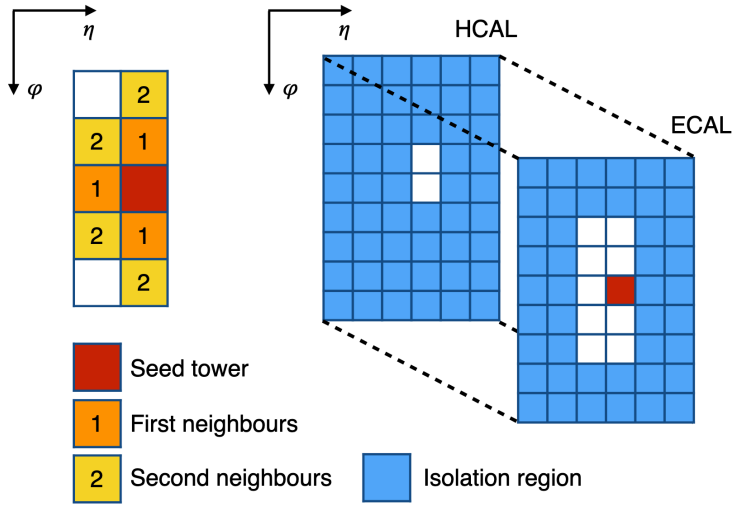


Figure 3.13: The L1 e/γ clustering and isolation definition [50]. Each square corresponds to a calorimeter TT.

In the absence of information from the tracker, L1 electrons and L1 pho-

tons are indistinguishable and are common referred to as L1 e/γ. The reconstruction of a L1 e/γ starts from a $E_T > 2\,\text{GeV}$ TT which represents the local maximum in its vicinity. This is called the *cluster seed*. Following the identification of the cluster seed, surrounding TT with $E_T > 2\,\text{GeV}$ are accumulated in such a way that only contiguous TTs, matching the electron footprint, are kept. A maximum $\text{TT}(\eta) \times \text{TT}_(\phi) = 2 \times 4$ size is used to reject contributions from the PU. The elongation in the ϕ direction takes into account energy deposits due to the bremsstrahlung of electrons, caused by the bending of their path by the CMS magnetic field. To minimize the background rate, additional identification criteria are applied to the L1 e/γ clusters: A shape veto against PU clusters, a veto against hadron-induced showers and a requirement for low ratio of HCAL to ECAL energy in the seed TT. An isolation requirement is also applied. It is calculated based on the energy deposits on a wider $\text{TT}(\eta) \times \text{TT}_(\phi) = 6 \times 9$ window around the seed and its threshold is optimized for different E_T ranges. An illustration of the clustering algorithm and isolation definition are shown in Fig. 3.13.

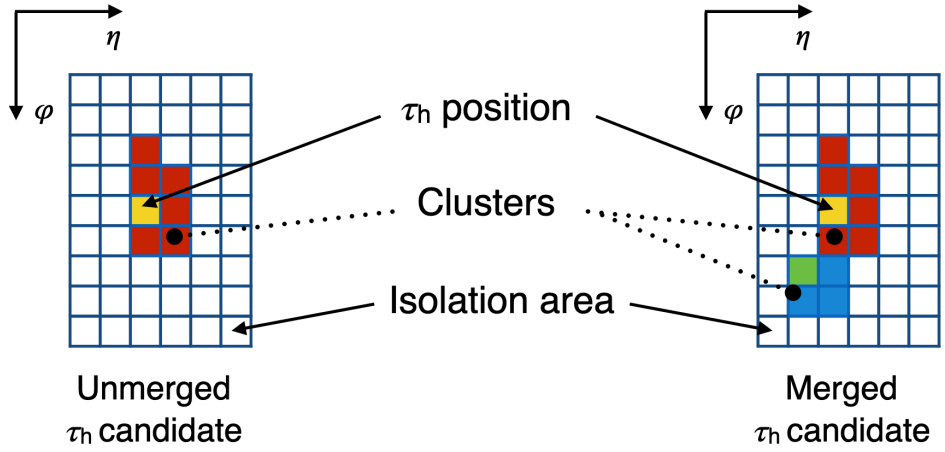


Figure 3.14: The L1 τ clustering algorithm and isolation definition [50].

While leptonically decaying τ leptons are reconstructed at the L1T as L1 muons or L1 e/γ, hadronically decaying taus, τ_h, require special reconstruction considerations. They can decay to one, two or even three charged or neutral pions which create calorimeter clusters that are generally separated due to the effects of the magnetic field. As a result, e/γ clustering algorithm is modified to accommodate the τ_h footprint and possibly merge neighbouring clusters (Fig. 3.14). An isolation criterion is again exploited to

discriminate genuine τ_h from PU-induced energy deposits but the threshold is generally relaxed with respect to the e/γ one.

The reconstruction of L1 jets follows a fixed-size clustering logic: Starting from the jet seed, defined as the local maximum TT with $E_T > 4\,\text{GeV}$, the total L1 jet energy is defined as the sum of the energy of all TTs in a 9×9 TT window around it. The window size is chosen such that it approximately matches the anti-k_T clustering algorithm (Sec. 3.4.7) size of 0.4 in the barrel. Several geometric restrictions are applied during the jet reconstruction in order to avoid the double counting of jets and to prevent TTs with the same energy to veto one another when considered as jet seeds. In addition to the L1 jets, calorimeter energy sums are computed with the full calorimeter granularity as well: L1 H_T is defined as the total scalar transverse energy of all jets and L1 $p_{\text{T}}^{\text{miss}}$ is defined as the magnitude of the vector sum of transverse energy of all the TTs. To lower the background rate and to improve the efficiency of these reconstruction algorithms for fixed rate, PU mitigation techniques are utilized. For L1 jets, a local PU E_T estimation is performed by considering the three lowest energy 3×9 areas around a jet window, as shown in Fig. 3.15, and is subsequently subtracted by the L1 jet energy. This correction is also propagated to the L1 H_T. The estimation of the PU correction on L1 $p_{\text{T}}^{\text{miss}}$ is based on low energy deposits in the central ($|\eta| < 0.34$) part of the detector.

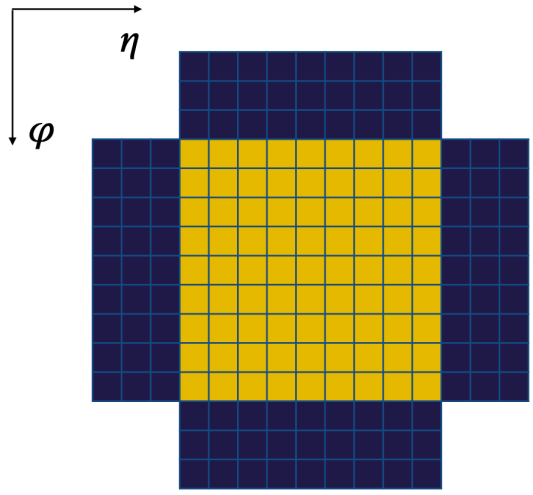


Figure 3.15: The definition of the areas used to measure the energy of a L1 jet (yellow) and to estimate the energy deposit by local PU (blue) [50].

The final decision of accepting or rejecting an event takes place in the μGT. The properties of the L1 candidates of an event are checked in parallel against the approximately 400 seeds of the L1T menu. Out of these, around 50 seeds are used for physics analysis that use the full luminosity recorded by CMS, around 100 are contingency seeds with more stringent conditions, while the rest are primarily used for ancillary measurements, such as calibration or monitoring. An event is accepted if it satisfies the requirements of at least one seed and this initiates a full detector readout for processing in the HLT.

3.3.2 The High-Level Trigger System

Using only the events selected by the L1T system, the CMS HLT system further reduces the event rate down to approximately 1 kHz to match the input capability of the data acquisition system (DAQ). The selection of events by the HLT is more refined due to the usage of streamlined versions of the full offline reconstruction. Owing to the looser timing constraints, the HLT system has access to tracker information, which permits the identification of tracks, vertices (even displaced ones) and of jets arising from b-quarks.

The HLT reconstruction must happen within $\sim$300 ms and is performed by $\sim$ 32000 CPU cores (2018 figure). It starts from objects identified by the L1T to avoid having to read out the full, raw event at once and at first processes the muon and calorimeter data before initiating the CPU-intensive task of processing the tracker information. It is worth noting that it is impossible to run the full vertex and track finding procedure withing the latency of the HLT and, as a result, targeted simplifications are employed: The primary vertex is found with information only from the pixel detector, while the tracks are built in several steps which differ based on the specific HLT algorithm they are going to be used for. The HLT algorithms, also known as *HLT paths*, are formed by a series of reconstruction modules and selection filters of increasing algorithmic refinement and physics sophistication that are progressively applied to HLT physics objects, in order to minimize the computational resource needs.

If any event manages to pass at least one full HLT path out of the approximately 600 that comprise the *HLT trigger menu*, it is accepted for offline analysis and is eventually sent to the CMS Tier-0 computing center for offline processing and permanent storage. Accepted events are categorized in non-mutually exclusive *HLT streams*, according to the HLT path that they passed. Apart from the physics HLT streams, additional calibration and

monitoring HLT streams are provisioned. To extend the physics acceptance even further, two additional data taking strategies have been implemented in the HLT: The *scouting* technique can achieve higher event rates by cutting down on the event size. This is accomplished by storing a reduced event content for the scouted events. The *parking* technique also allows for extra events with full event content to be stored on tape. However, these events do not follow the prompt reconstruction schedule of regular events but they are only reconstructed a posteriori, in cases where there is a strong physics motivation for extra data. The HLT trigger menu is highly flexible and constantly changes to accommodate the shifting priorities and goals of the CMS physics program.

3.4 Physics Object Reconstruction in CMS

The full reconstruction of a collision event involves the proper grouping of all the detector signals to sets that are compatible with the signature of physics objects. Fig. 3.16 shows the signature of different particles: Muons are charged particles and, as a result, they leave a trace in the tracker. However, due to their large mass, they are minimum ionizing particles and pass through the rest of CMS detector without interacting much. Apart from the tracker, they can only be detected at the muon chambers and this makes their signatures very characteristic. Electrons are also detected in the tracker system but, in contrast with the muons, they are stopped at CMS ECAL. Photons, similarly to the electrons, are absorbed in the ECAL but leave no trace in the tracker, since they have no charge. Hadrons differ from the previously mentioned particles in that only a small fraction of their energy is deposited in the ECAL, while most of their energy is lost in the HCAL. Charged hadrons are differentiated from neutral hadrons by the trace they leave in the tracker. Neutrinos and any other weakly interacting particles rarely leave a signal in the detector. Nevertheless, their existence can be inferred by the energy imbalance in the transverse plane, $p_{\mathrm{T}}^{\mathrm{miss}}$. After a discussion on the vertex and track reconstruction algorithms in Sec. 3.4.1, the basic principles of the particle flow algorithm are laid out in Sec. 3.4.3. A detailed description of the offline CMS reconstruction algorithms is given for each physics object type in Secs. 3.4.4 through 3.4.9.

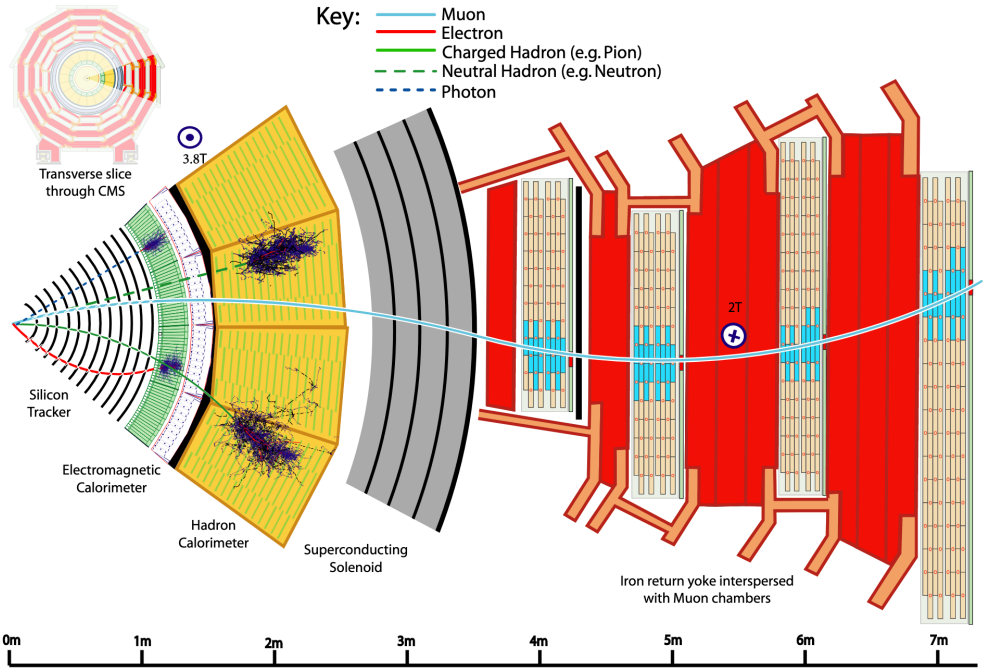


Figure 3.16: Illustration of the interactions of different particles in a slice of the CMS detector [52].

3.4.1 Tracks and Vertices

The track building procedure [53] is an extremely complicated part of the CMS reconstruction. The proximity of the tracker to the p-p IP leads to the very high occupancy of the subdetector, making the combinatorics difficult to handle. As a result, multiple reconstruction steps are performed before the full collection of tracks and vertices of an event is completed.

The first step of the track reconstruction is the *local reconstruction*, which involves the clustering of zero-suppressed pixel and strip channels, above a given threshold, into hits. The cluster position and its uncertainty is also computed. For the pixel detector, this is implemented with two algorithms: A fast algorithm, which creates *projected clusters* in the local detector coordinate system by summing all charges along an axis, is used for track seeding and pattern recognition. For the final track fit, a more precise, template-based algorithm is used, where the observed charge distributions are compared with expected from Monte Carlo (MC) simulation distributions. In the case of the strip detector, hits are formed by adding up signals from neigh-

bouring strips and their position is calculated from the average position of its constituents weighted by the charge of each one of them and corrected for several factors, e.g. the Lorentz drift. The hit efficiency for good modules reaches 99.8% and the hit resolution is between 20(10) and 50 μm for the pixel(strip) detector.

Before the actual track reconstruction can commence, two preparatory processes must be performed. First, the hits must be translated in the global coordinate system. During this procedure, discrepancies between the ideal and the real position of each subdetector element are taken into account. This is achieved by a process called *track alignment* [54]. Additionally, the position of the LHC beamspot, i.e. the 3-dimensional profile of the LHC luminous region, and an approximation of the event primary vertices are required, since they are used for constraining the track seed creation.

The CMS track reconstruction is performed by a software called *Combinatorial Track Finder* (CTF), which is a modified version of the combinatorial Kalman filter [55–57]. The software is based on the concept of *iterative tracking*: Multiple track finding iterations (maximum 6) are performed, with each one removing the hits that formed a track, therefore simplifying the procedure for next iterations. The first iteration, denoted as iteration 0, is responsible for finding most tracks, since it targets prompt tracks, i.e. the ones originating close to the p-p IP, with three pixel hits. Iteration 1 recovers prompt tracks with only two pixel hits. Iteration 2 is designed to reconstruct low-p_{T} prompt tracks, while iterations 3 through 5 aim at finding tracks coming outside the beamspot and tracks not covered by any previous iterations.

Each of the previously mentioned iterations can be broken into four separate stages. The first stage involves the generation of the track seeds. The seed generation is chosen to start from the innermost part of the tracker subdetector and it is subject to several weak restrictions that limit the hit combinations while ensuring excellent efficiency. Different combinations of seeds are created based on the type of hits used to form them, which are used in different iterations.

The next stage is the track finding module, based on the Kalman filter method. Starting from the track seed, a fast *analytical propagator* searches for hits in subsequent detector layers. Once compatible hits have been found, multiple passes with a more accurate *material propagator*, which takes into account the effect of multiple Coulomb scattering. This procedure repeats until a termination condition is met. Under conditions, also an inward track

building is performed. After all the tracks have been built, a trajectory cleaner procedure is applied that removes tracks corresponding to a single charged particle. Selection requirements on the numbers of hits, lost hits and the threshold to initiate the inward track building are applied depending on the iteration.

The third stage is the global track fitting. Starting from inside out, the position of hits is reevaluated by the means of a Kalman filter fit based on current track parameters. When the outermost hit has been reached, the procedure is repeated in the reverse order. After the fitting has been completed, a search for spurious hits is initiated and these are removed. Finally, the last stage involves the selection of tracks depending on their p_{T}, η, number of layers/hits, χ^2 of the fit and impact parameters. Loose criteria are designed as the minimum requirements for a track to be included in the general track collection. Progressively stricter criteria that reduce the efficiency and fake rate are applied for the tight and high-purity track collections, the latter of which are mostly used by physics analysis. The cumulative efficiency of the overall tracking performance from the six iterations of track reconstruction, as a function of the transverse distance from the beam axis, is shown in Fig. 3.17.

Vertex reconstruction used the reconstructed previously tracks to accurate measure the location of the p-p interaction vertices, along with its uncertainty. It proceeds in three steps: First, a selection of tracks is applied, imposing requirements on the track transverse impact parameter, the number of hits and the χ^2 of the trajectory fit. Subsequently, the track clustering is performed using a deterministic annealing algorithm [58]. Finally, the position of each vertex is estimated using an adaptive vertex fitter. This is followed by a vertex cleaning procedure, where vertices with a lot of tracks with high probability of being fake are discarded. The primary vertex (PV) is defined as the candidate vertex with the largest value of summed physics-object p_{T}^2, where the physics objects are the jets (discussed in Sec. 3.4.7) and the associated missing transverse momentum (discussed in Sec. 3.4.9). The PV reconstruction probability is approximately 98(100)% for vertices with (more than) two tracks and the vertex position resolution ranges from 10 to 100 μm depending strongly on the number of associated tracks.

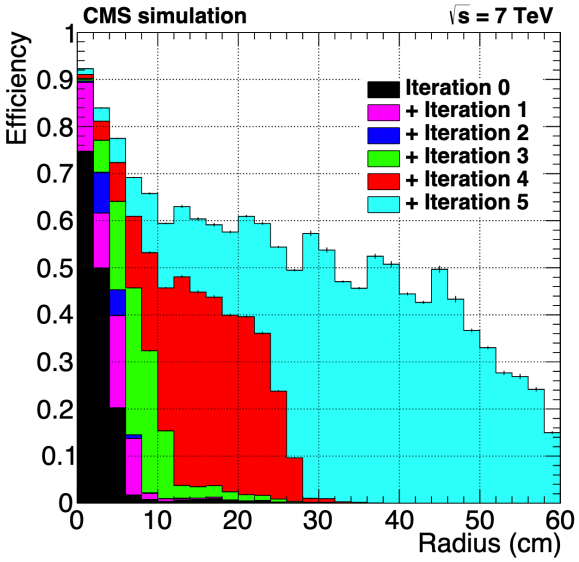


Figure 3.17: Cumulative efficiency of track reconstruction iterations as a function of the transverse distance from the beam axis to the production point of each particle. Tracks are required to pass the high-purity requirements, have $p_{\mathrm{T}} > 0.9\,\mathrm{GeV}$, $|\eta| < 2.5$ and transverse(longitudinal) impact parameter $< 60(30)$ cm. Simulated $\mathrm{t\bar{t}}$ events have been used [53].

3.4.2 Calorimeter Clusters

The creation of clusters out of the calorimeter deposits is a pivotal part of the CMS reconstruction workflow [52]. The formation of accurate, in terms of position and energy, clusters is extremely important for the detection stable neutral particles, such as photons and neutral hadrons, since these are not detected by the tracker system. Moreover, the combination of cluster properties with corresponding track parameters, as described in Sec. 3.4.3, assists in the proper reconstruction of charged particles. The clustering is performed separately in different layers of the CMS calorimeters: ECAL barrel and endcaps, HCAL barrel and endcaps, and the two preshower subdetectors. No clustering is applied for energy deposits in the HF.

The first step of the cluster building algorithm is the identification of *cluster seeds*. These are defined as the cells with the maximum energy, over a given threshold, with respect to their 4 or 8 nearest neighbours, depending on the subdetector. Following the determination of the seeds, *topological*

clusters are created by accumulating cells that share at least one corner with the cluster and have an energy deposit at least two times higher than the typical noise level subdetector. To account for the increase of the noise level in the endcaps, an E_T requirement is also set.

When the topological clusters have been formed, standalone clusters are identified within them. This is achieved with the usage of a Gaussian-mixture model, effectively breaking up the energy of the individual cells of the topological cluster into N Gaussian energy deposits, where N is the number of clusters seeds. An expectation-maximization algorithm is utilized which iteratively applied to find the parameters A_i and $\vec{\mu}_i$ of the postulated Gaussian i which is in the form $g_j = A_i e^{-(\vec{c}_j - \vec{\mu}_i)^2/(2\sigma)^2}$. The parameter $\vec{c}_j$ is the position of the jth cell and the parameter σ is the energy resolution of the respective subdetector, as given by Eq. 3.3 and 3.4. After the algorithm converges, the positions and energies of the Gaussian functions are taken as cluster parameters. Finally, the clusters are calibrated separately for electromagnetic and hadronic deposits in the different subdetectors to ensure good identification of neutral particles with very low probability of misreconstructed clusters.

3.4.3 The Particle-Flow Algorithm

In principle, physics objects can be reconstructed from signals registered in individual detectors. For example, the existence of photons and electrons can be solely inferred by the energy deposits in the ECAL and muons can be identified by information in the muon chambers, where they practically the only signal. However, a substantial improvement in the determination of the properties of the physics objects can be attained if elements from different subdetectors are correlated to identify them. This approach is called *particle-flow* (PF) reconstruction [52] and it is a core aspect of the CMS physics object reconstruction.

The PF algorithm can be conceptually split in two steps: First, the *building blocks* of the PF objects have to be reconstructed. These are the tracks and the clusters, described in Sec. 3.4.1 and 3.4.2 respectively. Then, these building blocks are associated with one another based on topological considerations to form the *PF blocks*. The properties of the PF blocks are evaluated and PF blocks are promoted to the proper final-state object accordingly.

The PF linking algorithm tests in principle any pair of PF building blocks to check whether they can be associated. To avoid a long computing time and a waste of computational resources from testing building blocks that

are separated by a large distance, only the nearest neighbours in the (η,ϕ) plane are considered in the linking procedure. Tracks are associated with calorimeter clusters if its extrapolated position, which varies on the type of cluster to be linked, falls within the cluster area. The combination of track-cluster with the smallest distance is kept in case of ambiguity. To efficiently reconstruct photons, which can be converted to an e^+e^- pair, and electrons, which can emit photons due to bremsstrahlung radiation, special linking considerations are taken into account. A cluster-to-cluster linking is also performed among different types of clusters: ECAL-HCAL and ECAL-preshower. A link is established if the ECAL cluster is within the HCAL cluster in the former case and if the preshower cluster is within the ECAL cluster in the latter. Ambiguities are again solved on the principle of smallest distance. Finally, links are also formed between central tracks and tracks in the muon system. These are described in the context of muon reconstruction in Sec. 3.4.4.

To begin with, muon candidates, which give the cleanest signature in the detector, are identified. The PF building blocks from which they were reconstructed are excluded from further consideration. The electron identification and reconstruction follows, with energetic and isolated photons being identified in the same step. Again, their constituents are removed from the PF building blocks list. Eventually, charged hadrons are reconstructed, along with extra photons and neutral hadrons, which are identified by clusters without a corresponding track or by the excess of calorimeter energy with respect to the energy measured by linked tracks.

3.4.4 Muons

Muons are usually the easiest to be identified, since there are the only particles of passing through the whole detector leaving a signal at the muon chambers. Their reconstruction [49] commences with "local" reconstruction that involves the transformation of the muon subdetector signals to *muon hits*. Muon hits indicate the passage of a muon from a specific part of the detector. In contrast with the RPCs, DTs and CSCs are multi-layer detector and, as a result, hits come in groups within the same chamber. These groups form straight-line tracks called *segments*.

Muon track reconstruction follows. There are multiple types of muon tracks: *Tracker muon tracks* are created from tracks in the inner silicon tracker if they are matched with a muon segment. *Standalone muon tracks*

are tracks formed using only information from the muon chambers. Muon segments are used as seed to aggregate multiple hits into a track, to which a Kalman-filter technique is applied. Finally, *global muon tracks* are built "outside-in" by matching standalone muon tracks to tracker muon tracks. A Kalman-filter algorithm is used in this case as well to merge the two tracks in a combined one. 99% of the muon within the geometric acceptance of the muon chambers are reconstructed as one of the aforementioned muon track types.

Standalone muon provide the worse resolution among the three muon types and contain a non-negligible fraction of cosmic ray muons. Tracker and global muons profit from the excellent resolution of the silicon tracker. The fact that tracker muons are required to match to only one muon segment, which is most usually in the innermost muon station, makes them prone to moderate fake rates from punch-through hadrons, i.e. hadronic showers that manage to reach the edge of the muon subsystem. Global muons are designed to significantly decrease the rate of such misreconstructions.

In the PF algorithm, muons are subject to a series of selection criteria. These are related to track fit χ^2, number of hits per track, segment compatibility, compatibility with the primary vertex, etc. Muon isolation is another powerful handle for muon identification. For the computation of PF muon isolation, the energy contribution of charged and neutral particles other than the muon itself is measured by combining information from all subdetectors in a cone around it. Utilizing the selection criteria mentioned above, different identification types of muons are defined with varying efficiency and purity: Loose muon identification (ID), medium muon ID, tight muon ID, soft muon ID and high-momentum muon ID. Fig. 3.18 shows the reconstruction and identification efficiency for the the loose and tight muon IDs.

3.4.5 Electrons

Electrons are reconstructed by linking together energy deposits from the (electromagnetic) calorimeters and tracks from the inner tracker, with the added complication caused by the high energy loss due to radiation (bremsstrahlung) [59]. More specifically, an electron at $\eta \simeq 0(1.4)$ radiates away 33(86)% of its energy before reaching the ECAL, with the η variation of this fraction originating from amount of intervening material in different part of the detector. Due to this fact, specialized algorithms are used for the reconstruction of electron clusters and tracks.

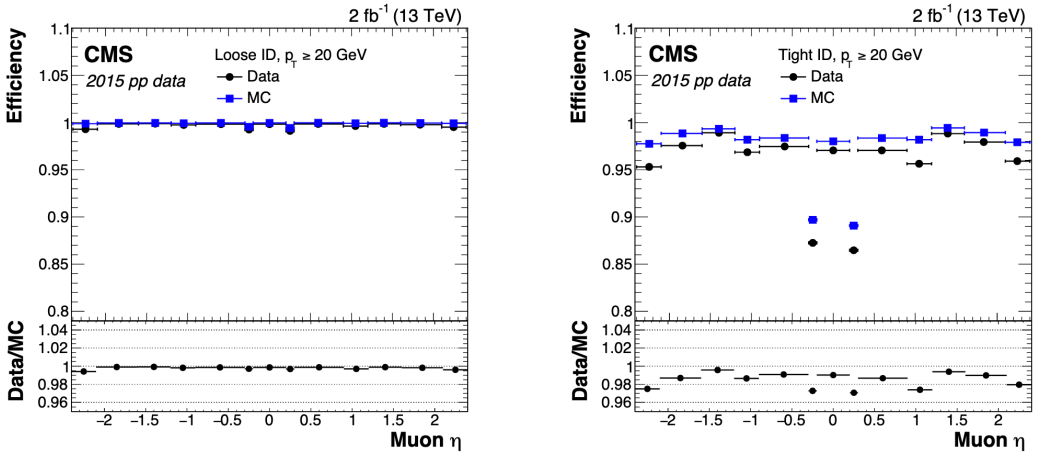


Figure 3.18: Reconstruction and identification efficiency for muons in 2015 data (black circles), simulation (blue squares) for loose (left) and tight (right) muon ID. Muons with $p_{\mathrm{T}} > 20\,\mathrm{GeV}$ have been used. The lower efficiency close to $|\eta| = 0.3$ corresponds to the connecting regions between the central muon wheel and the two neighboring wheels, where less instrumentation could be installed [49].

Accurately reconstructing electron clusters means collecting all of the clusters created by photon radiated away due to bremsstrahlung. These primarily spread in the ϕ direction, due to the direction magnetic field lines. There are two algorithms used for this purpose. In the barrel, the "hybrid" algorithm is implemented. Arrays of $\eta \times \phi = 5 \times 1$ crystals are accumulated around the seed crystal if their energy exceeds a specific threshold, for a specific number of steps or until no more arrays can be added. The extended cluster formed is called *supercluster* (SC). In the endcaps, the "multi-5 $\times$ 5" is used. Around the cluster seed, 5×5 crystals are checked and merged with the seed, if their sum of E_T exceeds a threshold value. The SC created with this procedure cannot be larger than 0.07 in the η direction and 0.3 in the ϕ direction. The energy of the preshower crystals within the SC range are also added in the SC. In both regions, the SC energy is the sum of the energies of all its clusters, while the SC position corresponds to the energy-weighted mean of the cluster positions.

To reconstruct the track formed by the passage of an electron in the silicon tracker requires dedicated methods to make sure that the efficiency and resolution stay good despite the large energy losses. Because this specialized

tracking is a time-consuming process, a specialized seeding of the tracks is also applied. Electron track seeds are found either by the extrapolating the position of ECAL SCs back to the beamspot or by identifying tracks in the standard track collection that are compatible with energy losses.

These seeds are used in a track finding procedure similar to the standard tracking one, as described in Sec. 3.4.1. However, for the electron tracks, the fit χ^2 requirements are relaxed in order to accommodate for trajectory deviations due to bremsstrahlung and a penalty is applied for missing hits in order to reduce combinations of genuine electron tracks with tracks from electron conversions. The tracks created this way are then subject to a final track fit. In contrast with standard tracks, the energy loss distribution of an electron is described by the Bethe–Heitler formula, which is non-Gaussian. As a consequence, the standard Kalman-filter techniques are not suitable for its description. A modified version of the Kalman filter is implemented, which approximates the Bethe–Heitler distribution with a mixture of multiple Gaussians. This method is called *Gaussian sum filter* (GSF) and greatly improves the track building performance for electrons. The left plot of Fig. 3.19 shows the GSF tracking electron reconstruction efficiency as a function of p_{T} for multiple η ranges.

The electron candidates are formed by the association of a GSF track and a cluster in the ECAL. In the case that the GSF track seed was found by the extrapolation of a SC back to the tracker, the corresponding ECAL cluster is simply the SC itself (ECAL-seeded electrons). In the case of tracker-based seeding, the GSF track is linked with a PF electron cluster. Ambiguities are resolved on the basis of missing inner hits, the E_{SC}/p ratio or the type of seed. Multiple techniques are implemented to increase the performance of charge and momentum estimation of electron candidates [60].

Genuine electrons are selected by applying requirements on variables that quantify the cluster-track association (e.g. cluster energy-track momentum compatibility), on calorimeter variables (e.g. the transverse shape of electromagnetic showers in the ECAL) and on GSF track variables (e.g. the χ^2 of the track fit). Apart from simple but robust cut-based strategies, multivariate (MVA) techniques are exploited to maximize the selection efficiency while minimizing the background rate. More specifically, a *boosted decision tree* (BDT) algorithm is implemented to combine the separation power of multiple variables and improve the discrimination between genuine (signal) and misreconstructed (background) electrons. For the same signal efficiency, the background rate is halved when the MVA-based selection is used instead

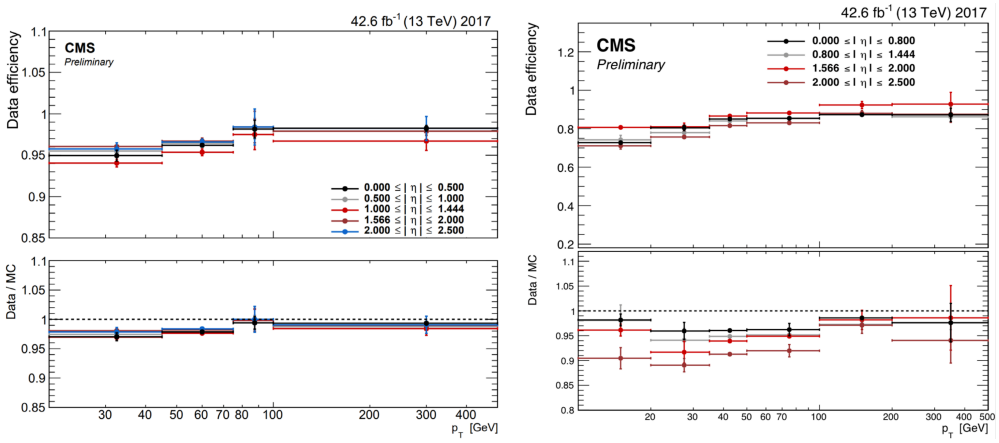


Figure 3.19: Left: Electron reconstruction (GSF tracking) efficiency. Right: Electron identification efficiency for the BDT ID working point of 90% of signal efficiency, including the cut on isolation. Both: Upper subplots show the efficiency in data and lower subplots show the data to MC efficiency ratio, in both cases as a function of the supercluster p_{T}, for different supercluster η ranges [60].

of the cut-based one. The efficiency of the MVA-based electron ID is shown as a function of p_{T} for multiple η ranges in the right plot of Fig. 3.19 for the working point that ensures 90% inclusive efficiency. Similarly to the muons, non-prompt electrons, e.g. those produced inside jets, can be rejected by taking advantage of the electron isolation.

3.4.6 Photons

The reconstruction of photons is very similar to that of electrons. In particular, the calorimeter clustering algorithms used for the creation of SC for electrons are applied for the creation of SC for photons [61]. The R_9 variable is defined as the energy sum of the 3×3 crystals centred on the most energetic crystal in the SC divided by the energy of the SC and can be exploited to separate converted photons from unconverted or late converted (i.e. converted at a radius > 85 cm) photons, as it can be seen in the left plot of Fig. 3.20. In the case of unconverted or late converted photons, there is no track that can be related to the SC. In the case of converted photons, the SC can be associated with the tracks of the electron pair, which must

fulfil special requirements, e.g. the fact that the photon is massless leads to the requirement that the electron tracks from conversions are almost parallel to each other. In both cases, the shower shape must be compatible with a photon signature. Multiple calibrations and corrections are applied on the photon energy and resolution, described in detail in [61]. After those, the resolution for unconverted photons is 1.5–2.5%($\sim$ 3.5%) and for converted photons is 2.0–4.0%($\sim$4.0%) in the barrel(endcap).

In terms of separation from electrons, a "conversion-safe electron veto" can be applied to the photons. According to this veto, photons are rejected if there is track with a hit in the inner layer of the pixel detector pointing to the photon SC that is not matched to a reconstructed conversion vertex. A more stringent veto, denoted as "pixel track seed veto", is also available. This veto trades lower efficiency for higher purity by rejecting photons whose SC is associated with a track initiated from a pixel track seed.

Neutral mesons, primarily π^0 decaying to two photons, usually pose the most important background in physics analyses using isolated photons in the final. To discriminate prompt photons (signal) against those originating from isolated neutral mesons or neutral mesons within jets (background), both a cut-based and an MVA-based strategy are utilized. Both methods profit from using shower shape and isolation variables to separate signal from background. The MVA-based method implements a BDT algorithm which manages to better capture and exploit non-trivial correlations among the input variables and achieves a better performance with respect to the cut-based one. The efficiency for MVA-based ID is shown in the right plot of Fig. 3.20.

3.4.7 Jets

Jets are physics objects designed to encompass the multitude of particles produced during the hadronization of quarks and gluons from the hard scattering. They are reconstructed by clustering the PF charged and neutral hadrons with the anti-kT algorithm [62, 63]. The algorithm defines "distances" between particles i, j and the beamline B as follows:

$$d_{ij} = \min\left(p_{Ti}^{-2}, p_{Tj}^{-2}\right) \frac{\Delta R_{ij}^2}{R^2} \tag{3.5}$$

$$d_{iB} = p_{Ti}^{-2} \tag{3.6}$$

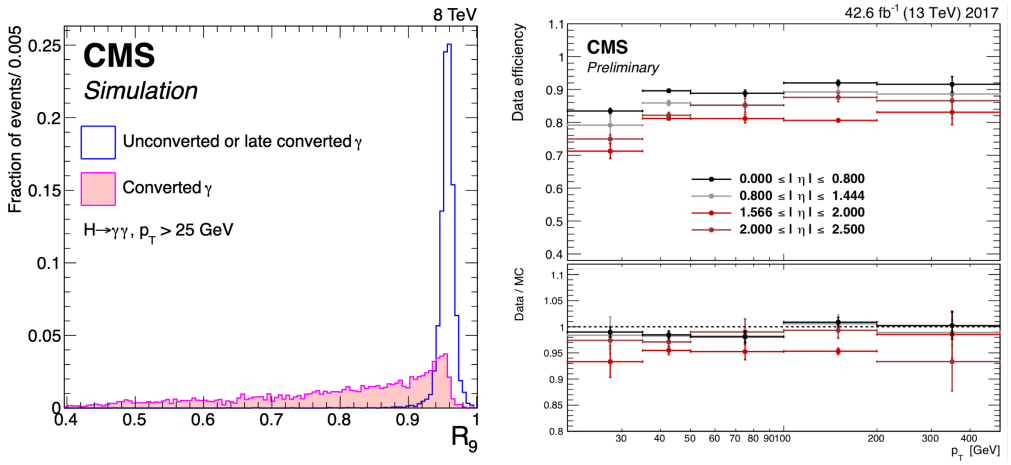


Figure 3.20: Left: Distributions of the R_9 variable for photons in the ECAL barrel that convert in the material of the tracker before a radius of 85 cm (converted γ) and those that convert later or do not convert at all before reaching the ECAL (unconverted or late converted γ) [61]. Right: Photon identification efficiency for the BDT ID working point of 90% of signal efficiency. The upper subplot shows the efficiency in data and the lower subplots shows the data to MC efficiency ratio, in both cases as a function of the supercluster p_{T}, for different supercluster η ranges [60].

where R is a distance parameter that controls the size of jets. Jets in CMS are usually reconstructed with $R = 0.4$ (*AK4 jets*), targeting showers from light quarks and gluons. "Fat" jets of $R = 0.8$ (*AK8 jets*) are also used when attempting to reconstruct Lorentz-boosted W, Z, and Higgs bosons or for t quark identification.

The distances defined above are calculated for all particles in an event. Starting from the smallest d_{iB} and the smallest d_{ij}, if $d_{ij} < d_{iB}$, particle j is combined with particle i into a single entity i and all the relevant distances are recalculated. The process is repeated until no more merging with entity i is possible. Entity i is now excluded from the collection of objects processed by the algorithm and is promoted to a jet. The procedure is repeated with the particle with the second smallest $d_{i'B}$ and so on, until no particles are left unchecked.

A series of corrections are applied to jets to ensure excellent response and resolution. Before the jet clustering mentioned in the previous paragraph, the

PF candidates undergo a PU cleaning procedure that is implemented either with the *charged hadron subtraction* (CHS) or with the *pileup per particle identification* (PUPPI) algorithm [64]. The reconstructed jets with the CHS- or the PUPPI-based PU mitigation are referred to as CHS and PUPPI jets, respectively.

After the jets have been formed, a multi-stage *jet energy correction* (JEC) procedure is initiated, as shown in Fig. 3.21. First, another correction against PU is applied (L1 correction). The *PU offset corrections* are extracted from dijet events in QCD simulation, generated with and without PU overlay. They are parametrized as a function of the jet p_{T}, the jet η, the PU energy density, ρ, and the jet area, A, with the aim of removing any data set dependence on luminosity. Apart from the corrections based on MC simulation, an extra correction, implementing the random cone (RC) method in zero-bias events, is applied to the data. The improvement in the jet p_{T} response achieved with this correction is demonstrated with the comparison of the left (before any JEC) and the middle (after the PU offset corrections) plots of Fig. 3.22. Next, the jet response in both data and simulation is ameliorated with correction factors determined by comparing the reconstructed jet p_{T} to the particle-level one in a QCD dijet sample (L2L3 corrections). The goal of these correction is to make the jet p_{T} and η response uniform. Residual non-uniformities in data are taken into account with the *L2L3 residual corrections*. These rely on the conservation of momentum in dijet and Z/γ+jets to correct any small differences of the jet response in data and MC. Optional corrections relating to the flavor of the jet are also obtained. The application of the above JECs to the reconstructed jets leads to the collection of the calibrated jets [65]. The excellent response of calibrated jets can be seen in the right plot of Fig. 3.22.

Dedicated algorithms have been developed to identify jets originating from the hadronization of b or c quarks [23]. In contrast with hadrons formed through the hadronization of light (u, d or s) quarks and gluons, hadrons containing the b(c) quark can have long lifetimes that reach up to 1.5(1) ps. Due to this fact, they tend to decay at a non-negligible distance from the PV that ranges from a few mm to one cm, depending on the hadron momentum. This gives rise to displaced tracks that form *secondary vertices* (SVs). The displacement of the SVs with respect to the PV is quantified by a number impact parameters, which measure the distance between the PV and the SV in different directions. The impact parameter value can be defined in three spatial dimensions, denoted as IP_{3D}, in the plane transverse

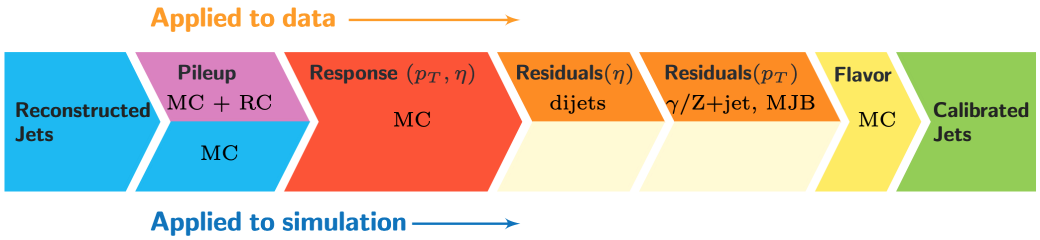


Figure 3.21: The different stages of JECs, applied from left to right, for data (upper row) and simulation (lower row). All corrections marked with "MC" are derived from simulation studies, "RC" stands for random cone, and "MJB" refers to the analysis of multijet events [65].

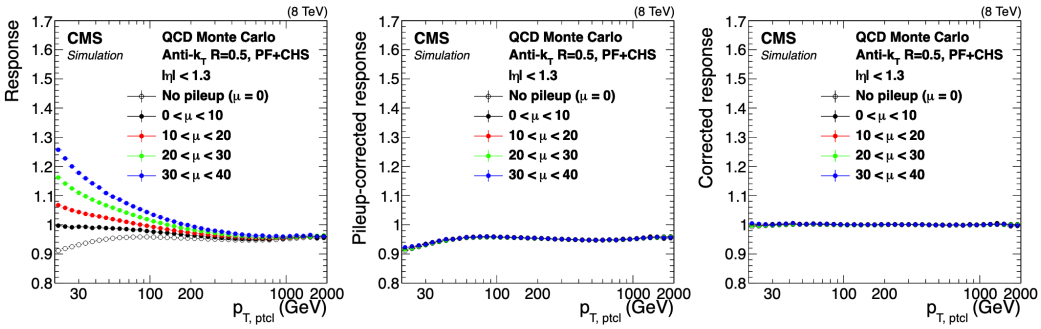


Figure 3.22: Average value of the measured jet p_{T} divided by the jet $p_{T,ptcl}$ at particle-level in QCD MC simulation, as a function of $p_{T,ptcl}$. The same plot is shown before any corrections (left), after only pileup offset corrections (middle), and after all JECs (right). μ stands for the average number of PU interactions per bunch crossing [65].

to the beam line, denoted as d_{xy} or along the longitudinal direction, denoted as d_z. The significance of those parameters is also defined, e.g. for the IP_{3D} variable: $SIP_{3D} = IP_{3D}/\sigma_{IP_{3D}}$, where σ_X stands for the uncertainty in the measurement of the variable X. b and c quarks also have larger masses than light quarks and gluons and have harder fragmentation. Additionally, their decay is usually characterized by the presence of a charged lepton. b and c tagging algorithms profit from all of the above properties of the respective quark by combining them in a neural network (NN) (CSVv2 algorithm), in a BDT (cMVAv2 algorithm) and a deep NN (DeepCSV algorithm). Comparison of the performance, in terms of b quark identification, of these heavy flavor tagging algorithms is shown in Fig. 3.23.

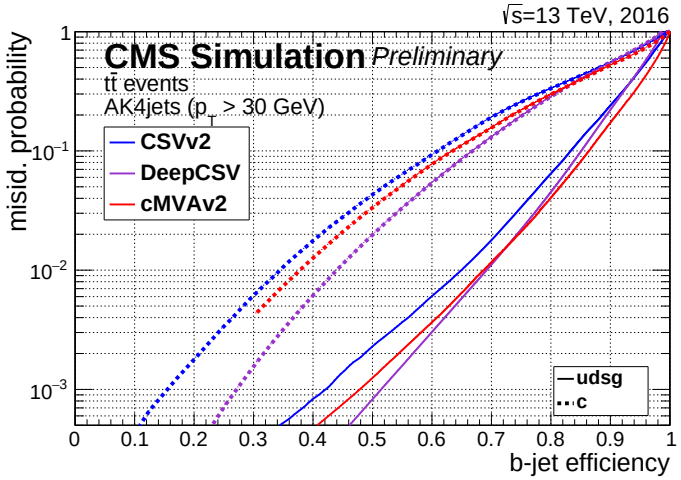


Figure 3.23: Performance of the b jet identification algorithms demonstrating the probability for a non-b jet to be misidentified as b jet as a function of the efficiency to correctly identify b jets. The curves are obtained on simulated $t\bar{t}$ events using jets within tracker acceptance with $p_T > 30\,\text{GeV}$ [66].

3.4.8 Hadronic τ Leptons

Due to its large mass, the τ is the only lepton that can decay to hadrons. Even though leptonic τ decays can be reconstructed as μ or e leptons through the algorithms mentioned in Secs. 3.4.4 and 3.4.5 respectively, hadronic τ decays require dedicated algorithms to be reconstructed. In approximately 2/3 of the cases, τ leptons decay hadronically and multiple decay modes are possible: τ decays to one charged hadron (primarily $\pi^{\pm}$), called *1-prong*, are the most common and they can also include up to two extra neutral hadrons (π^0). τ decays to three charged hadrons, called *3-prongs*, with up to one neutral hadron occur $\sim 15\%$ of the time. Any hadronic decays other than the ones mentioned above have $\sim 3\%$ branching fraction (BR) in total. All of the decays produce a ν_τ, which cannot be detected, and some proceed through an intermediate resonance (Table 3.2).

Hadronic τ are reconstructed in CMS with the *hadrons-plus-strips* (HPS) algorithm [67]. Starting from the constituents of reconstructed AK4 jets, the algorithm identifies the prongs of the τ decay from the PF charged hadron within the jet. These are required to have $p_T > 0.5\,\text{GeV}$ and to be loosely consistent with the PV of the event, so that PU contamination is minimal

Decay mode	Resonance	BR (%)
$\tau^{\pm} \to e^{\pm}\nu_e\nu_\tau$		17.8
$\tau^{\pm} \to \mu^{\pm}\nu_\mu\nu_\tau$		17.4
$\tau^{\pm} \to h^{\pm}\nu_\tau$		11.5
$\tau^{\pm} \to h^{\pm}\pi^0\nu_\tau$	$\rho(770)$	25.9
$\tau^{\pm} \to h^{\pm}\pi^0\pi^0\nu_\tau$	$a_1(1260)$	9.5
$\tau^{\pm} \to h^{\pm}h^{\mp}h^{\pm}\nu_\tau$	$a_1(1260)$	9.8
$\tau^{\pm} \to h^{\pm}h^{\mp}h^{\pm}\pi^0\nu_\tau$		4.8
Other hadronic		3.3

Table 3.2: Decay modes of the τ lepton with their branching fraction. $h^{\pm}$ denotes a charged hadron. When the decay occurs via an intermediate resonance, the corresponding meson is shown [67].

while allowing for longer lifetime τ lepton to be reconstructed. The neutral pions can be identified via their decay to a pair of photons, which also have a high probability of converting to a pair of electrons, when interacting with the material of the detector. As a result, the full energy of the neutral pions is contained in an elongated in the ϕ direction region, called a *strip*. Strips are formed in a dynamically adjusted $\Delta\eta \times \Delta\phi$ window, depending on the p_{T} of the e/γ objects to be clustered and of the e/γ objects already clustered. The prongs and strips are then combined to reconstruct the τ candidates of the $h^{\pm}$, $h^{\pm}\pi^0$, $h^{\pm}\pi^0\pi^0$ or $h^{\pm}h^{\mp}h^{\pm}$ decay modes (the $h^{\pm}h^{\mp}h^{\pm}\pi^0$ mode is not considered due to its large contamination by jets). All of the τ candidate constituents are required to fall within a p_{T}-dependent cone, their invariant mass must be consistent with the corresponding resonance, if the decay proceeds via one, and their total charge has to be ± 1.

Multiple algorithms have been designed for the identification of τ leptons. Isolation variables prove to be a strong handle for the separation of τ candidates from jets in cut-based selections [67]. BDT-based methods have been developed to discriminate τ leptons from electrons, muons and jets, combining a large number of dedicated variables in each case [67, 68]. Finally, superior performance in every discriminant is achieved with the use of deep NN (DNN) algorithm, called *DeepTau*, that takes as input almost every variable available connected to the τ candidate [69]. The improvement attained with the DeepTau algorithms compared to the BDT-based methods is shown in Fig. 3.24.

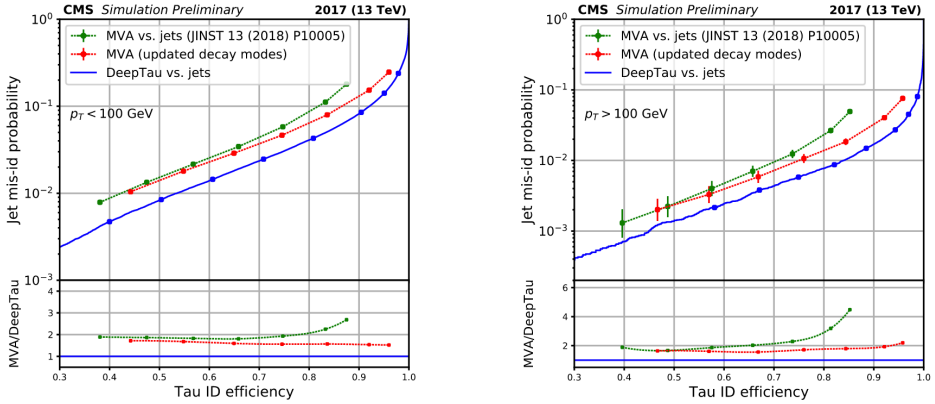


Figure 3.24: Jet misidentification probability of τ candidates as a function of the τ ID efficiency in a W+jets sample. The dots correspond to the working points chosen for the discriminators [69].

3.4.9 Missing Transverse Energy

As previously mentioned, no statement can be about the initial longitudinal momentum of the colliding partons, since they carry an unknown fraction of the proton momentum In contrast, it is known from the initial conditions of a collision that the total transverse energy is zero. As a result, the amount of *missing transverse momentum* ($p_{\mathrm{T}}^{\mathrm{miss}}$) and its direction can provide a useful handle for spotting the presence of weakly interacting particles, such as neutrinos, in the final state of a collision and extracting information about their properties.

The reconstruction of $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ in CMS is performed by taking the negative sum of the vector momentum of all PF reconstructed objects. This defines the PF $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$. A second algorithm, which applies the PUPPI method to the PF objects before adding up their momenta, is also used in order to minimize the PU dependence in the sum. This defines the PUPPI $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$. The simple but very robust PF $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ is used by most analyses in CMS, while the PUPPI $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ can be used in case where the PU effects are important [70].

Due to the inclusion of all event objects for its calculation, $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ is very sensitive to a number of detector effects. Potential sources of inaccurate estimation include the energy/momentum thresholds, non-linearities in response and inefficiencies of different subdetectors. To mitigate these effects, the JECs mentioned in Sec. 3.4.7 are propagated to the calculation of $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$:

$$\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}} = \vec{p}_{\mathrm{T}}^{\,\mathrm{miss,raw}} - \sum_{\mathrm{j}\in\mathrm{jets}} \left(\vec{p}_{\mathrm{T}}^{\,\mathrm{j,corr}} - \vec{p}_{\mathrm{T}}^{\,\mathrm{j}}\right) \tag{3.7}$$

After all corrections, the performance of $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ algorithms can be tested in Z/γ boson events. Taking the example of $\mathrm{Z} \to \mu\mu/\mathrm{ee}$ events, these have a single event axis and precise momentum scale, which can be accurately reconstructed due to the excellent reconstruction of leptons in the final state. By defining the p_{T} of the Z boson as $\vec{q}_T$ and the p_{T} of the hadronic recoil of the event as $\vec{u}_T$, with components $u_\parallel$, in the direction of $\vec{q}_T$, and $u_\perp$, in the perpendicular direction, the momentum conservation in the transverse plane imposes $\vec{q}_T + \vec{u}_T + \vec{p}_{\mathrm{T}}^{\,\mathrm{miss}} = 0$. A direct consequence of this is that the variable $-\langle u_\parallel \rangle / q_T$ gives an estimate of the response of the $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$, while the root mean square (RMS) of the $u_\parallel + q_T$, $\sigma(u_\parallel)$, and the $u_\perp$, $\sigma(u_\perp)$, are used for the estimation of the $u_\parallel$ and $u_\perp$ resolution, respectively. The performance of the PF $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ algorithm is shown in Fig. 3.25. The left plot shows the $p_{\mathrm{T}}^{\mathrm{miss}}$ response, which is close to unity for $q_T > 100\,\mathrm{GeV}$. The underestimation for values of $q_T < 100\,\mathrm{GeV}$ mainly comes from uncalibrated components of the $p_{\mathrm{T}}^{\mathrm{miss}}$, such jets with $p_{\mathrm{T}} < 15\,\mathrm{GeV}$. The middle and right plots show the $p_{\mathrm{T}}^{\mathrm{miss}}$ resolution parallel and perpendicular to q_T respectively. Due to the improved energy resolution at higher energies, the $p_{\mathrm{T}}^{\mathrm{miss}}$ resolution gets down to 13% for $u_\parallel$ and 9% for $u_\perp$ at $q_T > 200\,\mathrm{GeV}$.

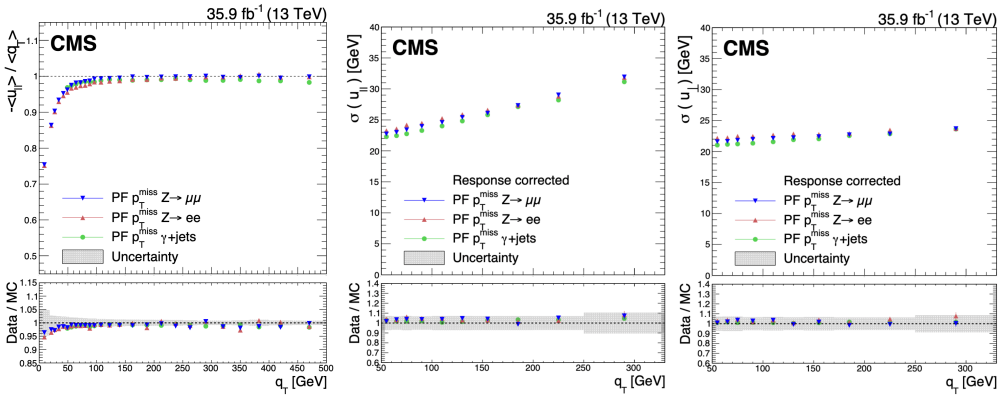


Figure 3.25: Left: Response of $p_{\mathrm{T}}^{\mathrm{miss}}$, defined as $-\langle u_{\parallel}\rangle/q_T$. Middle(Right): Resolution of the $u_{\parallel}(u_{\perp})$ components of the hadronic recoil. All: The variables are plotted as a function of q_T in Z $\rightarrow \mu\mu$/ee (blue/red)) and γ+jets (green) data events. The lower panels show the ratio of data to simulation, with the band corresponding to the total systematic uncertainties due to the JECs, estimated from the Z $\rightarrow$ ee sample [70].

Chapter 4

Search for supersymmetry in final states with two or three soft leptons and missing transverse momentum in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$

This chapter describes a search for SUSY particles in final states consisting of two oppositely charged or three low-p_T (soft) leptons and p_T^miss. The search is conducted at a centre-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$ with the full Run 2 data set recorded by the CMS experiment, corresponding to an integrated luminosity of up to $137\,\text{fb}^{-1}$. A similar search, targeting only final states with two leptons of opposite-sign (OS) charge, has been published by the CMS experiment with the p-p collision data recorded in 2016 [71]. The ATLAS experiment has also performed two searches with the data collected during the full Run 2, investigating SUSY signatures with two same-flavor, oppositely charged, soft leptons [72] and with three leptons [73]. The analysis presented in this chapter has been published in the Journal of High Energy Physics [74].

The chapter is organized as follows: Sec. 4.1 motivates the search and discusses the different SUSY scenarios that this physics analysis can probe.

Sec. 4.2 mentions the data and simulation samples used and Sec. 4.3 describes in detail the definition of the physics objects and the analysis strategy. The methods of background estimation are explained in Sec. 4.4, while Sec. 4.5 gives an account of all the systematic uncertainties that are applied. Finally, the results of the search and their interpretation to various simplified SUSY models are discussed in Sec. 4.6.

4.1 Motivation and Signal Models

As previously mentioned in Sec. 1.5, the SM exhibits signs of incompleteness, leading both the theoretical and the experimental high energy physics community to explore potential extensions for it. One of the most attractive theories beyond the SM (BSM) is SUSY, described in detail in Chapter 2. The experimental interest in SUSY is evident by a large number of searches that have been conducted with the aim of discovering particles compatible with it (Sec. 2.4). Focusing on R-parity conserving models, a common characteristic of all SUSY searches is the presence of large amounts of $p_{\mathrm{T}}^{\mathrm{miss}}$. This originates from the amount of energy that the neutral and stable LSP carries away while escaping the detector. The LSP is often accompanied by high p_{T} visible objects, hadronic jets and/or leptons[1], and these serve as a very effective signature for probing SUSY processes.

Fig. 4.1 shows the state of mass exclusion limits for $\widetilde{\chi}_2^0\widetilde{\chi}_1^{\pm}$ pair production in the middle of 2018, before this analysis was performed. The results come from multiple searches conducted by the CMS Collaboration. The figure shows that, for the vast majority of the SUSY parameter, from the line corresponding to $m_{\widetilde{\chi}_2^0} = m_{\widetilde{\chi}_1^0} + m_{\mathrm{Z}}$ to the right, the decay of EWKino particles to on-shell boson, possibly high p_{T} ones, is possible. These can subsequently decay to high p_{T} jets/leptons. This parameter space has been the obvious target of the majority of SUSY EWKino searches, due to the great discovery potential it provides while avoiding the experimental challenges that low p_{T} objects, coming from off-shell bosons, pose. However, as evident by Fig. 4.1 and the summary of results presented in Sec. 2.4, the lack of evidence for SUSY particles, especially in models with energetic particles in the final state, has resulted in strong constraints on their masses.

Given the exclusion of a wide range of SUSY particle mass combinations (masspoints) with a large mass difference (Δm) between the LSP and the

[1]In the following, the term "lepton" only refers to e and μ, not to τ leptons

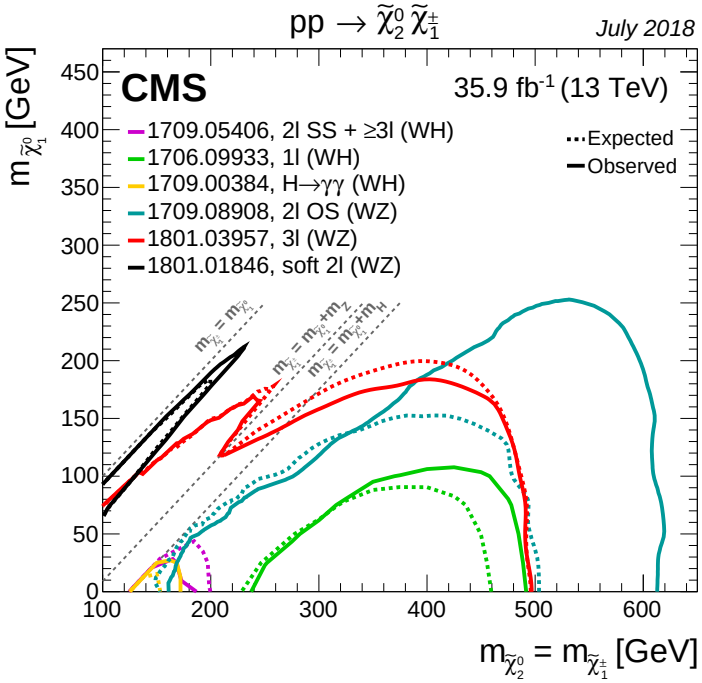


Figure 4.1: Results by CMS Collaboration for the lower mass limits, at 95% C.L., on $\widetilde{\chi}_2^0\widetilde{\chi}_1^\pm$ pair production with decay to multiple final states, as described in the plot legend. The plot shows the summary of results with only the 2016 data set, i.e. before the analysis presented in this thesis was conducted [21].

NLSP, Fig. 4.1 also demonstrates a different direction that SUSY searches can move towards. The parameter space between the $m_{\widetilde{\chi}_2^0} = m_{\widetilde{\chi}_1^0}$ line, which is the physical lower boundary of the parameter space, and the $m_{\widetilde{\chi}_2^0} = m_{\widetilde{\chi}_1^0} + m_{\mathrm{Z}}$ line is less well-explored, mainly due to the experimental difficulties it entails. Scenarios that predict SUSY particles in this part of the parameter space are often referred to as *compressed*. This characterization arises from the fact that, in such scenarios, the Δm between the LSP and the NLSP is small ($< 10\%$) compared to the mass of the NLSP, making Δm seem "compressed" with respect to the rest of the mass scales of the scenario.

In SUSY scenarios with compressed mass spectra, the available energy (equal to Δm) for the final state particles is small and the majority of it is carried away by the massive LSP. Still the $p_{\mathrm{T}}^{\mathrm{miss}}$ produced in the event

due to this is not significant enough to separate it from the bulk of the SM processes. As a result, an initial state radiation (ISR) jet is usually required to boost the final state objects, leading to moderate or even high amounts of $p_{\mathrm{T}}^{\mathrm{miss}}$. Despite this, the rest of the final state objects, which can be considered massless when compared to the LSP, tend to be soft, with p_{T} that does not exceed a few tens of GeV. Less than $\sim 10\,\mathrm{GeV}$ of p_{T} is at the limit of the object detection, reconstruction and identification capabilities of general purpose detectors. On top of that, low p_{T} physics is swamped with multijet processes from QCD, diminishing the sensitivity of hadronic searches. To decrease the huge QCD backgrounds, the presence of leptons is required.

The analysis presented in this thesis is explicitly designed to explore compressed SUSY models. The signature comprises of multiple soft leptons along with $p_{\mathrm{T}}^{\mathrm{miss}}$, induced by an ISR jet. The details on the physics object and event selection are given in Sec. 4.3. More specifically, except for the goal of pushing the sensitivity in the compressed regime to even higher NLSP masses, this analysis extends the low Δm acceptance of the previous CMS search in similar phase space [71]. An extension to even lower Δm, down to 1 GeV, targets to probe the thin slice of parameter space between the $m_{\widetilde{\chi}_2^0} = m_{\widetilde{\chi}_1^0}$ line and the black exclusion line in Fig. 4.1, while the addition of trilepton final states aims at covering the gap between the black and the red lines in the same figure.

Apart from the experimental incentive coming from the fact that SUSY might have eluded detection by hiding in a portion of the parameter space that is difficult to probe, the motivation for exploring the signatures of compressed SUSY models is solidly supported from the theoretical side as well. The theoretical motivation is different for different models and it is discussed for each one separately below. The analysis considers five signal models, which can be split in two major categories: The models that involve EWKino pair production and the models that include stop pair production. Although the salient features of those categories of compressed models remain the same, small differences between them lead to the creation of slightly different event selections for each category, as elaborated in Sec. 4.3. In all of the simplified models (SMS) below, the masses of all SUSY particles except for those explicitly mentioned, are set to very high values, being effectively decoupled from all the analysis observables.

In terms of EWKino pair production, the first model considered is the *wino-bino* SMS, denoted as TChiWZ. In this scenario, the bino mass parameter is only slightly smaller than the wino mass parameter ($M_1 \leq M_2$) and the

higgsino mass parameter is much larger than both of them ($M_1, M_2 \ll |\mu|$). According to the conclusions of Sec. 2.3, this mass parameter configuration leads to a neutralino LSP ($\widetilde{\chi}_1^0$) with mass $\sim M_1$ and a mass-degenerate pair of neutralino ($\widetilde{\chi}_2^0$) and chargino ($\widetilde{\chi}_1^\pm$) as the NLSPs, with masses $\sim M_2$, only slightly heavier. As a result, the LSP is mostly bino and the NLSPs are mostly winos, hence the name of the model. The mass configuration of the model is shown on the left of Fig. 4.2. The diagram of the process studied is shown on the right of Fig. 4.2. It is assumed that the pair of $\widetilde{\chi}_2^0$ $\widetilde{\chi}_1^\pm$ is produced with a pure wino production cross section (Table 4.1), computed at next-to-leading-order (NLO) plus next-to-leading-log (NLL) precision [75–77], and that the branching fraction of the decays $\widetilde{\chi}_2^0 \to \mathrm{Z}^*\widetilde{\chi}_1^0$ and $\widetilde{\chi}_1^\pm \to \mathrm{W}^*\widetilde{\chi}_1^0$ is 100%. The wino-bino model is motivated due to the fact that it can account for the observed DM relic density [78, 79]. Moreover, direct DM experiments lose sensitivity for the wino-bino model at $m_{\widetilde{\chi}_1^0} > 200\,\mathrm{GeV}$ with a blind spot at $m_{\widetilde{\chi}_1^0} \approx 350\,\mathrm{GeV}$ [80]. As a result, the masspoints of this model reached by collider experiments are relatively free from direct DM detection constraints and probing them at the LHC is of major importance.

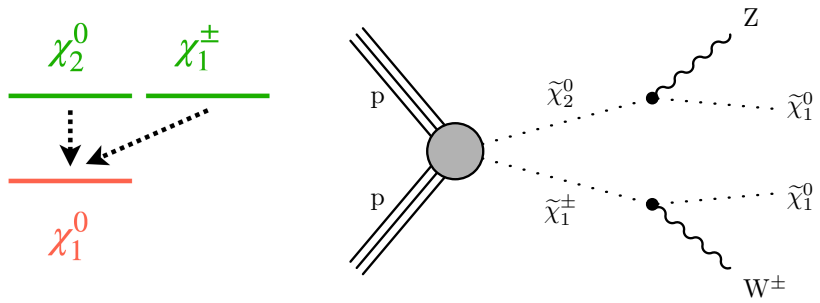


Figure 4.2: Left: Mass configuration of the non-decoupled particles of the TChiWZ SMS model. Right: Diagram of the decay of SUSY particles in the case of $\widetilde{\chi}_2^0$ $\widetilde{\chi}_1^\pm$ pair production, as included in the SMS under study.

The production of EWKinos can also be realized through neutralinos that are mostly higgsinos. This is possible if the higgsino mass parameter is smaller than the bino and wino mass parameters ($|\mu| < M_1, M_2$). With this mass configuration, the two lightest neutralinos, $\widetilde{\chi}_1^0$ and $\widetilde{\chi}_2^0$, as well as the lightest chargino, $\widetilde{\chi}_1^\pm$, have a mass that is approximately equal to $|\mu|$. The mass differences between them are of the order of $m_\mathrm{W}^2 / \min(M_1, M_2)$,

NLO-NLL wino-like $\widetilde{\chi}_2^0\,\widetilde{\chi}_1^\pm$ cross sections			
$m_{\widetilde{\chi}_2^0} = m_{\widetilde{\chi}_1^\pm}$ [GeV]	100	200	300
σ [fb]	22670.1	1807.39	386.936

Table 4.1: Pure wino-like $\widetilde{\chi}_2^0\widetilde{\chi}_1^\pm$ pair production cross sections in p-p collisions at 13 TeV for benchmark masspoints. The cross sections are computed at NLO plus NLL precision in a limit of mass-degenerate wino $\widetilde{\chi}_2^0$ and $\widetilde{\chi}_1^\pm$, light bino $\widetilde{\chi}_1^0$, and with all the other sparticles assumed to be heavy and decoupled [9].

which can be made arbitrarily small, depending on the values of M_1 and M_2. For the higgsino SMS, it is assumed that $m_{\widetilde{\chi}_1^\pm} = \frac{1}{2}(m_{\widetilde{\chi}_1^0} + m_{\widetilde{\chi}_2^0})$, with the mass configuration shown on the left of Fig. 4.2. Both the $\widetilde{\chi}_2^0\ \widetilde{\chi}_1^\pm$ (Fig. 4.2, right) and the $\widetilde{\chi}_2^0\ \widetilde{\chi}_1^0$ (Fig. 4.3, right) processes are considered for this model with pure higgsino cross sections (Table 4.1), computed at NLO+NLL precision [75–77]. The branching fraction of the decays modes shown is again set to 100%. For all the EWKino pair production SMS models, the Δm between the LSP and the NLSP is varied in the range 1–50 GeV.

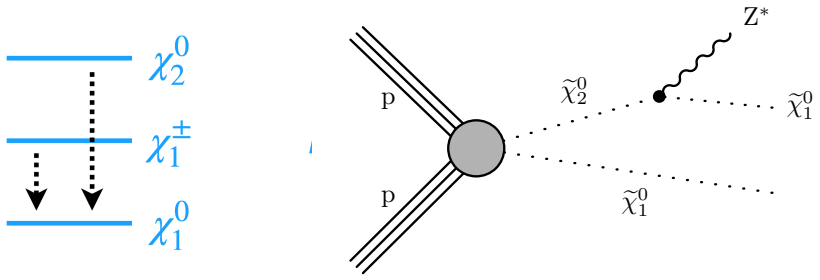


Figure 4.3: Left: Mass configuration of the non-decoupled particles of the higgsino SMS model. Right: Diagram of the decay of SUSY particles in the case of $\widetilde{\chi}_2^0\ \widetilde{\chi}_1^0$ pair production, as included in the SMS under study.

The case of light higgsinos is highly favored theoretically and, for higgsino models specifically, mass splittings of 5–10 GeV are a common occurrence in realistic SUSY models, as shown in Refs. [81–84]. The higgsino SMS is a useful interpretation due to its close mapping onto the physical observables measured. Nevertheless, it is highly interesting to understand the SUSY phase space covered by these realistic models. In general, this is a daunting

NLO-NLL higgsino-like $\widetilde{\chi}_2^0\,\widetilde{\chi}_1^\pm$ cross sections			
$m_{\widetilde{\chi}_2^0}$ [GeV]	100	150	200
σ [fb]	5325.95	1215.47	424.166

NLO-NLL higgsino-like $\widetilde{\chi}_2^0\,\widetilde{\chi}_1^0$ cross sections			
$m_{\widetilde{\chi}_2^0}$ [GeV]	100	150	200
σ [fb]	3277.01	715.14	244.213

Table 4.2: Pure higgsino-like $\widetilde{\chi}_2^0\,\widetilde{\chi}_1^\pm$ and $\widetilde{\chi}_2^0\,\widetilde{\chi}_1^0$ pair production cross sections in p-p collisions at 13 TeV for benchmark masspoints. The cross sections are computed at NLO plus NLL precision in a limit of mass-degenerate higgsino $\widetilde{\chi}_2^0$ and $\widetilde{\chi}_1^\pm$, and $\widetilde{\chi}_1^0$, with all the other sparticles assumed to be heavy and decoupled [9].

task, given the huge number of free parameters (120) that the MSSM introduces. However, by making a few simple assumptions, i.e. first and second-generation universality, no flavor-changing NC and no new sources of CP violation, the phenomenological MSSM (pMSSM) can be constructed [85], which reduces the number of free parameters down to 19. On top of that, for the Higgsino phenomenology, it turns out that only a small subset of those free parameters have a significant effect. In particular, only μ, M_1, M_2 and $\tan\beta$ have an important impact.

To approach the full theory, a pMSSM-inspired model with higgsino LSP is also studied. In the *higgsino pMSSM model*, the masses of the physical neutralino and chargino eigenstates are fully determined by the neutralino and chargino mass matrices, respectively. All EWKino production modes are allowed and the full cross section calculation is performed at NLO+NLL precision for each model masspoint separately via PROSPINO2 computational package [86]. The $\tan\beta$ parameter is fixed to 10, since its residual dependence is diminished for such large values. The gluino mass parameter M_3 is set to sufficiently high values to decouple from any physics observable, trilinear couplings are removed and the GUT bino-wino mass parameter relation is assumed, $M_1 = 0.5 \times M_2$, so that the parameter space is reduced to a μ-M_1 two-dimensional (2D) scan. The parameter space scanned is delimited by the relations: $100 < \mu < 240\,\text{GeV}$ and $0.3 < M_1 < 1.2\,\text{TeV}$. The dependence of fundamental quantities, such as the inclusive cross section (upper left), the $m_{\widetilde{\chi}_1^0}$ (upper right), $\Delta m(\widetilde{\chi}_2^0, \widetilde{\chi}_1^0)$ (lower left) and $\Delta m(\widetilde{\chi}_1^\pm, \widetilde{\chi}_1^0)$ (lower right), on

the 2D parameter space variables are shown in Fig. 4.4. Specialized computational tools [87–91] were utilized for the calculation of the particle mass spectrum and the decay rates.

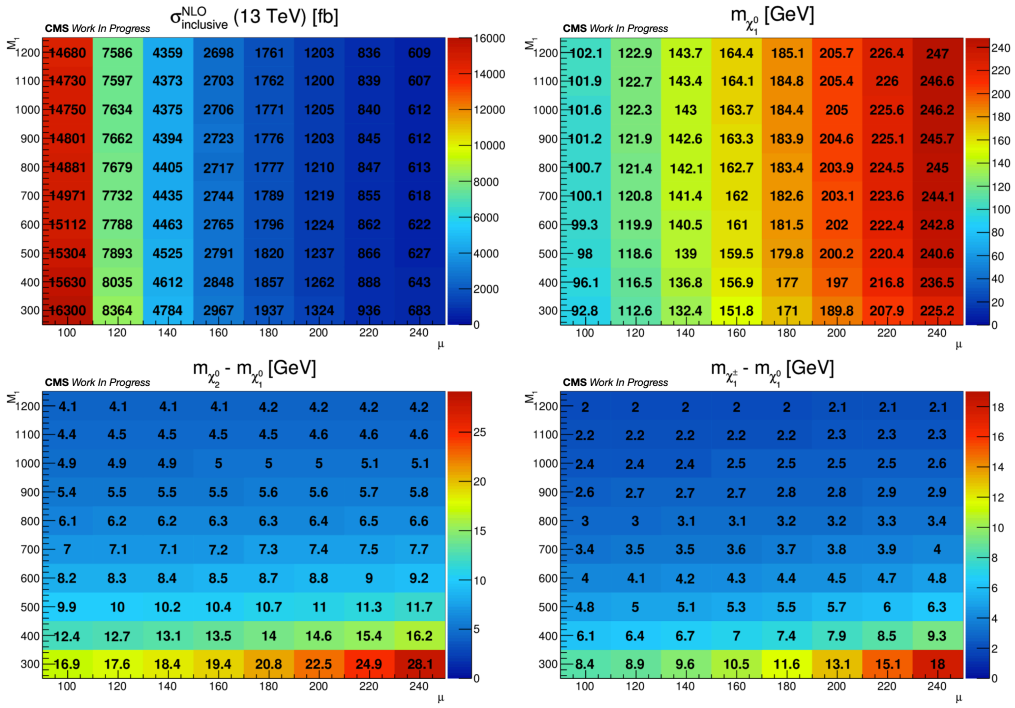


Figure 4.4: The inclusive cross section (upper left), the $m_{\widetilde{\chi}_1^0}$ (upper right), $\Delta m(\widetilde{\chi}_2^0, \widetilde{\chi}_1^0)$ (lower left) and $\Delta m(\widetilde{\chi}_1^{\pm}, \widetilde{\chi}_1^0)$ (lower right) for each masspoint of the higgsino pMSSM model.

The realization of SUSY compressed mass spectra is also possible in the case of stop pair production. In fact, the scenario of a light stop in combination with an almost mass-degenerate bino LSP is strongly motivated by the fact that this mass configuration creates the proper conditions for $\widetilde{\mathrm{t}}_1$-$\widetilde{\chi}_1^0$ co-annihilation, making the $\widetilde{\chi}_1^0$ LSP the dominant source of DM [92]. Depending on the exact mass configuration of the low mass particles (shown on the upper left of Fig. 4.5), two different compressed stop SMS are studied: In the case where no intermediate particle exists with mass between the stop and the LSP, the stop can only decay to the LSP directly. The stop undergoes the four-body decay $\widetilde{\mathrm{t}}_1 \to \mathrm{t}^*\widetilde{\chi}_1^0 \to \mathrm{bW}^*\widetilde{\chi}_1^0 \to \mathrm{bf}\bar{\mathrm{f}}'\widetilde{\chi}_1^0$, via a virtual top quark and a virtual W boson, as shown on the upper right of Fig. 4.5.

This model is denoted as T2bff$\widetilde{\chi}_1^0$. In the case where the $\widetilde{\chi}_1^\pm$ is lighter than the $\widetilde{\mathrm{t}}_1$, then the chargino-mediated decay, $\widetilde{\mathrm{t}}_1 \to \mathrm{b}\widetilde{\chi}_1^\pm \to \mathrm{bW}^*\widetilde{\chi}_1^0 \to \mathrm{bff}'\widetilde{\chi}_1^0$, is possible (lower of Fig. 4.5) and dominates in naturalness-inspired models. For this second model, denoted as T2bW, the mass of the chargino is set to $m_{\widetilde{\chi}_1^\pm} = \frac{1}{2}(m_{\widetilde{\chi}_1^0} + m_{\widetilde{\mathrm{t}}_1})$. For both models, the cross section of some benchmark masspoints is given in Table 4.3, the $\Delta m(\widetilde{\mathrm{t}}_1, \widetilde{\chi}_1^0)$ probed is between 10 and 80 GeV and the branching fractions of the decays described above are assumed to be 100%.

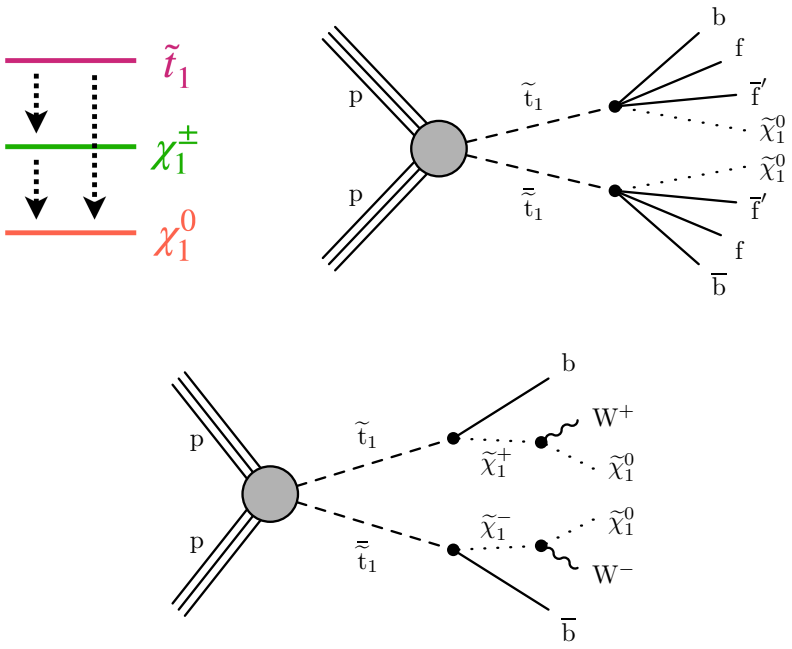


Figure 4.5: Upper left: Mass configuration of the non-decoupled particles of the T2bW SMS model. In the case of the T2bff$\widetilde{\chi}_1^0$ SMS model, the $\widetilde{\chi}_1^\pm$ is omitted from the diagram and only the direct $\widetilde{\mathrm{t}}_1 \to \widetilde{\chi}_1^0$ decay is possible. Upper right: Diagram of the decay of SUSY particles in the case of T2bff$\widetilde{\chi}_1^0$ SMS model. Lower: Diagram of the decay of SUSY particles in the case of T2bW SMS model.

To ensure the most accurate theoretical description, a couple of signal modelling refinements are applied to the analysis signal samples. The first refinement concerns the invariant mass of the leptons originating from the

NLO-NLL stop-antistop cross sections				
$m_{\widetilde{t}_1}$ [GeV]	300	400	500	600
σ [fb]	8516.15	1835.37	518.48	174.60

Table 4.3: Stop-antistop pair production cross sections in p-p collisions at 13 TeV for benchmark masspoints. The cross sections are computed at NLO plus NLL precision in the limit where all the sparticles, except for $\widetilde{t}_1$, $\widetilde{\chi}_1^0$ and $\widetilde{\chi}_1^\pm$, are assumed to be heavy and decoupled [9].

$\widetilde{\chi}_2^0 \to \widetilde{\chi}_1^0$ decay. As it is explained in detail in Sec. 4.3, this variable plays a central role in the analysis strategy, since it acts as a proxy to the Δm between the LSP and the NLSP in the EWK signal models. As a result, it shows great sensitivity in discriminating the SUSY signal from the SM backgrounds. Precisely reproducing its shape in simulation from the full theory is of major importance. It is shown in Ref. [93] that the differential decay rate of the $\widetilde{\chi}_2^0 \to \widetilde{\chi}_1^0$ decay depends on the structure of the couplings of the neutralinos to the Z boson:

$$\begin{aligned}\frac{d\Gamma_{\widetilde{\chi}_2^0 \to \widetilde{\chi}_1^0 \ell\bar{\ell}}}{dM(\ell\ell)} \propto\ & M(\ell\ell)\frac{\sqrt{M(\ell\ell)^4 - M(\ell\ell)^2(\mu^2+M^2)+(\mu M)^2}}{(M(\ell\ell)^2 - m_{\mathrm{Z}}^2)^2} \\ & \times [-2M(\ell\ell)^4 + M(\ell\ell)^2(2M^2-\mu^2)+(\mu M)^2]\end{aligned} \tag{4.1}$$

where $\mu = \tilde{m}_{\widetilde{\chi}_2^0} - \tilde{m}_{\widetilde{\chi}_1^0}$ and $M = \tilde{m}_{\widetilde{\chi}_2^0} + \tilde{m}_{\widetilde{\chi}_1^0}$ are constructed from the signed eigenvalues of the diagonalized neutralino mass matrix ($\tilde{m}$). Even after the values of the physical masses are set, the expression in Eq. 4.1 can still depend on the relative sign of the mass matrix eigenvalues and that leads to different dilepton invariant mass distributions, as shown in the left plot of Fig. 4.6. In particular, the relative sign of $\tilde{m}_{\widetilde{\chi}_2^0}$ and $\tilde{m}_{\widetilde{\chi}_1^0}$ can be either positive or negative in the wino-bino (TChiWZ) SMS, while it can only be negative in the higgsino models.

In the default simulated samples, the decay of EWKinos is handled by PYTHIA [94], which performs the calculation of the differential decay rate considering only the phase space component with a flat matrix element. To account for this simplification that alters the dilepton invariant mass distribution (Fig. 4.6, left), a reweighting procedure is applied. More specifically, for each scenario of $\tilde{m}_{\widetilde{\chi}_2^0} \times \tilde{m}_{\widetilde{\chi}_1^0}$ sign, numerical factors are applied on the dilepton invariant mass values computed by PYTHIA to correct its distribution.

The result of the reweighting procedure is shown in the right plot of Fig. 4.6, where the comparison between the theoretical prediction and the PYTHIA reweighted distribution shows excellent agreement for both scenarios.

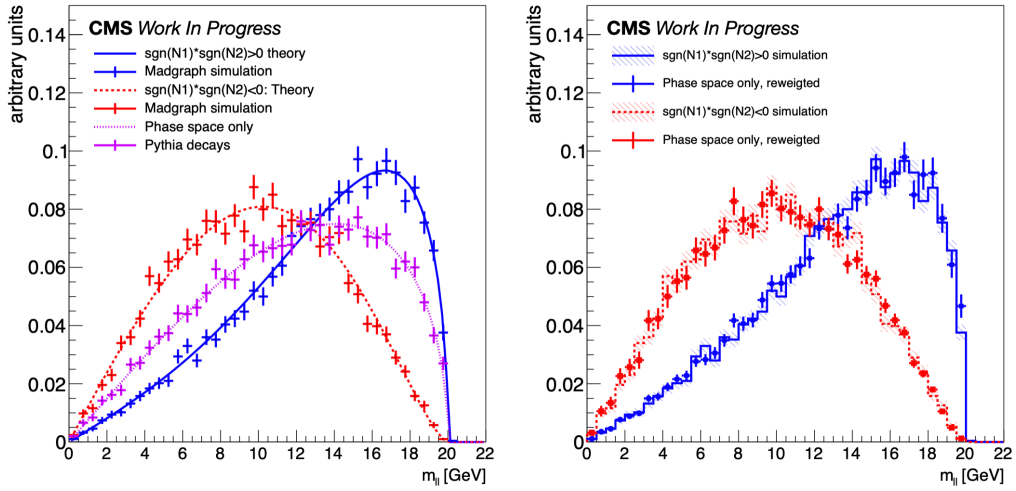


Figure 4.6: Left: Differential decay rate for the $\widetilde{\chi}_2^0 \to \widetilde{\chi}_1^0 \ell\bar{\ell}$ process as a function of the invariant mass of the daughter leptons. Lines correspond to theoretical calculations, while points correspond to the result of MC simulations, with statistical errors only shown. The differential decay rate is shown under the simplified flat matrix element assumption of PYTHIA (purple line) and with the inclusion of the full matrix element in the case of $\tilde{m}_{\widetilde{\chi}_2^0}\tilde{m}_{\widetilde{\chi}_1^0} > 0$ (blue) and in the case of $\tilde{m}_{\widetilde{\chi}_2^0}\tilde{m}_{\widetilde{\chi}_1^0} < 0$ (red). Right: Comparison between the simulation of the full theory (lines) and re-weighted simulation (points). Both cases using the full matrix element are shown: $\tilde{m}_{\widetilde{\chi}_2^0}\tilde{m}_{\widetilde{\chi}_1^0} < 0$ (blue) and $\tilde{m}_{\widetilde{\chi}_2^0}\tilde{m}_{\widetilde{\chi}_1^0} < 0$ (red).

As a result of the low mass difference of the lightest SUSY particles of the signal models considered, the virtual Z* and W* vector bosons mediating the decay of the NLSP to the LSP can be very off-shell. When their mass is low, smaller than $\sim 20\,\text{GeV}$, some of their decay modes can be heavily suppressed due to the no-longer negligible mass of the decay products. In order to include this effect in the signal modelling, a reweighting procedure of the branching fraction of the Z* and W* boson decays has been implemented. The correction factors to the default simulated samples have been extracted by comparing the branching fractions, as computed in the TCHIWZ sample, with the branching fractions calculated in 3-body decays of $\widetilde{\chi}_2^0$ and $\widetilde{\chi}_1^\pm$ at

tree-level with the SDecay module of SUSYHIT 1.5a [90]. The improvement comes from the fact that the SUSYHIT calculation takes into account second and third generation fermion masses. Loop-induced radiative decays of the $\widetilde{\chi}_2^0$ have been omitted. It is worth mentioning that nonpertubative QCD effects of the $\widetilde{\chi}_1^\pm \to \widetilde{\chi}_1^0 \mathrm{q}\mathrm{q}'$ decay become important starting at $\Delta m(\widetilde{\chi}_1^\pm, \widetilde{\chi}_1^0) \sim 1.5\,\mathrm{GeV}$, hence overestimating the branching fraction of $\mathrm{W}^* \to \ell\nu$ decays for these mass differences [95]. However, assuming that the same effect is true for the $\widetilde{\chi}_2^0$ and given that the sensitivity of this analysis reaches down to $\Delta m(\widetilde{\chi}_2^0, \widetilde{\chi}_1^0) = 3\,\mathrm{GeV}$ (Sec. 4.6), this effect can be safely neglected. Fig. 4.7 shows the comparison between the SUSYHIT and TCHIWZ SMS branching fraction for the Z^* (left) and W^* (right) decays to quantify the magnitude of the refinement in their branching fractions.

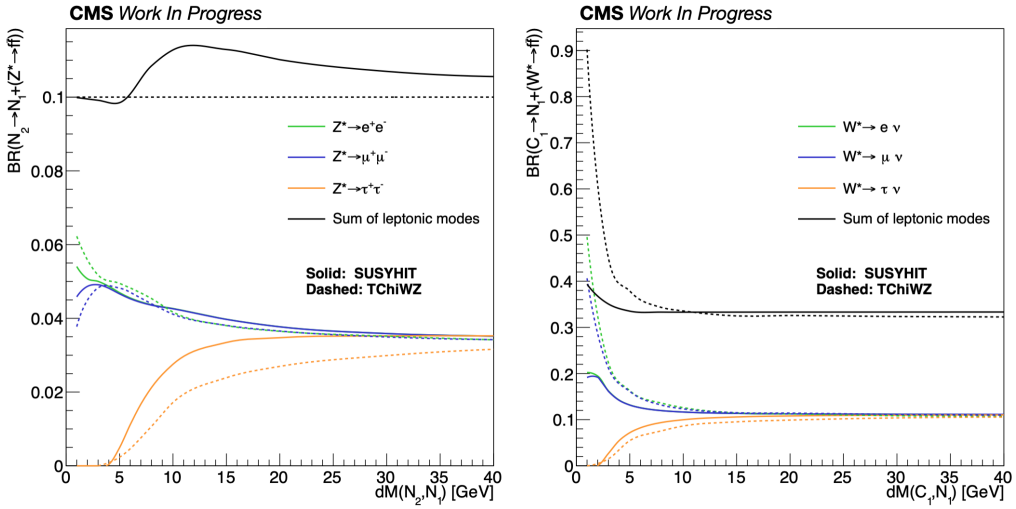


Figure 4.7: Left: Branching fraction of the $\widetilde{\chi}_2^0 \to \widetilde{\chi}_1^0 \mathrm{Z}^* \to \widetilde{\chi}_1^0 \mathrm{f}\bar{\mathrm{f}}$ decay versus $\Delta m(\widetilde{\chi}_2^0, \widetilde{\chi}_1^0) = m_{\mathrm{Z}^*}$. Right: Branching fraction of the $\widetilde{\chi}_1^\pm \to \widetilde{\chi}_1^0 \mathrm{W}^* \to \widetilde{\chi}_1^0 \mathrm{f}\nu_{\bar{\mathrm{f}}}$ decay versus $\Delta m(\widetilde{\chi}_1^\pm, \widetilde{\chi}_1^0) = m_{\mathrm{W}^*}$. Both: Comparison between the branching fraction modulations computed with the SUSYHIT package (solid line) and in the TCHIWZ SMS (dashed line), in which the $\widetilde{\chi}_1^\pm$ and $\widetilde{\chi}_2^0$ decays are computed by PYTHIA.

4.2 Data and Simulated Samples

The analysis relies on data and simulated samples produced centrally by CMS to extract its final results. This section describes the computational packages and configurations used for the production of the simulated samples (broken down to individual processes), as well as the trigger algorithms applied to record the data.

Starting from the signal samples, these have been generated with the MadGraph5_aMC@NLO [96, 97] event generator. The precision in perturbative QCD used for these samples is at leading order (LO), while the consolidation of additional partons from the event generator with parton shower generator is achieved with the MLM merging scheme [98]. The same setup has been used for the production of the major background processes of the analysis, which include $t\bar{t}$, DY, W+jets and Z+jets. The generation of the WZ process, which is dominant in trilepton final states, is performed at NLO with POWHEG v2.0 [99–104] event generator. The rest of diboson processes (WW, ZZ and Wγ) are generated with MadGraph5_aMC@NLO at NLO precision but using the FxFx merging scheme [97]. Other background processes that do not contribute significantly to the analysis yields, such as $t\bar{t}W$, $t\bar{t}Z$, WWW, ZZZ, WZZ or WWZ, are also produced using MadGraph5_aMC@NLO at NLO precision. Exceptions to this are the background processes from single top quarks in association with a W boson which are generated at NLO precision with POWHEG v1.0 [105]. The MadGraph5_aMC@NLO versions used are the 2.2.2 (2.3.2 for $t\bar{t}Z$) for 2016 and 2.4.2 (2.6.5 for $t\bar{t}Z$) for 2017 and 2018.

Different LO and NLO (depending on the generation details of each sample) parton distribution functions (PDFs) have been used for samples of different years: The 2016 simulated samples use the NNPDF3.0 [106] PDF set, while the NNPDF3.1 [107] PDF set is implemented for the 2017 and 2018 simulated samples. Similarly, the underlying event tune of the PYTHIA 8.212 package utilized for showering, hadronization and the underlying event description differs from year to year: The CUETP8M1 [108, 109] and the CP5 [110] underlying event tune are applied for 2016 and 2017/2018 respectively.

The simulation of the CMS detector is performed in the best precision possible with the Geant4 [111] package. Whereas the full detector simulation (FullSIM) is used for the majority of simulated samples, the T2bff$\widetilde{\chi}_1^0$ SMS, the T2bW SMS and the pMSSM higgsino samples have been produced

utilizing a fast detector simulation (FastSIM) [112, 113].

Three different trigger algorithms (paths) have been utilized for collecting the data used in this search. The main trigger path relies exclusively on the missing transverse energy (MET). To recover data events with moderate $p_{\mathrm{T}}^{\mathrm{miss}}$, a second trigger path has been designed specifically for this analysis. In combination with $p_{\mathrm{T}}^{\mathrm{miss}}$, this trigger path also requires a pair of soft muons. As a consequence, the extra dimuon requirement permits the relaxation of the $p_{\mathrm{T}}^{\mathrm{miss}}$ threshold to lower values. Finally, a complementary trigger path relying only on a dimuon pair with higher p_{T} threshold is used in the cases where the analysis cuts require it. Table 4.4 lists the trigger paths used in each year along with their corresponding luminosity. The luminosity is reduced for some trigger paths, because of technical reasons that caused their unavailability for short time periods. A discussion on the requirements of each trigger path and on the corrections applied to each on of them due to data-MC differences is included in the following subsection.

Year	HLT path	Luminosity
2016	`HLT_DoubleMu3_PFMET50`	$33.2\,\mathrm{fb}^{-1}$
	`HLT_PFMETNoMu120_PFMHTNoMu120_IDTight`	$35.9\,\mathrm{fb}^{-1}$
	`HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ`	$35.9\,\mathrm{fb}^{-1}$
	`HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL_DZ`	$35.9\,\mathrm{fb}^{-1}$
2017	`HLT_DoubleMu3_DZ_PFMET50_PFMHT60`	$36.7\,\mathrm{fb}^{-1}$
	`HLT_PFMETNoMu120_PFMHTNoMu120_IDTight`	$41.5\,\mathrm{fb}^{-1}$
	`HLT_PFMETNoMu120_PFMHTNoMu120_IDTight_PFHT60`	$41.5\,\mathrm{fb}^{-1}$
	`HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass3p8`	$36.7\,\mathrm{fb}^{-1}$
2018	`HLT_DoubleMu3_DZ_PFMET50_PFMHT60`	$59.2\,\mathrm{fb}^{-1}$
	`HLT_PFMETNoMu120_PFMHTNoMu120_IDTight`	$59.7\,\mathrm{fb}^{-1}$
	`HLT_PFMETNoMu120_PFMHTNoMu120_IDTight_PFHT60`	$59.7\,\mathrm{fb}^{-1}$
	`HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass3p8`	$59.7\,\mathrm{fb}^{-1}$

Table 4.4: List of HLT paths used in the analysis.

4.2.1 Triggers

As in the majority of SUSY analyses, the trigger strategy for this search involves capturing events based on the amount on $p_{\mathrm{T}}^{\mathrm{miss}}$. However, $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ is a complicated object, calculated by the (vectorial) sum of all the other objects

in the event, and measuring it at trigger level involves some subtleties. For the L1T, the signals from different subdetectors are not combined. As a result, the L1 MET is measured only from the calorimeters, without the inclusion of the muons (which are, in general, only detected in the muon chambers) in the calculation. The $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$, without the inclusions of the muons, is denoted as $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss,corr}}$ and it is used through the analysis, even offline, since it guarantees compatibility with the trigger variable. The term "MET" will be used to denote $p_{\mathrm{T}}^{\mathrm{miss}}$ and/or $p_{\mathrm{T}}^{\mathrm{miss,corr}}$, when there is no possibility of confusion between the two terms.

Starting from the trigger paths relying solely on MET, only pure $p_{\mathrm{T}}^{\mathrm{miss,corr}}$ trigger paths were chosen. This choice simplifies the offline selection and permits full coverage of the $p_{\mathrm{T}}^{\mathrm{miss}}$-$p_{\mathrm{T}}^{\mathrm{miss,corr}}$ phase space. The trigger paths with the lowest threshold covering the full luminosity of each year were used, backed up (in 2017 and 2018) with trigger paths with an extra H_{T} requirement at 60 GeV. The online $p_{\mathrm{T}}^{\mathrm{miss,corr}}$ threshold is, therefore, set at 120 GeV for all years.

For each trigger path, the efficiency for an offline selected event to also be selected online is measured. The efficiency of a specific quantity (e.g. $p_{\mathrm{T}}^{\mathrm{miss,corr}}$) of a trigger path is measured as a ratio of number of events: The denominator includes the events with the full collection of the trigger path requirements except for the one on the quantity measured, while the numerator counts the events of the denominator that also fulfil the extra requirement of having fired the trigger path. In all the cases, the efficiency is measured in both data and simulated background events arising from a mixture of DY, $\mathrm{t\bar{t}}$ and diboson processes.

For the pure $p_{\mathrm{T}}^{\mathrm{miss,corr}}$ trigger paths, due to the limited resolution, especially at the L1 trigger, an approximately stable efficiency plateau is reached at the offline value of $p_{\mathrm{T}}^{\mathrm{miss,corr}}$ of $\sim$200 GeV compared to the 125 GeV online threshold. This is what drives the offline $p_{\mathrm{T}}^{\mathrm{miss,corr}}$ requirements of the analysis (more details in Sec. 4.3.4). More details on the trigger efficiency of the pure $p_{\mathrm{T}}^{\mathrm{miss,corr}}$ trigger paths are given in Appx. A.1.

As already mentioned, a particularity of this search is that the signal final state can often produce only moderate amounts of MET. As a consequence, there is sensitivity to be gained even at offline $p_{\mathrm{T}}^{\mathrm{miss,corr}}$ values lower than 200 GeV. To retrieve the efficiency at those $p_{\mathrm{T}}^{\mathrm{miss,corr}}$ values, which is impossible with the pure $p_{\mathrm{T}}^{\mathrm{miss,corr}}$ trigger paths, a custom double-μ+MET trigger path was developed. The L1 seed of this trigger path requires two muons with $p_{\mathrm{T}} > 3$ GeV and online $p_{\mathrm{T}}^{\mathrm{miss,corr}} > 50$ GeV. The requirements were made

stricter for 2017 and 2018 by requiring one jet with $p_{\mathrm{T}} > 60\,\mathrm{GeV}$ or two jets with $p_{\mathrm{T}} > 30\,\mathrm{GeV}$, in order to reduce the online rate without any impact on the offline analysis. The double-μ+MET HLT path features extra requirements to control the trigger rate: The muons are required to be of opposite charge, to have a combined dimuon transverse momentum, $p_{\mathrm{T}}(\ell\ell)$, greater than 3 GeV and an invariant dimuon mass in the range $4 < M(\ell\ell) < 56\,\mathrm{GeV}$. The *distance of closest approach* (DCA) of the muon pair, which is equal to the minimum three-dimensional (3D) distance between the two muon tracks, is also set to be less than 0.5 cm. In 2017 and 2018, the DCA criterion was replaced by the requirement that the minimum distance along the z-axis (Δz) between the two muon tracks is less than 0.2 cm, which was found to be more efficient for triggering prompt muons. Finally, the HLT path sets the lower online $p_{\mathrm{T}}^{\mathrm{miss}}$ threshold at 60 GeV for all years, while a lower threshold of the online $p_{\mathrm{T}}^{\mathrm{miss,corr}}$ at 60 GeV is also required in 2017 and 2018, for compatibility with the L1 seed.

The efficiency measurement is performed separately for the different aspects of the double-μ+MET trigger path: The efficiency of the requirements pertaining the muons, primarily the p_{T}, DCA or Δz and $M(\ell\ell)$ requirements, is relatively low, around 50% for 3 GeV muons in the endcap but quickly increases to reach a plateau of $\sim 90\%$ for muons with $p_{\mathrm{T}} > 10\,\mathrm{GeV}$ over the whole η range. The hadronic part of the efficiency, which accounts for the $p_{\mathrm{T}}^{\mathrm{miss}}$ and $p_{\mathrm{T}}^{\mathrm{miss,corr}}$ requirements, reaches a plateau of more than 90% up to 100% for $p_{\mathrm{T}}^{\mathrm{miss}}, p_{\mathrm{T}}^{\mathrm{miss,corr}} > 125\,\mathrm{GeV}$. Appx. A.2 explains in greater detail the specifics of the double-μ+MET trigger path efficiency measurement.

In all of the efficiency measurements above, potential small differences between the data and the simulation performance are taken into account by applying correcting scale factors (SFs) on the simulation offline. These SFs are extracted as the ratio of the data efficiency over the simulation efficiency for each year separately and are a function of the variables that the efficiency measurements are performed on (Figs. A.1-A.11 in Appx. A). This is the case for the FullSim samples, which include the emulation of the trigger. In contrast, the trigger emulation is not applied on the FastSIM samples, so the raw data efficiency of each measurement above is applied.

Finally, the complementary pure double-μ trigger paths were found to have an efficiency close to unity for both data and MC. Given this and the fact that they have extremely limited usage in the analysis, no SFs are applied for these trigger paths.

4.2.2 Data Quality Issues

During the data taking period of Run 2 (2016–2018), problems with the CMS detector caused issues to the quality of the data. All of these issues have been checked for their impact on the analysis yields. The issues along with the checks performed are described below.

At the beginning of the 2018 data taking period, the timing of ECAL L1 trigger primitives was found to have gradually shifted towards early values in the data of 2016 and 2017. This effect, called *L1T prefire issue*, caused a significant portion of e/γ objects at $|\eta| > 2.5$ to be associated with a previous bunch crossing. Corrections to account for the issue have been derived for the 2016 and 2017 data, while no corrections are needed for the 2018 data set, since the issue was fixed before 2018 data taking started.

Although this search does not use forward ($\eta > 2.5$) e/γ objects directly, it can still be affected by the L1T prefire issue. This comes from the fact that L1 rules do not allow for consecutive bunch crossings to fire. As a consequence, if an event of interest for this analysis, that would normally be selected by one of the L1 seeds described in the previous section, contains a forward e/γ object with $E_T > 30\,\text{GeV}$, then there is a chance that it is not recorded, because the previous bunch crossing would have prefired and the L1 rule would have skipped the bunch crossing of interest. Moreover, it is possible that ECAL trigger towers associated to a high p_T jet in the forward region prefire and lead to an isolated ECAL deposit in bunch crossing -1, leading to the same outcome as above.

To correct for the indirect effect of the L1T prefire issue on the analysis, a set of *prefire weights* have been centrally computed by the CMS experiment and have been applied on the simulation yields of this analysis. By comparing the analysis yields pre- and post-applying the prefire weights, the impact of the issue on the analysis was estimated: The yields of 2016 changed by 1–2%, while the average change was slightly larger in general in 2017, reaching up to $\sim$4% in a couple of cases.

In 2017, especially at the end of the data taking period, a large enhancement of the instrumental MET tail was observed in the data. This translated to large disagreements between data and simulation for $p_\text{T}^\text{miss} \gtrsim 100\,\text{GeV}$. The effect was visible both at the trigger level (L1 and HLT) and offline. Its cause was tracked down to ECAL ageing at the endcap region, which lead to increased noise in the $2.65 < |\eta| < 3.139$ range.

All analyses exploiting MET for the extraction of their results were af-

fected by the issue and this is particularly true for this search, since one of its major prompt backgrounds, DY, mostly contains instrumental MET. To mitigate the issue, a special recipe was devised centrally by CMS for the computation of MET in 2017. More specifically, for 2017 only, MET (both the uncorrected one and the JECs involved in its correction) does not include jets and unclustered PF candidates with $p_{\mathrm{T}} < 50\,\mathrm{GeV}$ in the problematic η region. The resulting MET is referred to as *EE fixed MET*. The data-simulation agreement was significantly improved after the application of the EE fixed MET (plots on the data-MC agreement in control regions in Sec. 4.4). In the following, the EE fixed MET is used in all of the 2017 analysis.

In the summer of the 2018, a power interruption, resulting from a false fire alarm, caused a 40° section on one side of the endcap HCAL to permanently lose power and to no longer be operable for the rest of the data taking period. The incident is called *HEM issue* and a central recipe was designed to be applied on the post-HEM data to reject spurious electron, photon and jet objects that could arise in the affected area. The recipe involved vetoing events with any electron with $p_{\mathrm{T}} > 30\,\mathrm{GeV}$, $-3.0 < \eta < -1.4$ and $-1.57 < \phi < -0.87$. In addition to that, events were vetoed if they contained a jet in the same detector region, enlarged by half the default jet cone, i.e. if any jet with $p_{\mathrm{T}} > 30\,\mathrm{GeV}$ and $\Delta\phi(\mathrm{jet}, p_{\mathrm{T}}^{\mathrm{miss}}) < 0.5$ is present in the region $-3.2 < \eta < -1.2$ and $-1.77 < \phi < -0.67$.

The impact of the HEM issue was evaluated on this search by checking the distribution of forward ($\eta > 2.4$) jets for regions with abnormally increased activity and by checking the yield difference when applying the recipe prescribed above. To fully investigate the issue, the aforementioned vetos were made specific to this analysis, which uses soft electrons, by loosening the electron veto p_{T} threshold down to 5 GeV. Only a very slight excess of electrons was observed in the η-ϕ region affected by the issue, amounting to percent level difference. The application of the vetos did not improved the data-MC agreement and, as a result, the HEM issue was not applied on the analysis, following the central recommendations.

4.3 Physics Objects and Event Selection

The analysis attempts to exploit the kinematical differences between the SUSY signal and the background SM processes by applying a selection tai-

lored to soft leptons, as explained in Secs. 4.3.1 and 4.3.2. Jets, b-tagged jets and the MET (outlined in Sec. 4.3.3) are also utilized to increase the signal and background separation. Finally, the signal-rich regions of the analysis are designed, optimized for maximum sensitivity in the compressed Δm regime, requiring a set of selection criteria devised to specifically reject the dominant, reducible backgrounds (Sec. 4.3.4).

4.3.1 Muons

The muons used in this analysis are either global or tracker muons (defined in Sec. 3.4.4). Two muon collections are formed: The *tight ID* muons pass kinematical requirements that muons coming from the SUSY signature are expected to have. These muons are the ones used in all the regions of the analysis, unless stated otherwise. A *loose ID* muon category is also designed. As the name implies, this category contains muons that have, in general, less strict kinematical requirements and they are utilized in the nonprompt background estimation, as described in detail in Sec. 4.4.1.

To facilitate its optimization, the loose ID is factorized in two terms: A muon is categorized as a loose muon if it passes the tight ID or the *loose-not-tight ID*. The loose-not-tight ID is not a separate ID but rather a set of extra requirements that a muon not passing the tight ID must fulfil for it to be considered loose. In mathematical terms, the loose ID is defined as:

$$\text{loose ID} = \text{tight ID} \;||\; \text{loose-not-tight ID} \tag{4.2}$$

For convenience, the selection criteria of the loose-not-tight ID are given. These, along with the tight ID requirements are summarized in Table 4.5 and are explained below.

Given that this analysis targets soft leptons, the lower p_{T} bound for muons is set as low as 3.5 GeV to maximize the acceptance of very compressed signal masspoints, while maintaining sufficient trigger, reconstruction and identification efficiency. Muon candidates must be located within the extent of the inner tracking detectors and the muon chambers, i.e. $|\eta| < 2.4$.

The leptons originating from the SUSY signal are expected to be prompt and isolated. As a result, tight requirements are set for 3D impact parameter, IP_{3D}, and its significance, SIP_{3D}, as defined in Sec. 3.4.7. A combination of absolute isolation ($\mathrm{Iso}_{\mathrm{abs}}$), defined as the sum of energy deposited by PF candidates, other than the muon, in a cone of $\Delta R = 0.3$ around it, and of

Variable	Loose-not-tight ID	Tight ID
p_{T}	$\geq 3.5\,\mathrm{GeV}$	
$\lvert\eta\rvert$	≤ 2.4	
IP_{3D}	< 0.0175 cm	< 0.01 cm
SIP_{3D}	< 2.5	< 2.0
$\mathrm{Iso}_{\mathrm{abs}}$	$< 20.0\,\mathrm{GeV} + \frac{300.0\,\mathrm{GeV}^2}{p_{\mathrm{T}}}$	$< 5\,\mathrm{GeV}$
$\mathrm{Iso}_{\mathrm{rel}}$	< 1.0	< 0.5
DeepCSV veto	-	custom loose WP
Loose muon ID	✓	
Soft muon ID	✓	

Table 4.5: List of all identification criteria imposed to muon candidates for the loose-not-tight and tight ID selections. Though the criteria are the same for all years, the DeepCSV has different WPs and training depending on the year.

relative isolation ($\mathrm{Iso}_{\mathrm{rel}}$), defined as the $\mathrm{Iso}_{\mathrm{abs}}$ divided by the muon p_{T}, is used to ensure maximal performance in the whole p_{T} of the analysis. More specifically, the $\mathrm{Iso}_{\mathrm{rel}}$ requirement provides high efficiency for soft muons ($p_{\mathrm{T}} \leq 10\,\mathrm{GeV}$), while the $\mathrm{Iso}_{\mathrm{abs}}$ requirement ensures that even higher p_{T} muons remain isolated. Since leptons from the nonprompt background are expected to potentially be produced at larger displacements and may originate from within jets and, hence, be less isolated, the impact parameter and isolation criteria are relaxed for the loose-not-tight ID. The complicated expression of the $\mathrm{Iso}_{\mathrm{abs}}$ is discussed in Sec. 4.4.1.

Even though nonprompt leptons tend to come from decays of heavy flavor quarks, this is not the case for the leptons of the signal. As a consequence, a veto on the DeepCSV variable (Sec. 3.4.7) is set to reject leptons that are associated with a b-tagged jet. The loose b tagging DeepCSV working point (WP) is used to this end, corresponding to 10% misidentification probability. However, for the 2016 DeepCSV training, this WP causes an efficiency drop for leptons selected by the analysis with p_{T} between 10 and 25 GeV, as shown in Fig. 4.8, left. Because maintaining high efficiency is extremely important for the soft leptons of this analysis, the 2016 b tagging DeepCSV WP was modified to approximately match the efficiency of the other years. The original 2016 loose b tagging DeepCSV WP, set at 0.22, provides an average efficiency of 94.3%, while the same WP of 2017(2018), set at 0.1522(0.1241)

has an efficiency of $\sim 99\%$. To mitigate the inefficiency, the 2016 b tagging DeepCSV WP used for the rejection of leptons coming from b-tagged jets was moved to the value 0.4, achieving an average efficiency of 99.1%, comparable with the other years (Fig. 4.8, right).

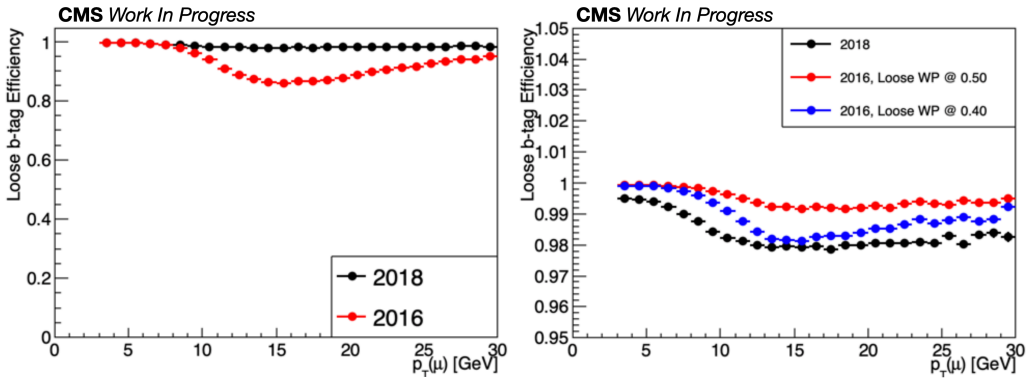


Figure 4.8: Loose b tagging DeepCSV WP efficiency as a function of the muon p_{T}, after all the tight ID requirements have been applied. Left: Comparison of the centrally provided loose WP for 2016 and 2018. Right: Comparison of the centrally provided loose WP for 2018 and the retuned values of the WP for 2016. The 0.40 value is chosen for the analysis.

Finally, the centrally provided loose and soft muon IDs are applied to both tight and loose-not-tight muon selections [49]. The loose muon ID is designed to identify prompt muons with high efficiency, while maintaining a low rate of the misidentification of charged hadrons as muons. The soft muon ID is optimized for low p_{T} muons, ensuring high purity for tracker muons by imposing tight track segment matching requirements [49].

The muon reconstruction and identification efficiencies are measured in both data and simulation for each year separately, taking advantage of the total integrated luminosity, i.e. $35.9\,\mathrm{fb}^{-1}$ in 2016, $41.4\,\mathrm{fb}^{-1}$ in 2017 and $59.7\,\mathrm{fb}^{-1}$ in 2018. To better account for geometric effects, the measurements are performed independently for barrel ($|\eta| < 1.2$) and endcap ($1.2 < |\eta| < 2.4$). The muon efficiency is factorized in three components:

$$\epsilon(\mu) = \epsilon(\mu)_{\text{Tracking}} * \epsilon(\mu)_{\text{Loose } \mu \text{ ID|Tracking}} * \epsilon(\mu)_{\text{Tight ID|Loose } \mu \text{ ID}} \tag{4.3}$$

where $\epsilon(\mu)_{\text{Tracking}}$ is the efficiency to reconstruct a track in the inner tracking detectors, $\epsilon(\mu)_{\text{Loose } \mu \text{ ID|Tracking}}$ is the efficiency to identify that track as

a loose muon candidate, after an inner track has been reconstructed, and $\epsilon(\mu)_{\text{Tight ID}|\text{Loose }\mu\text{ ID}}$ is the efficiency of all the remaining selection criteria associated to the tight muon selection of the analysis (identification, isolation, and impact parameter requirements), after a muon has passed the loose muon ID. The data over simulation efficiency ratio is applied as SF on the simulation for each individual muon selected by the analysis.

The tracking and the loose muon ID efficiencies are provided centrally by CMS, since they apply to the vast majority of analyses that involve muons. The tracking efficiency is very close to unity ($> 99.5\%$) and its effect is considered negligible, so no correction SF is taken into account for it. The loose muon ID efficiency is ~ 1 as well. However, as shown in Fig. 4.9, the correction SFs can be a few percent different from unity, especially for muons that pose reconstruction difficulties, e.g. muons with $p_{\text{T}} = 3.5\,\text{GeV}$ in the barrel region that barely reach the muon chambers, due to their track curvature by the magnetic field. The loose muon ID efficiencies and SFs have been measured using muons from the J/ψ peak for $p_{\text{T}} < 20\,\text{GeV}$ and from the Z peak for higher p_{T}.

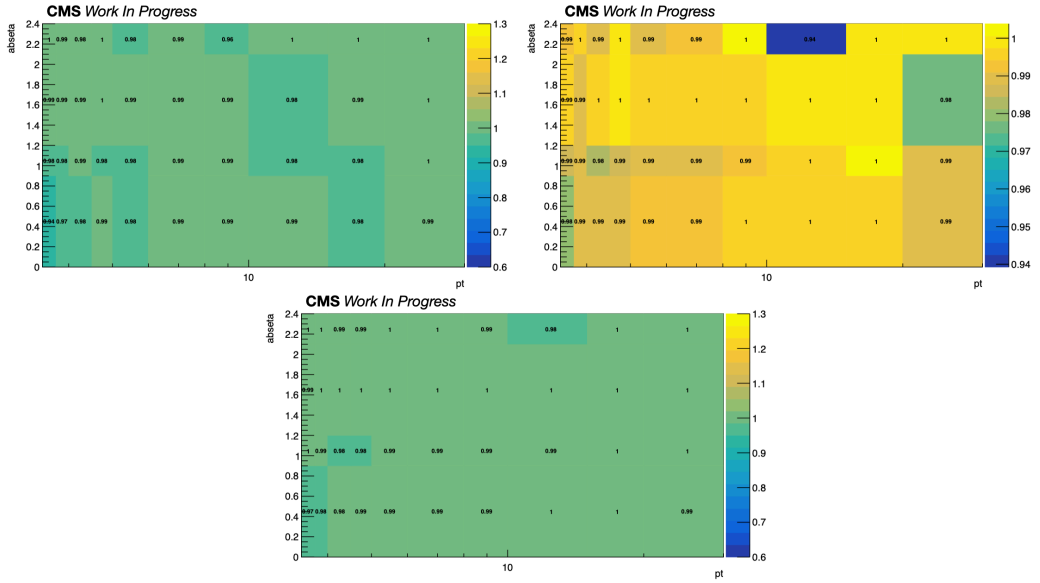


Figure 4.9: The 2016 (upper left), 2017 (upper right) and 2018 (bottom) loose muon ID SFs for $3.5 < p_{\text{T}}(\mu) < 30\,\text{GeV}$, measured on muons from the J/ψ peak using the full integrated luminosity of each year.

The tight ID efficiency for the muons is measured with the tag-and-probe method [114], using muons from Z boson decays. After the tag muon has been identified with strict requirements to ensure that it is originating from the Z boson, the probe muon is tested against passing the requirements of the tight ID of the analysis. The efficiency is calculated by dividing the probe muons from Z passing the tight ID by the total number of probe muons from Z. These values are extracted using a double Voigtian function for the signal distribution and an exponential function multiplied by an error function "turn-on" for the background distribution. The systematic uncertainty on the efficiency is estimated by taking into account the efficiency variation when using an alternate signal fit, in this case a double Gaussian function, as well as an alternate background fit, in this case an exponential function. The statistical uncertainty is found to be negligible.

In general, muons coming from J/Ψ decays could in principle be used for low p_{T} muons and could provide accurate efficiency measurements in regions where the statistical and systematic uncertainties of muons from the Z boson are relatively large. However, the tight ID criteria contain strict requirements on the muon impact parameters and isolation. Because muons from J/Ψ production are expected to be isolated and slightly displaced, they provide a low purity sample compared to the muons from Z in this case.

The usage of double peak functions for the signal distribution is due to a secondary bump appearing at $M(\ell\ell) < m_{\mathrm{Z}}$. Probe muons emitting photons as final state radiation tend to give rising to mass bumps at masses lower than that of the Z boson, depending on the amount of energy they radiate away. This effect is more striking for probe muons failing the tight ID, since the extra photon makes those muon to no longer be isolated. The phenomenon is present for electrons as well and it is even more prominent in their case, due to the higher final state radiation probability of electrons. An example of the secondary bump and the relevant signal fit is shown in Fig. 4.14 for electrons in simulation.

The efficiency measurement is exceptionally challenging in the low p_{T} range especially for the probe muons failing the tight ID. There, the background, originating mainly from W+jet events, is extremely large, leading to large systematic uncertainties from its modelling. An example of the passing and failing probe distributions of the dimuon invariant mass for probe muons with $3.5 < p_{\mathrm{T}} < 10\,\mathrm{GeV}$ is given in Fig. 4.10 and showcases the effect.

In general, the tight ID muon efficiency starts off at $\sim 70\%$ for muons with $p_{\mathrm{T}} \sim 3.5\,\mathrm{GeV}$ 4.11. The SFs obtained are in the range of 0.9–1.1. A

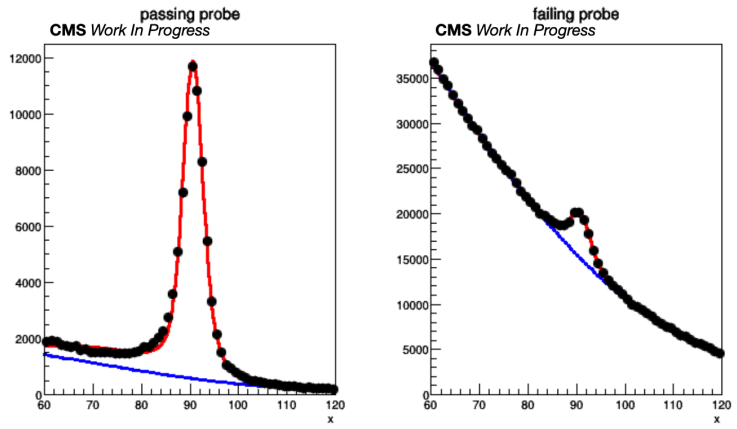


Figure 4.10: The dimuon mass distribution in data for the passing and failing probes with $3.5 < p_{\mathrm{T}} < 10\,\mathrm{GeV}$ in the endcap, used for the tight ID efficiency and scale factor measurement.

small gap in the SFs is observed at $p_{\mathrm{T}} = 10\,\mathrm{GeV}$ due to the the fact that this p_{T} value is the one when the $\mathrm{Iso}_{\mathrm{abs}}$ requirement becomes stricter than the $\mathrm{Iso}_{\mathrm{rel}}$ one.

Similar procedure has been followed for the calculation of the leptonic efficiency on FastSIM samples as well. The result is divided the FullSIM leptonic efficiency to give the "leptonic FastSIM SFs". These are applied, on top of the leptonic FullSIM SFs, to FastSIM samples only to correct the imperfect modeling of lepton quantities in the FastSIM. The leptonic FastSIM SFs are shown for muons in barrel and endcap, as a function of p_{T}, in Fig. 4.12.

4.3.2 Electrons

Electron candidates used in this analysis are formed by associating ECAL clusters to GSF tracks, as described in Sec. 3.4.5, and are selected according to the criteria summarized in Table 4.6.

Similarly to the muons, the lower p_{T} threshold for electrons is set to the lowest GSF electron value, at 5 GeV, while the maximum $|\eta|$ threshold is 2.5 to match the ECAL acceptance. The impact parameter, isolation and DeepCSV criteria are common between muons and electrons, backed up by the reasoning outlined in Sec. 4.3.1. As in the case of muons, electrons must be identified according to the relevant centrally provided ID. The electron

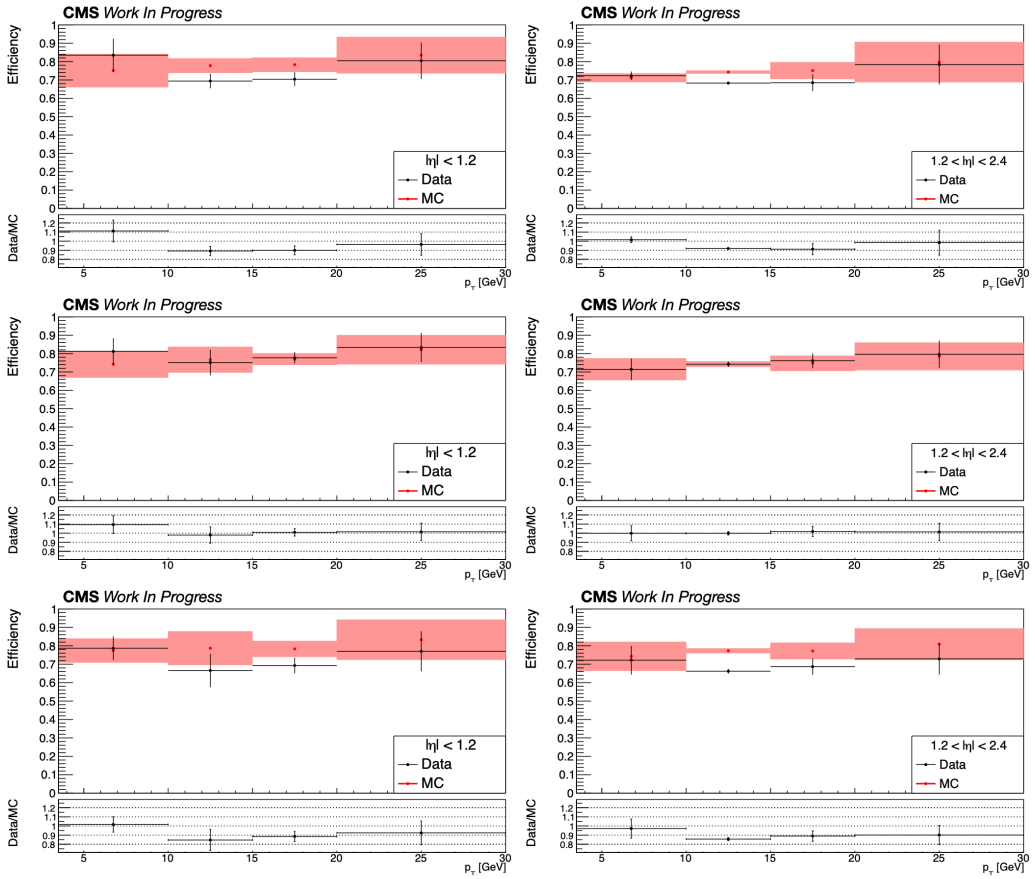


Figure 4.11: The 2016 (upper row), 2017 (middle row) and 2018 (lower row) muon efficiencies of the tight ID of the analysis in data and MC, obtained with the full integrated luminosity of each year, as function of p_{T} in the barrel (left) and the endcap (right). The uncertainty shown is both statistical and systematic for the data and the same for the MC. The data over MC scale factors are shown at the bottom of each plot.

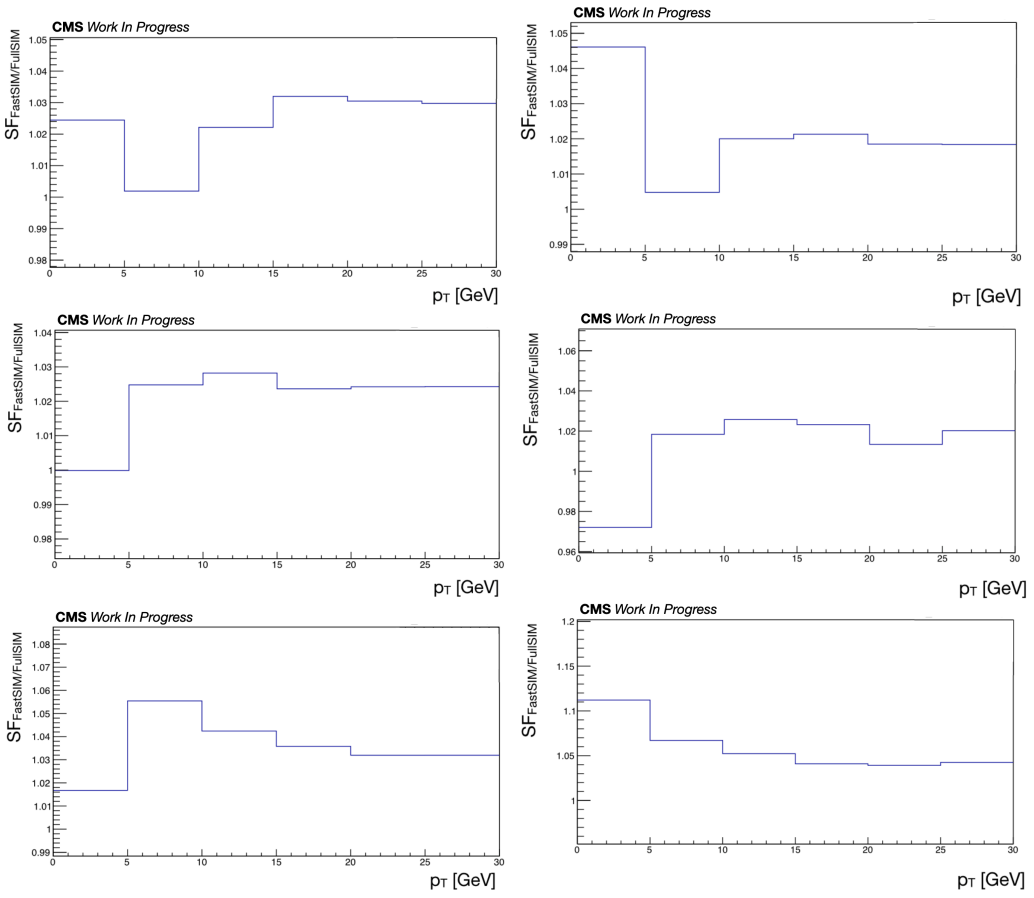


Figure 4.12: The 2016 (upper row), 2017 (middle row) and 2018 (lower row) leptonic FastSIM SFs, obtained with the full integrated luminosity of each year, as function of p_{T} in the barrel (left) and the endcap (right).

Variable	Loose-not-tight ID	Tight ID
p_{T}	$\geq 5\,\mathrm{GeV}$	
$\lvert\eta\rvert$	≤ 2.5	
IP_{3D}	< 0.0175 cm	< 0.01 cm
SIP_{3D}	< 2.5	< 2.0
$\mathrm{Iso}_{\mathrm{abs}}$	$< 20.0\,\mathrm{GeV} + \frac{300.0\,\mathrm{GeV}^2}{p_{\mathrm{T}}}$	$< 5\,\mathrm{GeV}$
$\mathrm{Iso}_{\mathrm{rel}}$	< 1.0	< 0.5
DeepCSV veto	-	custom loose WP
Electron MVA ID	custom ID	tight WP
No missing pixel hits	✓	
No conversion vertex	✓	

Table 4.6: List of all identification criteria imposed to electron candidates for the loose-not-tight and tight ID selections. Though the criteria are the same for all the years, the electron MVA ID and the DeepCSV have different WPs and trainings depending on the year.

MVA ID, mentioned in Sec. 3.4.5, is used with a WP that provides approximately 90% of efficiency inclusively for the tight ID. On the other hand, the electron MVA ID WP for the loose-not-tight ID was specifically created to minimize the flavor dependency of jets misreconstructed as electrons for the nonprompt background estimation, as discussed in Sec. 4.4.1. Finally, electrons selected by the analysis are required to be built from tracks that lack no pixel hits and to not be associated with a conversion vertex. The latter two requirements suppress backgrounds arising from misreconstructed electrons or photon conversion processes.

The strategy for measuring the electron reconstruction and identification efficiencies is similar to the one described for muons: They are evaluated per year, per detector region (barrel, $|\eta| < 1.5$, and endcap, $1.5 < |\eta| < 2.5$ and are factorized in separate components. The electron efficiency is broken down to $\epsilon(e)_{\mathrm{Reco}}$, which takes into account the efficiency for reconstructing a GSF electron from the association of a GSF track and a calorimeter cluster, and to $\epsilon(e)_{\mathrm{Tight\ ID|Reco}}$, which quantifies the efficiency for a GSF electron to be identified with the tight ID of the analysis:

$$\epsilon(e) = \epsilon(e)_{\mathrm{Reco}} * \epsilon(e)_{\mathrm{Tight\ ID|Reco}} \tag{4.4}$$

The ratio $\epsilon(e)_{\mathrm{data}}/\epsilon(e)_{\mathrm{MC}}$ is applied to the simulated events of the analysis in

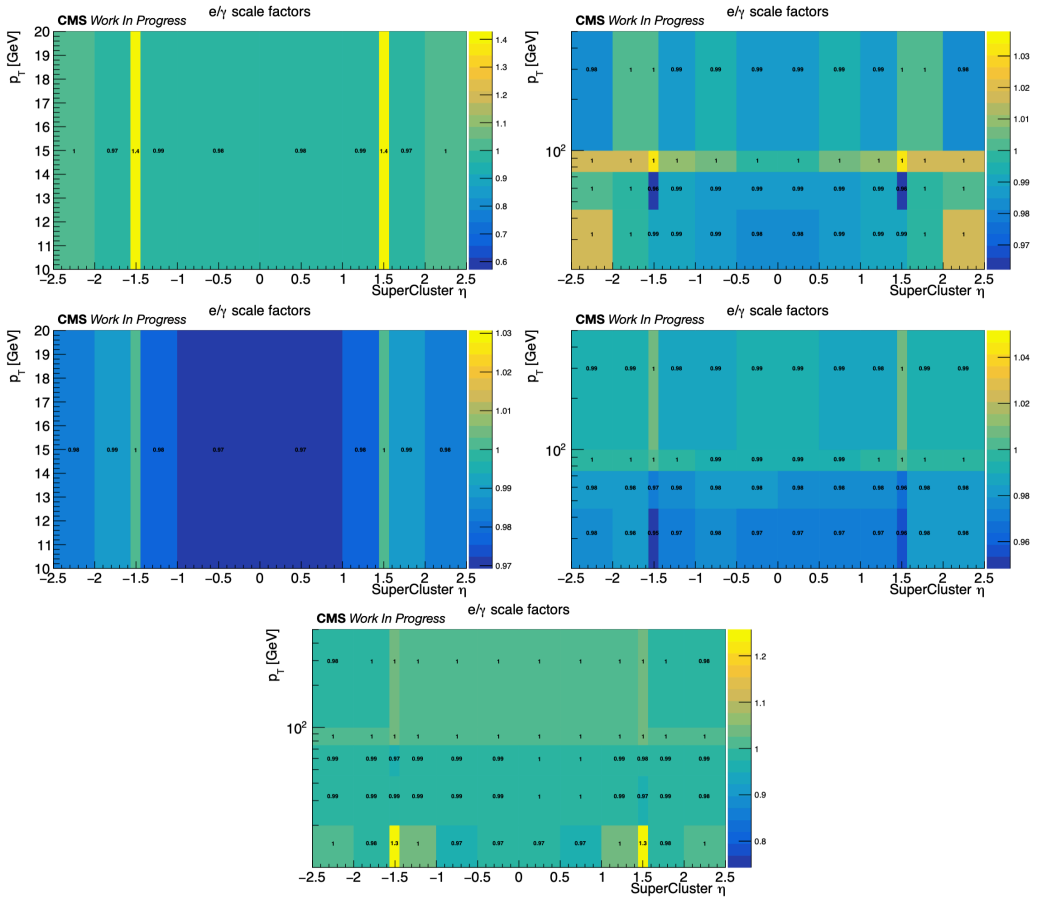


Figure 4.13: The 2016 (upper row), 2017 (middle row) and 2018 (bottom row) electron reconstruction SFs for $3.5 < p_{\mathrm{T}}(\mu) < 500\,\mathrm{GeV}$, measured on electrons from the Z peak using the full integrated luminosity of each year.

order to harmonize the electron selection efficiency in data and simulation.

The electron reconstruction efficiency is measured centrally by CMS for all interested analyses and is higher than 95% in the whole p_{T} range, reaching up to $\sim$100% for electrons above 100 GeV. The extracted data-to-simulation SFs are very close to unity for all years and detector regions, with the exception of the region $1.444 < |\eta| \sim 1.566$, where there is the crack between the ECAL barrel and endcap and the detector modeling is challenging. The electron reconstruction SF are shown in Fig. 4.13 for $10 < p_{\mathrm{T}} < 500\,\mathrm{GeV}$. The SFs for the range $5 < p_{\mathrm{T}} < 10\,\mathrm{GeV}$ are not explicitly measured. However, given

their mild p_{T} dependence, the SF values of the p_{T} range $10-20\,\mathrm{GeV}$ are used down to $p_{\mathrm{T}} = 5\,\mathrm{GeV}$ as a function of $|\eta|$.

The electron tight ID efficiencies are measured using the same tag-and-probe method as used for the muons, facing similar challenges due to the huge W+jets background at low p_{T} and the emission of final state photons that create a secondary mass bump. Fig. 4.14 shows an example of the latter effect.

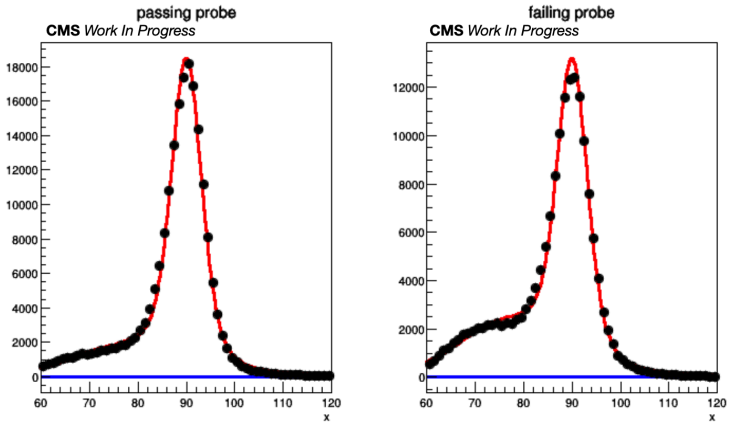


Figure 4.14: The dielectron mass distribution in simulation for the passing and failing probes with $20 < p_{\mathrm{T}} < 30\,\mathrm{GeV}$ in the endcap, used for the tight ID efficiency and scale factor measurement.

Fig. 4.15 demonstrates the electron tight ID values along with the respective SFs for all years and regions. Due to the difficulty of identifying isolated soft electrons with high purity, the efficiency starts at $\sim 40\%$ and grows to reach a plateau of $\sim 50\%$ for the endcap and $\sim 60\%$ for the barrel. The SFs correct deviations from unity up to 20%, with the SF gap at $p_{\mathrm{T}} = 10\,\mathrm{GeV}$ being observed also in this case.

As in the case of muons, the leptonic FastSIM SFs are computed and applied for the electrons using the methods mentioned above. The leptonic FastSIM SFs are shown for electrons in barrel and endcap, as a function of p_{T}, in Fig. 4.16.

4.3.3 Jets and Missing Transverse Energy

Apart from leptons, jets are also utilized in the search strategy. The jets used in the analysis are AK4 jets, as defined in Sec. 3.4.7, and the PU particles

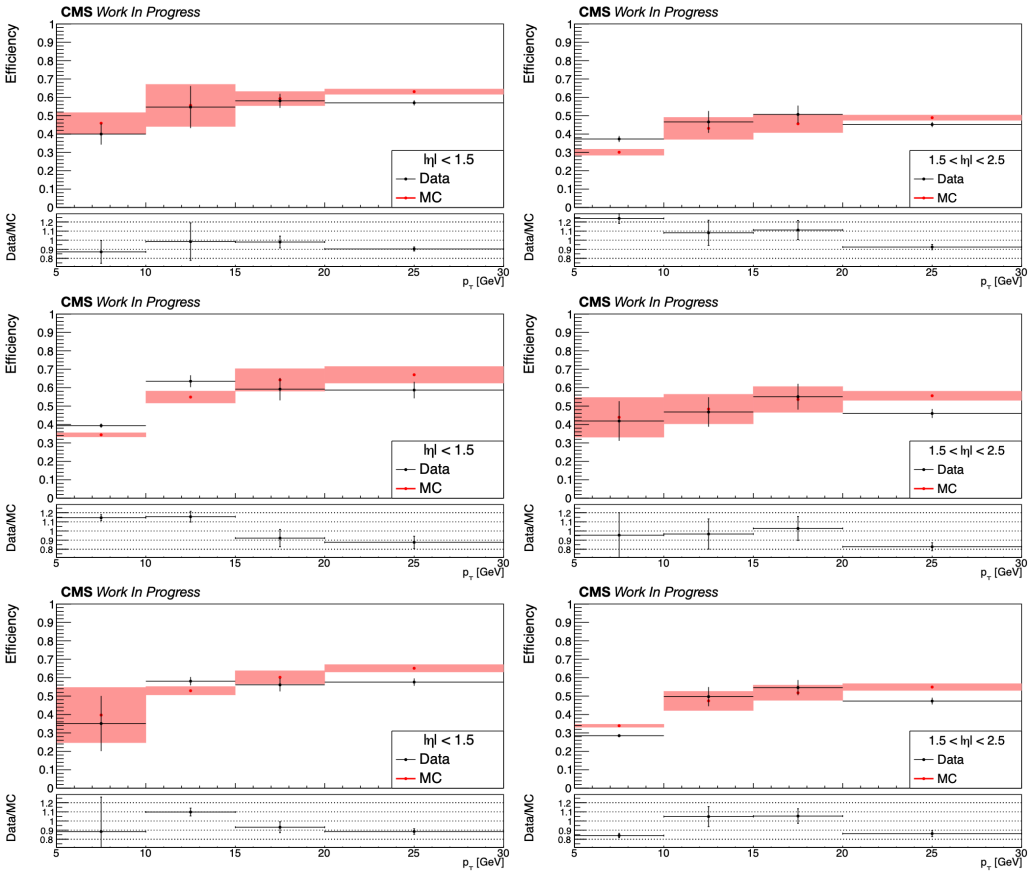


Figure 4.15: The 2016 (upper row), 2017 (middle row) and 2018 (lower row) electron efficiencies of the tight ID of the analysis in data and MC, obtained with the full integrated luminosity of each year, as function of p_{T} in the barrel (left) and the endcap (right). The uncertainty shown is both statistical and systematic for the data and the same for the MC. The data over MC scale factors are shown at the bottom of each plot.

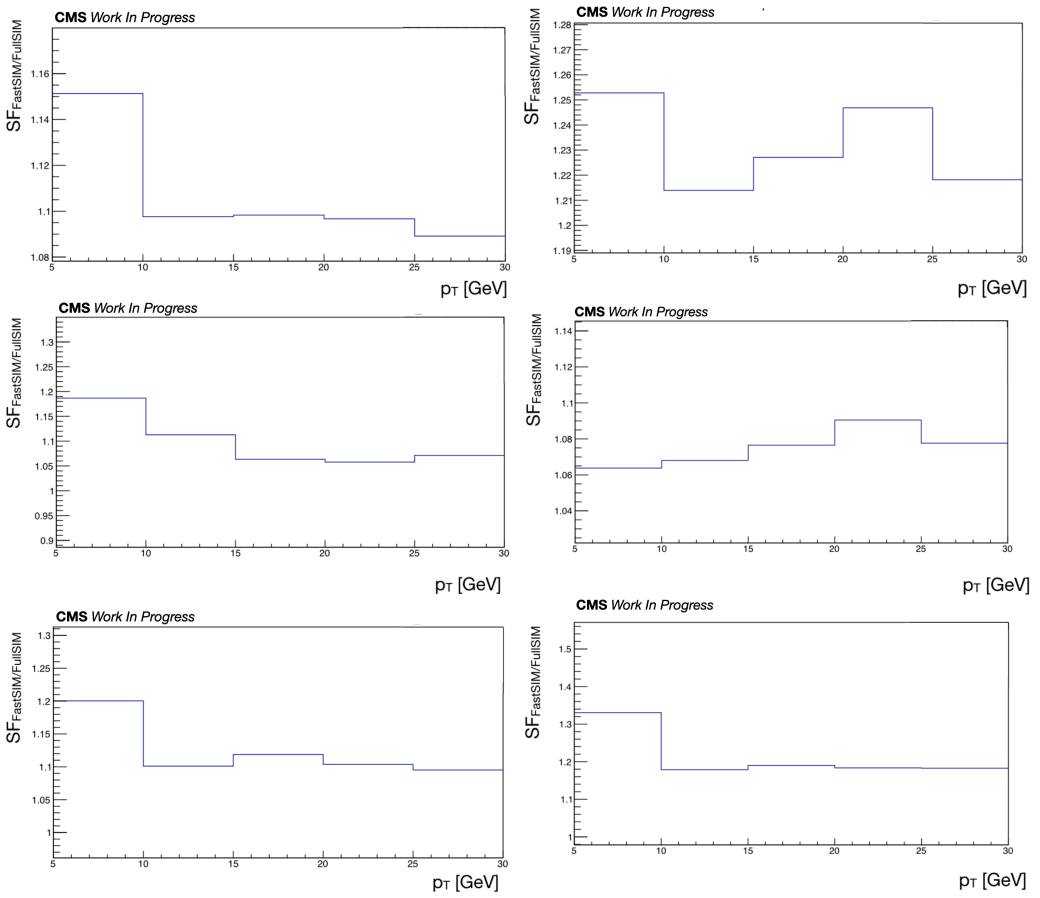


Figure 4.16: The 2016 (upper row), 2017 (middle row) and 2018 (lower row) leptonic FastSIM SFs, obtained with the full integrated luminosity of each year, as function of p_{T} in the barrel (left) and the endcap (right).

are removed with CHS method. Each jet must have p_{T} of at least 25 GeV and be located within the tracker acceptance ($|\eta| < 2.4$). The leading jet is also required to satisfy the *tight lepton veto* ID requirement, as defined in Ref. [115], which is effective in rejecting the background originating from misreconstructed muons and electrons. In order to resolve the ambiguity between lepton candidates and jets and to avoid potential double counting of objects, a procedure, called *jet cleaning*, is implemented: The jet closest to a lepton identified with the loose ID of the analysis is removed if they share a common PF candidate component. The standard L1 and L2L3 JECs (described in Sec. 3.4.7) are also applied.

Jets originating from the b quarks are tagged using the medium WP of the DeepCSV discriminant, mentioned in Sec. 3.4.7. The threshold value of the WP is extracted separately for the data set of each year, so that the mistag rate for light-flavor jets is of the order 1%. This WP corresponds to an approximate efficiency of 68%. Differences of b tagging efficiencies between data and simulation are corrected with SFs measured in dedicated data samples [23].

The $p_{\mathrm{T}}^{\mathrm{miss}}$ or $p_{\mathrm{T}}^{\mathrm{miss,corr}}$ objects (specified in Sec. 4.2.1) are reconstructed using the standard algorithm based on particle flow reconstructed objects with the JECs propagated, discussed in Sec. 3.4.9. Event filters are designed to remove data events with anomalously high values of $p_{\mathrm{T}}^{\mathrm{miss}}$. These can arise from multiple sources, such as detector noise, reconstruction inefficiencies or beam related effects (e.g. beam halos). The recommended event filters, listed and explained in Ref. [70], are applied on the events selected by the search. Finally, the distribution of PU interactions is corrected with dedicated SFs, extracted by the method described in Ref. [116].

Since the requirement of an ISR jet is of central importance for the analysis strategy, the modeling of the ISR p_{T} distribution is validated by comparing the simulation with data. Events that contain exactly one pair of OS, same-flavor leptons with $80 < M(\ell\ell) < 100$ GeV and at least one jet are selected. In 2016, deviations up to 20% are observed, which are corrected by applying p_{T}-dependent weights on the EWK signal samples. To keep the overall normalization of the signal sample constant after the ISR reweighting procedure, an extra, masspoint-dependent weight is applied. For the stop signal samples of 2016 and all of the signal samples (EWK and stop) of 2017 and 2018, studies showed that the data-simulation disagreement is of the order of 1%, so no dedicated weights are used. For all the EWK signal samples, a systematic uncertainty is implemented to account for the modeling of the

ISR jet.

Finally, corrections for the use of the fast detector simulation are also applied to the top squark pair production and the pMSSM higgsino signal samples. More specifically, the FastSIM $p_{\mathrm{T}}^{\mathrm{miss}}$ resolution is improved by using the average of the $p_{\mathrm{T}}^{\mathrm{miss}}$ calculated from the reconstructed objects and the $p_{\mathrm{T}}^{\mathrm{miss}}$ calculated with the generator objects:

$$p_{\mathrm{T}}^{\mathrm{miss}} = \frac{1}{2}\left(p_{\mathrm{T}}^{\mathrm{miss}}(\mathrm{Reconstructed}) + p_{\mathrm{T}}^{\mathrm{miss}}(\mathrm{Generator})\right) \tag{4.5}$$

Any potential additional mismodeling is covered by an extra systematic uncertainty, dedicated to account for this effect.

4.3.4 Baseline Event Selection and Categorization

The signal events that this analysis targets include two or three leptons with moderate to large amount of MET, induced by the presence of an ISR jet. Since the final states differ slightly between the EWK and the stop signal models and on the number of final state leptons, separate signal regions (SRs) are designed in each case: The 2ℓ-EWK and 3ℓ-EWK SRs are optimized for the EWK models with two and three leptons, respectively, while 2ℓ-Stop SRs target the signature produced by the stop models. A detailed discussion on the selection requirements of each SR and the differences between them is presented below.

Regarding the MET requirements, these are driven by the trigger requirements mentioned in Sec. 4.2.1. The pure $p_{\mathrm{T}}^{\mathrm{miss,corr}}$ trigger paths reach an efficiency plateau at approximately $p_{\mathrm{T}}^{\mathrm{miss,corr}} = 200\,\mathrm{GeV}$, setting the lower boundary for the higher MET bins of the analysis. Depending on the SR, the $p_{\mathrm{T}}^{\mathrm{miss,corr}} > 200\,\mathrm{GeV}$ phase space is split further into bins in order to maximize the sensitivity of the search to low mass differences ($\Delta m < 10\,\mathrm{GeV}$). The *medium*, *high* and *ultra* MET bins are defined for the 2ℓ SRs, with sightly different values for the EWK and Stop SRs, as shown in Table 4.7. The 3ℓ-EWK SR is not split any further because of the low event yield it contains. To recover events with lower values of MET, the double-μ+MET trigger path is utilized to extend the acceptance down to $p_{\mathrm{T}}^{\mathrm{miss,corr}} = 125\,\mathrm{GeV}$. The $125 < p_{\mathrm{T}}^{\mathrm{miss,corr}} < 200\,\mathrm{GeV}$ range defines the low MET bin of the analysis. Due to the online requirement also on $p_{\mathrm{T}}^{\mathrm{miss}}$ for the double-μ+MET trigger path, an extra offline criterion of $p_{\mathrm{T}}^{\mathrm{miss}} > 125\,\mathrm{GeV}$ is set only for the low MET

bin. The complete definition of the MET binning strategy of the search is given in Table 4.7.

SR	Low-MET		Medium-MET	High-MET	Ultra-MET
	$p_{\mathrm{T}}^{\mathrm{miss}}$	$p_{\mathrm{T}}^{\mathrm{miss,corr}}$	$p_{\mathrm{T}}^{\mathrm{miss,corr}}$	$p_{\mathrm{T}}^{\mathrm{miss,corr}}$	$p_{\mathrm{T}}^{\mathrm{miss,corr}}$
2ℓ-EWK	> 125	$(125, 200]$	$(200, 240]$	$(240, 290]$	> 290
3ℓ-EWK	> 125	$(125, 200]$	—	> 200	—
2ℓ-Stop	> 125	$(125, 200]$	$(200, 290]$	$(290, 340]$	> 340

Table 4.7: Definition of the MET bins of the SRs. The boundaries of $p_{\mathrm{T}}^{\mathrm{miss}}$ and $p_{\mathrm{T}}^{\mathrm{miss,corr}}$ (in GeV) of every bin are described.

Events are further categorized based on the most discriminating variable of each final state. For the 2ℓ-EWK SRs, which are designed to be sensitive to the processes on the right diagrams of Figs. 4.2 and 4.3, the invariant mass of the lepton (muon or electron) pair is used. Since selected leptons are required to have the same flavor (SF) and OS, they are bound to originate from the off-shell Z boson produced in the $\widetilde{\chi}_2^0 \to \widetilde{\chi}_1^0$ decay in the aforementioned processes. This leads to the relation $M(\ell\ell) \sim M_{\mathrm{Z}^*} \sim \Delta m(\widetilde{\chi}_2^0, \widetilde{\chi}_1^0)$, meaning that the $M(\ell\ell)$ signal distribution exhibits an endpoint at the mass differences value of each signal masspoint, as shown in Fig. 4.17. No such feature is present for the SM background processes. The $M(\ell\ell)$ distribution bins used for the categorization of the events, optimized to enhance sensitivity for masspoints with Δm between 5 and 10 GeV, are $[1, 4, 10, 20, 30, 50]$ GeV for the higher MET bins. The binning of the low-MET bin is the same with the omission of the $[1, 4]$ GeV $M(\ell\ell)$ bin. This is due to the intrinsic lower bound of the dilepton invariant mass at 4 GeV in the double-μ+MET trigger path, used to collect the data in this specific low-MET bin. To add to that, the double-μ+MET trigger path selects only dimuon events, therefore the low-MET bins in all the SRs contains only events with at least one muon pair.

Low mass resonances are also taken into account in the analysis event selection: A Υ-meson veto is applied in all the MET bins by rejecting events with $9 < M(\ell\ell) < 10.5$ GeV. Apart from that, in the higher MET bins, where the $M(\ell\ell)$ goes below 4 GeV and, due to that, leptons can be close together, a J/ψ veto and an overlapping-lepton veto are applied by rejecting events with $3 < M(\ell\ell) < 3.2$ GeV and $\Delta R(\ell\ell) < 0.3$, respectively.

In case of the 3ℓ-EWK SRs, the invariant mass of the leptons coming from the Z boson is again a probe for mass difference of the signal masspoints.

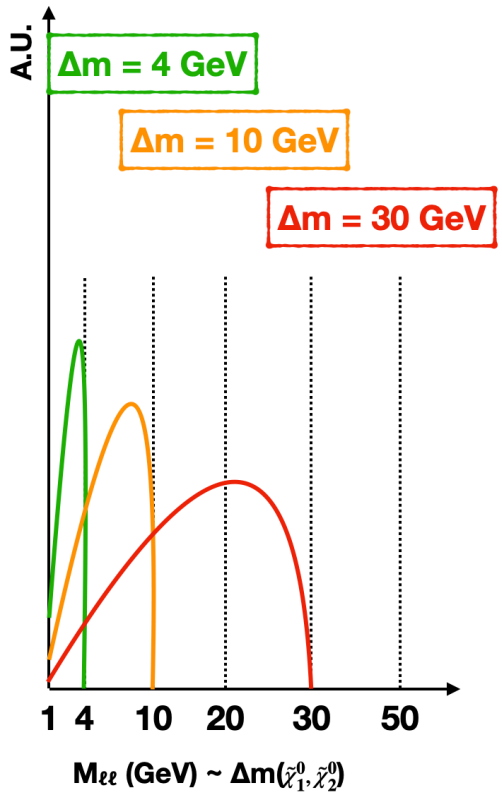


Figure 4.17: Example graphical representation of the endpoint exhibited in the dilepton invariant mass signal distribution for different signal masspoints.

However, the presence of a third lepton, coming from the leptonic decay of the W boson from the other decay leg of the process on the right diagram of Fig. 4.2, leads to the creation of another SFOS lepton pair in the final state. Due to the fact that this analysis targets compressed mass models, the ambiguity is resolved by selecting the $M(\ell\ell)$ of the SFOS lepton pair with the smallest value, denoted as $M^{\text{min}}_{\text{SFOS}}(\ell\ell)$. To be in accordance with the upper $M(\ell\ell)$ bound in the double-μ+MET, the maximum invariant mass of the SF and any (charge) sign (AS) lepton pairs, $M^{\text{max}}_{\text{SFAS}}(\ell\ell)$, is required to be less than 60 GeV in the trilepton low-MET bin.

In all the SRs, the leptons are selected to have low p_{T}, up to 30 GeV. Moreover, for the 2ℓ-Stop SRs, which target the stop production models, shown in upper right and lower diagrams of Fig. 4.5, the leading lepton p_{T} is used to categorize the events in lieu of the dilepton invariant mass. The

leptons in the case of stop models come from different decay legs, hence their invariant mass has no physical meaning, while the p_{T} of the leading lepton still provides some discrimination between the SUSY signal and the SM background processes. The binning used is $[3.5, 5, 12, 16, 20, 25, 30]$ GeV for the higher MET bins, with the omission of the lowest p_{T} bin in the low-MET bin, where such low lepton p_{T} values are still on the double-μ+MET trigger efficiency turn-on. The fact that the leptons in these models do not share a common production vertex also lifts the requirement for them to be of the same flavor in the 2ℓ-Stop SRs. However, this selection criterion is not lifted in the low-MET bin, where the intrinsic trigger path requirement for a dimuon pair still remains. In the 2ℓ SRs only, the p_{T} of the lepton pair is required to be greater than 3 GeV. The SR event categorization is graphically depicted in Fig. 4.18.

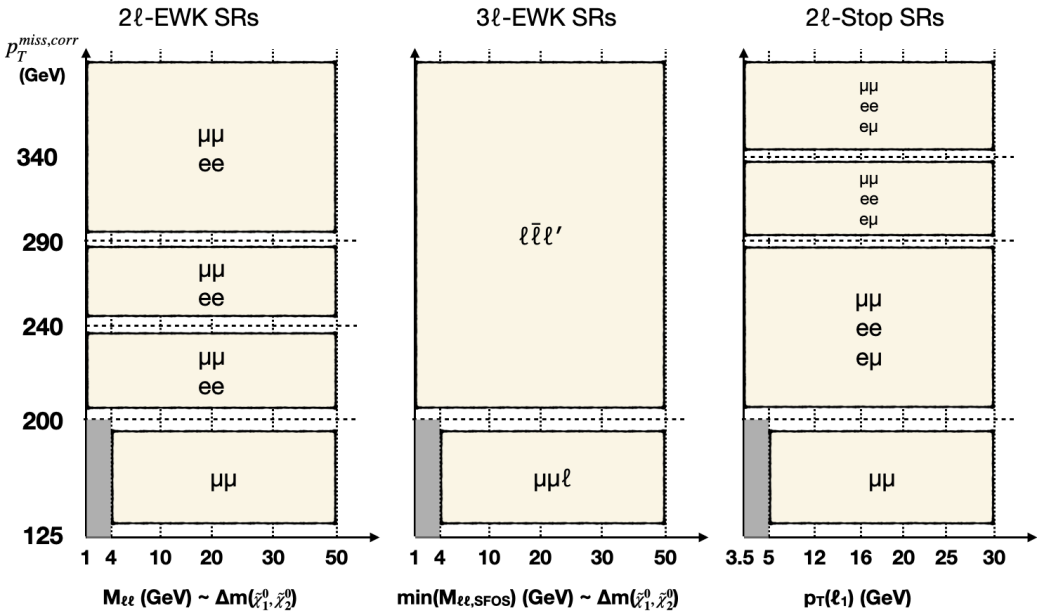


Figure 4.18: Graphical representation of the event categorization in the different SRs of the analysis.

An $H_T > 100$ GeV requirement is applied to all of the SRs as a consequence of the event selection at the trigger level. This requirement also suppresses background process with low hadronic activity. Furthermore, for the signal processes, no final state radiation can be produced with high momentum: In fact, photons or gluons cannot be radiated from the LSPs, which

are neutral and not strongly interacting, while the intermediate SM particles produced have low available energy in their decay to begin with. As a result, only initial state radiation can be responsible for large amounts of H_{T} required in the event selection. Imposing the requirement of "tight lepton veto" [115] identification criteria for the leading jet, which removes jets from calorimetric noise and from misreconstructed leptons, in combination with the H_{T} requirement, becomes equivalent with the requirement of an ISR jet, which boosts the final state objects and induces large enough amounts of MET to capture events by the trigger paths available. The tight identification criterion for the leading jet is only applied in the 2ℓ SRs of the analysis, while it is removed in the 3ℓ SRs to enhance the low event yield of those regions. In the dilepton SR selection, the relation $2/3 < p_{\mathrm{T}}^{\mathrm{miss}}/H_{\mathrm{T}} < 1.4$ is additionally required, since it retains the ISR boosted events from signal processes, while significantly reduces the contribution from QCD multijet events.

The event selection is also tailored to reject the main reducible background processes of the analysis: To reject the majority of the $\mathrm{t\bar{t}}$ background, a veto for b-tagged jets with $p_{\mathrm{T}} > 25\,\mathrm{GeV}$, identified by the DeepCSV algorithm, is applied. It is worth mentioning that the low p_{T} bound of this veto still allows for processes with soft bottom quarks in the final state, such as the ones of the stop production models of Fig. 4.5. The contribution of the $\mathrm{t\bar{t}}$ process along with that of other diboson processes, most importantly the WW process, are reduced by imposing an upper bound on the transverse mass of the each of the leptons and the $p_{\mathrm{T}}^{\mathrm{miss}}$ of the event at the scale of the mass of the W boson, i.e. $M_{\mathrm{T}}(\ell, p_{\mathrm{T}}^{\mathrm{miss}}) < 70\,\mathrm{GeV}$. However, this specific requirement is removed for the 2ℓ-Stop SRs in order to maintain a high signal acceptance, as well as for the 3ℓ-EWK SRs, where the aforementioned process are negligible.

The suppression of the DY process is achieved by the approximate reconstruction of the mass of the intermediate Z boson [117]. The Z boson decays to electrons or muons produce no invisible particles. Therefore, DY events with leptonic final states via these decay modes contain only instrumental $p_{\mathrm{T}}^{\mathrm{miss}}$, which is rarely enough to pass the MET thresholds of the analysis event selection. As a result, the Z boson decay mode that mostly contributes to the background of this analysis is the $\mathrm{Z} \to \tau\tau \to ee\nu_e\nu_e/\mu\mu\nu_\mu\nu_\mu\ \nu_\tau\nu_\tau$, which contains real $p_{\mathrm{T}}^{\mathrm{miss}}$ due to the presence of the neutrinos. Moreover, the ISR jet boosts the Z boson, causing the momentum directions of the neutrinos to be approximately the same as the ones of the final leptons. Under

this assumption and the assumption that the $p_{\mathrm{T}}^{\mathrm{miss}}$ of the event is predominantly due to the neutrinos, the momenta of the final leptons are rescaled against the hadronic recoil of the Z boson and then combined with the components of the $p_{\mathrm{T}}^{\mathrm{miss}}$ along their directions to estimate the momentum of the τ leptons. The momenta of the two τ leptons are subsequently added to approximate the momentum of the initial Z boson and the mass of the latter, $M_{\tau\tau}$, is computed. Due to long tails in the $M_{\tau\tau}$ distribution, the large range $0 < M_{\tau\tau} < 160\,\mathrm{GeV}$ usually signifies a DY event and it is vetoed from the SR event selection. The invalidity of either of the aforementioned assumptions, which is typically an evidence against the presence of a DY process, leads to unphysical values of the $M_{\tau\tau}$, either very large ones, $> 160\,\mathrm{GeV}$, or even negative ones, which are a consequence of the final leptons changing direction during the momentum rescaling procedure. Since the DY process leads to dilepton final states, the $M_{\tau\tau}$ veto is not applied in the 3ℓ-EWK SRs.

The event selection of the analysis is summarized in Table 4.8.

Variable	2ℓ-EWK		2ℓ-Stop		3ℓ-EWK	
	Low-MET	Higher-MET	Low-MET	Higher-MET	Low-MET	Higher-MET
Tight ID N_{lep}	2		2		3	
1 lepton pair	OS$\mu\mu$	OSSF	OS$\mu\mu$	OS	OS$\mu\mu$	OSSF
$p_{\mathrm{T}}^{\mathrm{miss,corr}}$ [GeV]	> 125		> 125		> 125	
$p_{\mathrm{T}}^{\mathrm{miss}}$ [GeV]	> 125	—	> 125	—	> 125	—
$M_{\mathrm{SFOS}}^{\mathrm{min}}(\ell\ell)$ [GeV]	$(4, 50)$	$(1, 50)$	$(4, 50)$	$(1, 50)$	$(4, 50)$	$(1, 50)$
$M_{\mathrm{SFOS}}^{\mathrm{min}}(\ell\ell)$ veto [GeV]	$(3, 3.2)$ and $(9, 10.5)$		$(3, 3.2)$ and $(9, 10.5)$		$(3, 3.2)$ and $(9, 10.5)$	
$M_{\mathrm{SFAS}}^{\mathrm{max}}(\ell\ell)$ [GeV]	—		—		< 60	—
$\Delta R(\ell\ell)$	—	> 0.3	—	> 0.3	—	> 0.3
p_{T} (ℓ_1) [GeV] for e(μ)	$(5, 30)$	$(5(3.5), 30)$	$(5, 30)$	$(5(3.5), 30)$	$(5, 30)$	$(5(3.5), 30)$
p_{T} (ℓ_2) [GeV] for e(μ)	$(5, 30)$	$(5(3.5), 30)$	$(5, 30)$	$(5(3.5), 30)$	$(5, 30)$	$(5(3.5), 30)$
p_{T} (ℓ_3) [GeV] for e(μ)	—		—		$(5, 30)$	$(5(3.5), 30)$
$p_{\mathrm{T}}(\ell\ell)$ [GeV]	> 3		> 3		—	
H_{T} [GeV]	> 100		> 100		> 100	
$p_{\mathrm{T}}^{\mathrm{miss}}/H_{\mathrm{T}}$	$(2/3, 1.4)$		$(2/3, 1.4)$		—	
Leading jet "tight lepton veto"	✓		✓		—	
$N_{\mathrm{b}}(p_{\mathrm{T}} > 25\,\mathrm{GeV})$	$= 0$		$= 0$		$= 0$	
$M_{\mathrm{T}}(\ell, p_{\mathrm{T}}^{\mathrm{miss}})$ [GeV]	< 70		—		—	
$M_{\tau\tau}$ [GeV]	veto $(0, 160)$		veto $(0, 160)$		—	

Table 4.8: List of all the event selection criteria of the analysis SRs. The label "Low-MET" refers to the low-MET bin, while the label "Higher-MET" refers collectively to the med-, high- and ultra-MET bins.

4.4 Background Estimation

The SM processes that contribute to the background of this analysis can be split into two major categories:

- The nonprompt background collectively includes SM processes that contain at least one lepton coming from a jet misidentified as a lepton or from a semileptonic decay of a heavy flavor quark. It is the most important background of the analysis. The dominant SM processes that contribute to it are the W+jets process in the 2ℓ SRs, where at least one of the jets is misidentified as a lepton, and the $t\bar{t}$ process for the 3ℓ SRs, where at least one of the bottom quarks decays into a lepton.

- The prompt background processes include all the SM processes leading to final states with two genuine leptons. The most dominant prompt backgrounds are the dileptonic $t\bar{t}$ and DY processes in the 2ℓ SRs and the fully leptonic final state of the WZ process in the 3ℓ SRs. Other diboson processes, i.e. the WW, the ZZ and the partially leptonic WZ ones, comprise the VV background of the analysis. Finally, there is a small contribution of SM processes with very low cross section, such as triboson processes or W, Z or H boson production in association with top quarks. These minor contributions are referred to as "Rare".

The general strategy for the background estimation is different depending on the nature of the background process and the significance of its contribution. The non-prompt background is typically not well-modeled in the simulation and, therefore, is estimated by a data-driven (DD) method, described in Sec. 4.4.1. The distribution shape of the major prompt SM processes, i.e. $t\bar{t}$, DY and fully leptonic WZ, is taken from simulation, while their normalization is corrected with DD scale factors. The details for the estimation of each process are given in Secs. 4.4.2, 4.4.3 and 4.4.4, respectively. Finally, the minor contributions from VV and Rare processes are taken purely from simulation and are discussed in Secs. 4.4.5 and 4.4.6, respectively.

The evaluation of the modeling of each of the major background processes is performed in specially designed control regions (CRs). These control regions are orthogonal to the SR selection with the inversion of at least one SR requirement and are enriched in the SM process under study. The same-sign (SS) CR is designed for the nonprompt background, the $t\bar{t}$ CR for the $t\bar{t}$

process, the DY CR for the DY process and the WZ enriched region (ER) for the fully leptonic WZ process. One more region, denoted as VV validation region (VR), is used for checking the data and simulation agreement and is enriched in processes with dibosons in the final state.

The event categorization in the CRs follows the one of the EWK SRs. Events are split into two MET bins:

- Low-MET bin: $p_{\mathrm{T}}^{\mathrm{miss}} > 125\,\mathrm{GeV}$ and $125 < p_{\mathrm{T}}^{\mathrm{miss,corr}} \leq 200\,\mathrm{GeV}$
- High-MET bin: $p_{\mathrm{T}}^{\mathrm{miss,corr}} > 200\,\mathrm{GeV}$.

Depending on the number of leptons in the final state (3 for the WZ ER, 2 otherwise), the CR events are also divided in bins of $M_{\mathrm{SFOS}}^{\mathrm{min}}(\ell\ell)$ or $M(\ell\ell)$, respectively, with the exactly the same binning as the SRs. The event selection for each CR is similar with the one presented in Table 4.8 with the modifications mentioned in Table 4.9 and discussed further in the following subsections.

Region	Modified selection criteria
SS(2ℓ) CR	SS lepton pair No requirement on $M_{\mathrm{T}}(\ell, p_{\mathrm{T}}^{\mathrm{miss}})$
$\mathrm{t\bar{t}}(2\ell)$ CR	$N_{\mathrm{b}}(p_{\mathrm{T}} > 25\,\mathrm{GeV}) > 0$ No requirement on $M_{\mathrm{T}}(\ell, p_{\mathrm{T}}^{\mathrm{miss}})$ No upper bound on $p_{\mathrm{T}}(\ell)$
DY(2ℓ) CR	$0 < M_{\tau\tau} < 160\,\mathrm{GeV}$ No upper bound on $p_{\mathrm{T}}(\ell)$
WZ(3ℓ) ER	No upper bound on $M_{\mathrm{SFOS}}^{\mathrm{min}}(\ell\ell)$ No $M_{\mathrm{SFAS}}^{\mathrm{max}}(\ell\ell)$ requirement $p_{\mathrm{T}}(\ell_1) > 30\,\mathrm{GeV}$ $p_{\mathrm{T}}(\ell_{2,3}) > 10(15)\,\mathrm{GeV}$ (for e in high-MET bin) At least one μ with $p_{\mathrm{T}} > 20\,\mathrm{GeV}$ in low-MET bin
VV(2ℓ) VR	$M_{\mathrm{T}}(\ell_i, p_{\mathrm{T}}^{\mathrm{miss}}) > 90\,\mathrm{GeV}$ No $M_{\mathrm{T}}(\ell, p_{\mathrm{T}}^{\mathrm{miss}})$ requirement No upper bound on $p_{\mathrm{T}}(\ell)$

Table 4.9: Summary of the modifications to the SR event selection criteria for the control, enriched and validation regions of the analysis.

4.4.1 Nonprompt Background

The nonprompt background, which is the dominant background of the analysis, is estimated with the "tight-to-loose" method, introduced in Ref. [118]. The method, along with modifications designed especially for this search, is explained in the following paragraphs.

The goal of the method is to estimate the number of events with both leptons passing the tight ID of the analysis, while, in reality, at least one of the leptons is nonprompt. To this end, the probability of a prompt lepton and of a nonprompt lepton to pass the tight ID of the analysis given that they have passed the analysis loose ID, called prompt rate (PR) and fake rate (FR) respectively, are computed. In the following, we denote as N_{TT}, N_{TL}, N_{LT} and N_{LL} the number of events with two tight ID leptons, with one tight ID lepton and one loose-not-tight ID lepton (two cases depending on whether it is the leading or the subleading lepton the tight one) and with two loose-not-tight ID leptons respectively. We also denote as N_{PP}, N_{PN}, N_{NP} and N_{NN} the number of events with two prompt leptons, with one prompt lepton and one nonprompt lepton (two cases as above) and with two nonprompt leptons respectively. With these definitions, the following system of equations is formed, relating the ID of the leptons with their promptness:

$$\begin{pmatrix} N_{TT} \\ N_{TL} \\ N_{LT} \\ N_{LL} \end{pmatrix} = \begin{pmatrix} \mathrm{PR}_1\mathrm{PR}_2 & \mathrm{PR}_1\mathrm{FR}_2 & \mathrm{FR}_1\mathrm{PR}_2 & \mathrm{FR}_1\mathrm{FR}_2 \\ \mathrm{PR}_1(1-\mathrm{PR}_2) & \mathrm{PR}_1(1-\mathrm{FR}_2) & \mathrm{FR}_1(1-\mathrm{PR}_2) & \mathrm{FR}_1(1-\mathrm{FR}_2) \\ (1-\mathrm{PR}_1)\mathrm{PR}_2 & (1-\mathrm{PR}_1)\mathrm{FR}_2 & (1-\mathrm{FR}_1)\mathrm{PR}_2 & (1-\mathrm{FR}_1)\mathrm{FR}_2 \\ (1-\mathrm{PR}_1)(1-\mathrm{PR}_2) & (1-\mathrm{PR}_1)(1-\mathrm{FR}_2) & (1-\mathrm{FR}_1)(1-\mathrm{PR}_2) & (1-\mathrm{FR}_1)(1-\mathrm{FR}_2) \end{pmatrix} \begin{pmatrix} N_{PP} \\ N_{PN} \\ N_{NP} \\ N_{NN} \end{pmatrix} \tag{4.6}$$

Eq. 4.6 and the discussion below target the case of two leptons in the final state. The logic can be naturally extended to the case of three final state leptons. Nevertheless, the solution becomes arduously complicated and its detailed presentation is beyond the scope of this thesis.

The first of equations 4.14 yields the number of events with two tight ID leptons in SR, $N_{TT} = f(N_{PP}, N_{PN}, N_{NP}, N_{NN})$. However, the number of events with two prompt leptons, N_{PP}, is taken into account in the prompt background processes of the analysis, so they are removed from the nonprompt background estimation. Hence:

$$N_{TT}^{nonprompt} = \mathrm{PR}_1\mathrm{FR}_2 N_{PN} + \mathrm{FR}_1\mathrm{PR}_2 N_{NP} + \mathrm{FR}_1\mathrm{FR}_2 N_{NN} \tag{4.7}$$

However, the promptness of the leptons in the data cannot be known, so

N_{TT} must be expressed in terms of whether the leptons pass or fail the tight ID of the analysis, i.e. $N_{TT} = f(N_{TL}, N_{LT}, N_{LL})$. This is achieved by solving the rest of equations 4.14 to yield N_{PN}, N_{NP} and N_{NN} as a function of N_{TT}, N_{TL}, N_{LT} and N_{LL}. The final result is:

$$N_{TT}^{nonprompt} = W_{TT} * N_{TT} + W_{TL} * N_{TL} + W_{LT} * N_{LT} + W_{LL} * N_{LL} \quad (4.8)$$

$$\begin{aligned}
W_{TT} &= \frac{-\mathrm{PR_1FR_2}(1-\mathrm{FR_1})(1-\mathrm{PR_2}) - \mathrm{FR_1PR_2}(1-\mathrm{PR_1})(1-\mathrm{FR_2}) + (1-\mathrm{PR_1})(1-\mathrm{PR_2})\mathrm{FR_1FR_2}}{(\mathrm{PR1}-\mathrm{FR_1})(\mathrm{PR_2}-\mathrm{FR_2})} \\
W_{TL} &= \frac{\mathrm{PR_1FR_2PR_2}(1-\mathrm{FR_1})}{(\mathrm{PR_1}-\mathrm{FR_1})(\mathrm{PR_2}-\mathrm{FR_2})} \\
W_{LT} &= \frac{\mathrm{PR_1FR_1PR_2}(1-\mathrm{FR_2})}{(\mathrm{PR_1}-\mathrm{FR_1})(\mathrm{PR_2}-\mathrm{FR_2})} \\
W_{LL} &= \frac{-\mathrm{PR_1PR_2FR_1FR_2}}{(\mathrm{PR_1}-\mathrm{FR_1})(\mathrm{PR_2}-\mathrm{FR_2})}
\end{aligned} \quad (4.9)$$

This and the next paragraphs are dedicated to the explanation of the measurement of the PR, the FR and the number of events with at least one loose-not-tight lepton. The PR is defined as the number of prompt leptons passing the analysis tight ID divided by the total number of prompt leptons that have passed the loose ID of the analysis. It is computed from simulation in a mixture of events with prompt leptons originating from Z or W bosons or from τ leptons, as the prompt background processes of the analysis. The measurement is performed separately for muons and electrons in bins of p_{T} and η. While the p_{T} binning is common for all leptons, the η bins are different for muons ($[0, 1.2]$ for the barrel bin and $[1.2, 2.4]$ for the endcap bin) and electrons ($[0, 1.5]$ for the barrel bin and $[1.5, 2.5]$ the endcap bin). Fig. 4.19 the PR separately for the η bins of muons and electrons and for different years overlaid.

The FR is measured as a function of the lepton p_{T} and η in a region enriched in nonprompt leptons, called measurement region (MR). For the FR measurement of muons, single muon trigger paths are used. The measurement for p_{T} greater than 10 GeV is performed with a trigger path capturing events with at least, one not necessarily isolated, muon with online $p_{\mathrm{T}} > 8$ GeV (`HLT_Mu8`), while the measurement below 10 GeV uses events with at least one, not necessarily isolated, muon in combination with a jet with online $p_{\mathrm{T}} > 40$ GeV, so that the muon online p_{T} threshold is reduced to

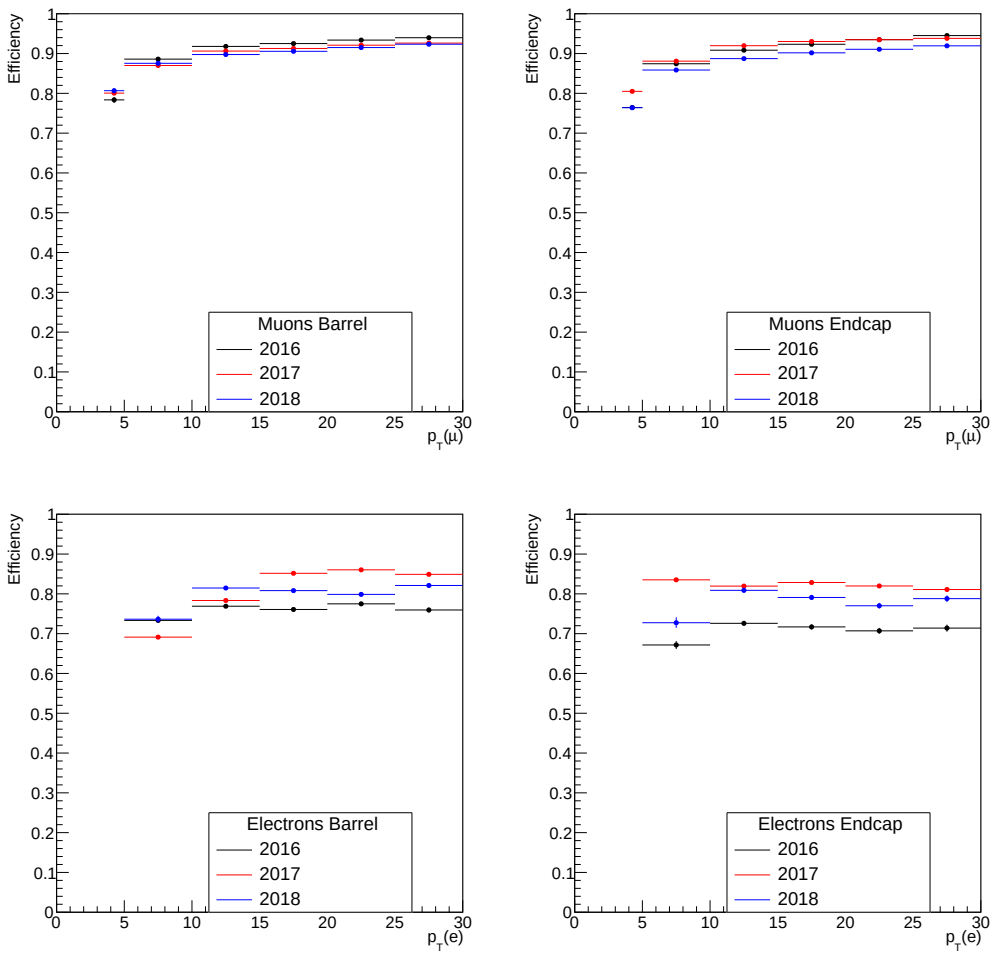


Figure 4.19: The PR of muons (upper row) and electrons (lower row) in the barrel (left) and endcap (right) for 2016 (black line), 2017 (red line) and 2018 (blue line), as a function of the lepton p_T.

3 GeV (`HLT_Mu3_PFJet40`). The MR for electrons utilizes a combination of trigger paths requiring at least one jet with variable p_T threshold for different years (`HLT_PFJet25` for 2018 and `HLT_PFJet40` for 2016 and 2017). On account of the trigger selections, the MR is defined with slight differences for muons and electrons. Event of the MR are selected if they contain exactly one jet with p_T greater than 50 GeV for muons and 30(40) GeV for electrons in 2018(2016 and 2017). The enrichment in nonprompt leptons is achieved

by requiring exactly one of loose-not-tight ID lepton, which is separated by the jet with a $\Delta R > 0.7$.

As a result of the MR selection, the MRs for both muons and electrons are dominated by QCD events, in which the FR should be measured. However, there is also a nonnegligible contamination from events with prompt leptons, such as W+jets and, to lower extent, Z+jets, that need to be subtracted. A useful handle to discriminate the QCD from the EW processes is the $M_{\mathrm{T}}(\ell, p_{\mathrm{T}}^{\mathrm{miss}})$ variable:

$$M_{\mathrm{T}}(\ell, p_{\mathrm{T}}^{\mathrm{miss}}) = \sqrt{2 \times p_{\mathrm{T}}(\ell) p_{\mathrm{T}}^{\mathrm{miss}} \left(1 - \cos \Delta\phi(\ell, p_{\mathrm{T}}^{\mathrm{miss}})\right)} \tag{4.10}$$

However, the $M_{\mathrm{T}}(\ell, p_{\mathrm{T}}^{\mathrm{miss}})$ is correlated with the lepton p_{T}. At the same time, the FR is also a function of the lepton p_{T}. To reduce the correlation between the $M_{\mathrm{T}}(\ell, p_{\mathrm{T}}^{\mathrm{miss}})$ selection for subtracting the EW processes and the FR, which could introduce nontrivial bias in the measurement of the latter, a modified definition of $M_{\mathrm{T}}(\ell, p_{\mathrm{T}}^{\mathrm{miss}})$. The modified variable, $M_{\mathrm{T}}^{\mathrm{fix}}(\ell, p_{\mathrm{T}}^{\mathrm{miss}})$, removes the explicit dependence on $p_{\mathrm{T}}(\ell)$ by using the most probable p_{T} value of a lepton coming from a W boson decay:

$$M_{\mathrm{T}}^{\mathrm{fix}}(\ell, p_{\mathrm{T}}^{\mathrm{miss}}) = \sqrt{2 \times 35 \times p_{\mathrm{T}}^{\mathrm{miss}} \left(1 - \cos \Delta\phi(\ell, p_{\mathrm{T}}^{\mathrm{miss}})\right)} \tag{4.11}$$

The $M_{\mathrm{T}}^{\mathrm{fix}}(\ell, p_{\mathrm{T}}^{\mathrm{miss}})$ distributions for the different processes contributing to the MR for muons in 2018 is shown in Fig. 4.20.

Three different methods are used for the subtraction of EW processes and the FR measurement in data.The first method ("cut & sub") involves selecting only the events with $M_{\mathrm{T}}^{\mathrm{fix}}(\ell, p_{\mathrm{T}}^{\mathrm{miss}}) < 20\,\mathrm{GeV}$. This phase space has a very high purity in QCD events ($> 90\%$) and the residual EW processes are subtracted using the estimation from simulation. The FR is then calculated with the formula:

$$\mathrm{FR}(p_{\mathrm{T}}, \eta) = \frac{\text{Events with nonprompt lepton passing the analysis tight ID}}{\text{Events with nonprompt lepton passing the analysis loose ID}} \tag{4.12}$$

The second method ("unfolded") exploits both the low and high range of $M_{\mathrm{T}}^{\mathrm{fix}}(\ell, p_{\mathrm{T}}^{\mathrm{miss}})$. The range of small (S) values, $0 < M_{\mathrm{T}}^{\mathrm{fix}}(\ell, p_{\mathrm{T}}^{\mathrm{miss}}) < 20\,\mathrm{GeV}$, drives the calculation of the FR, contributing to it a term FR_S. However, the subtraction of the EW contribution is now performed by utilizing the

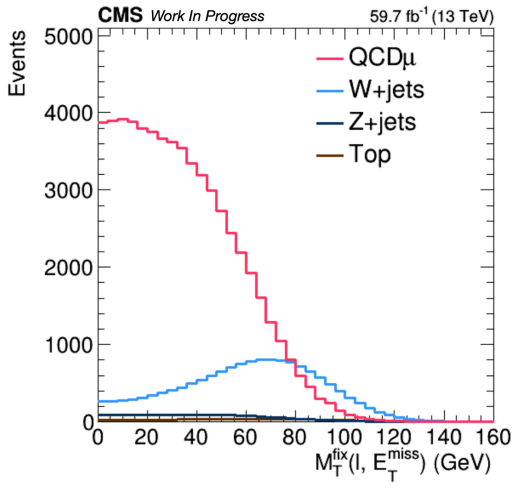


Figure 4.20: Overlaid $M_{\mathrm{T}}^{\mathrm{fix}}(\ell, p_{\mathrm{T}}^{\mathrm{miss}})$ distributions for the different processes of the MR for muons in 2018.

range of large (L) values, $70 < M_{\mathrm{T}}^{\mathrm{fix}}(\ell, p_{\mathrm{T}}^{\mathrm{miss}}) < 120\,\mathrm{GeV}$, which is rich in EW processes. The FR in this range, denoted as FR_L, is subtracted from FR_S, after it has been weighted with the relative S to L EW simulation yields ($N^{\text{prompt MC}}$), corrected by the data (N_S^{data}). Putting everything together, the FR in this method is calculated by the formula:

$$\mathrm{FR}(p_{\mathrm{T}}, \eta) = \frac{\mathrm{FR}_S - \mathrm{FR}_L \times r}{1 - r}, \text{ with } r = \frac{N_S^{\text{prompt MC}} / N_L^{\text{prompt MC}}}{N_S^{\mathrm{data}} / N_L^{\mathrm{data}}} \tag{4.13}$$

The third method ("sim. fit") takes into account the $M_{\mathrm{T}}^{\mathrm{fix}}(\ell, p_{\mathrm{T}}^{\mathrm{miss}})$ distribution as a whole. The measurement of the FR is achieved by simultaneously fitting the $M_{\mathrm{T}}^{\mathrm{fix}}(\ell, p_{\mathrm{T}}^{\mathrm{miss}})$ distribution of events with leptons passing the analysis tight ID (numerator of Eq. 4.12) and failing the analysis tight ID but passing the loose ID (denominator of Eq. 4.12 excluding the numerator). This method takes advantage of the QCD and EW templates from simulation. The fits are performed in bins of lepton p_{T} and η and the FR is calculated after the fits with Eq. 4.12.

The three different methods yield similar FR results within their uncertainties, as shown in Figs. B.1-B.3 of Appx. B. The final FR used in the analysis is the weighted average of the FR calculated on data with the methods mentioned above, while its uncertainty is given by the envelope of the FR

uncertainties of the three methods. The results of the FR measurement for muons and electrons in the different η bins is shown as a function of the lepton p_{T} in Figs. 4.21–4.23. The MC simulation FR measurements are shown as a reference and only the DD FR measurements are used in the analysis.

As already mentioned, the main sources of nonprompt background in the SRs are the W+jets and the $\mathrm{t\bar{t}}$ processes. The FR measured in the MR can depend on the flavor composition of the QCD events selected. Consequently, flavor-dependent differences in the measurement of the FR can potentially bias the nonprompt estimation of the different processes in the SRs. To minimize this effect, the analysis loose-not-tight ID, introduced in Secs. 4.3.1 and 4.3.2 for muons and electrons respectively, has been specifically tuned to harmonize the behavior of the FR for leptons originating from different flavor quarks. The results of this minimization are presented in Figs. B.4-B.6 and show good agreement between leptons coming from different quark flavor decays.

Another potential source of bias in the nonprompt background estimation method is dependence on the p_{T} of the mother quark of the lepton. The reason for this is that the mother quark p_{T} spectrum can be different in the QCD events of the MR and the nonprompt W+jets and $\mathrm{t\bar{t}}$ events of the SRs. To reduce the mother quark p_{T} dependence, the hyperbolic $\mathrm{Iso}_{\mathrm{abs}}$ requirement, mentioned in Table 4.5 and 4.6 for muons and electrons respectively, was found to be the optimal choice, since it ensures approximately constant energy in the lepton isolation cones.

The final ingredient for the tight-to-loose is the number of events with at least one loose-not-tight lepton. These events are taken from regions, called application regions (ARs), with the same selection as the SRs, but with the requirement of at least one loose-not-tight lepton replacing the requirement of 2(3) tight ID leptons applied to the $2\ell(3\ell)$ SRs. This change enriches the ARs in nonprompt background. More specifically, the numbers for N_{TL}, N_{LT} and N_{LL} are taken from mutually exclusively subsets of data events that comprise the ARs, called AR sidebands, with only the leading lepton passing the loose-not-tight ID, only the subleading lepton passing the loose-not-tight ID and both leptons passing the loose-not-tight ID respectively. The N_{TT} is taken from the data events of the SRs. To this end, the ARs follow the same binning strategy as the SRs of the analysis. The DD tight-to-loose method takes all of the inputs described above and uses Eqs. 4.8 and 4.9 to estimate the nonprompt background in the SRs, separately for each year.

Even though the nonprompt background estimation is performed for each

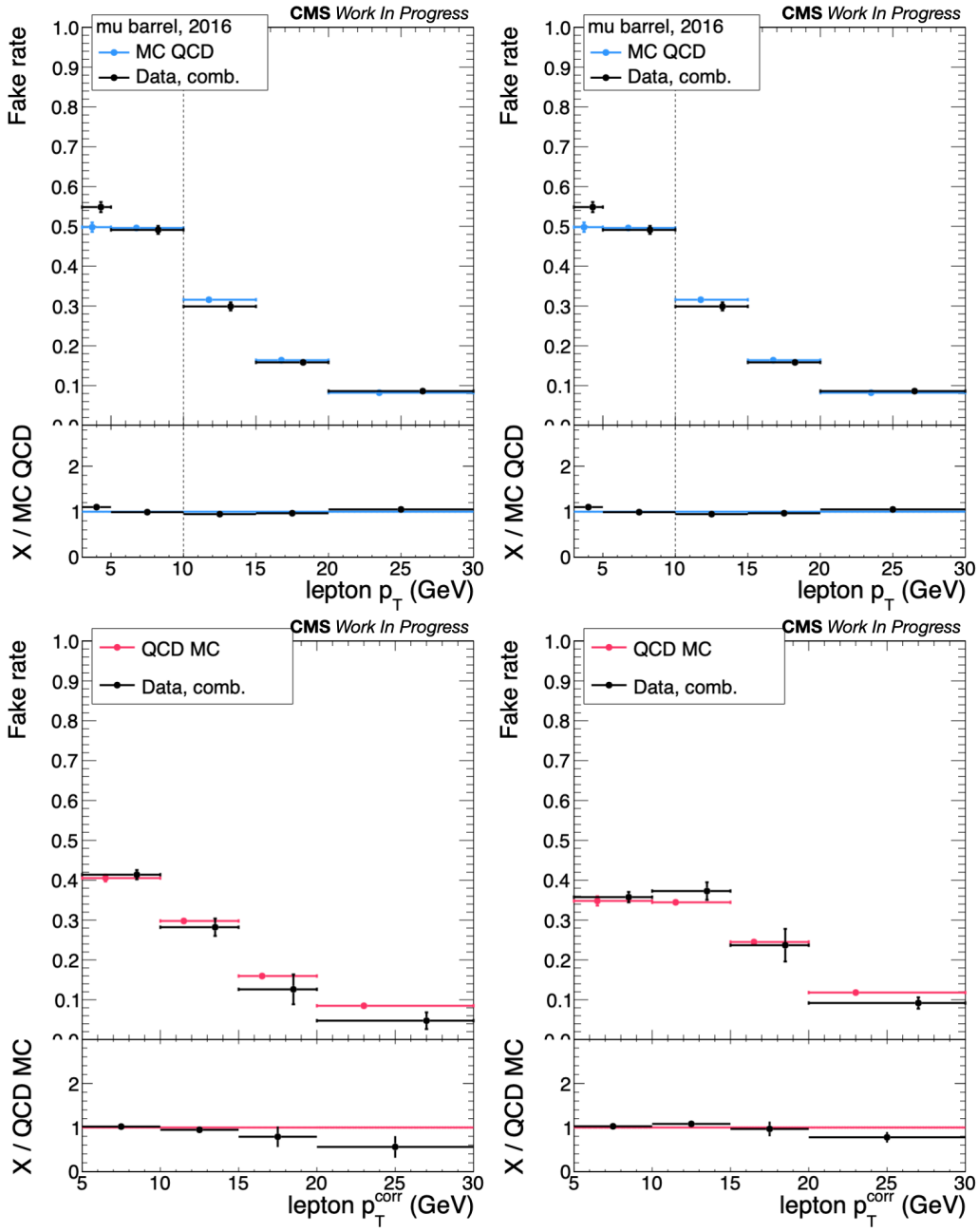


Figure 4.21: The FR of muons (upper row) and electrons (lower row) in the barrel (left) and endcap (right) for 2016, as a function of the lepton p_{T}.

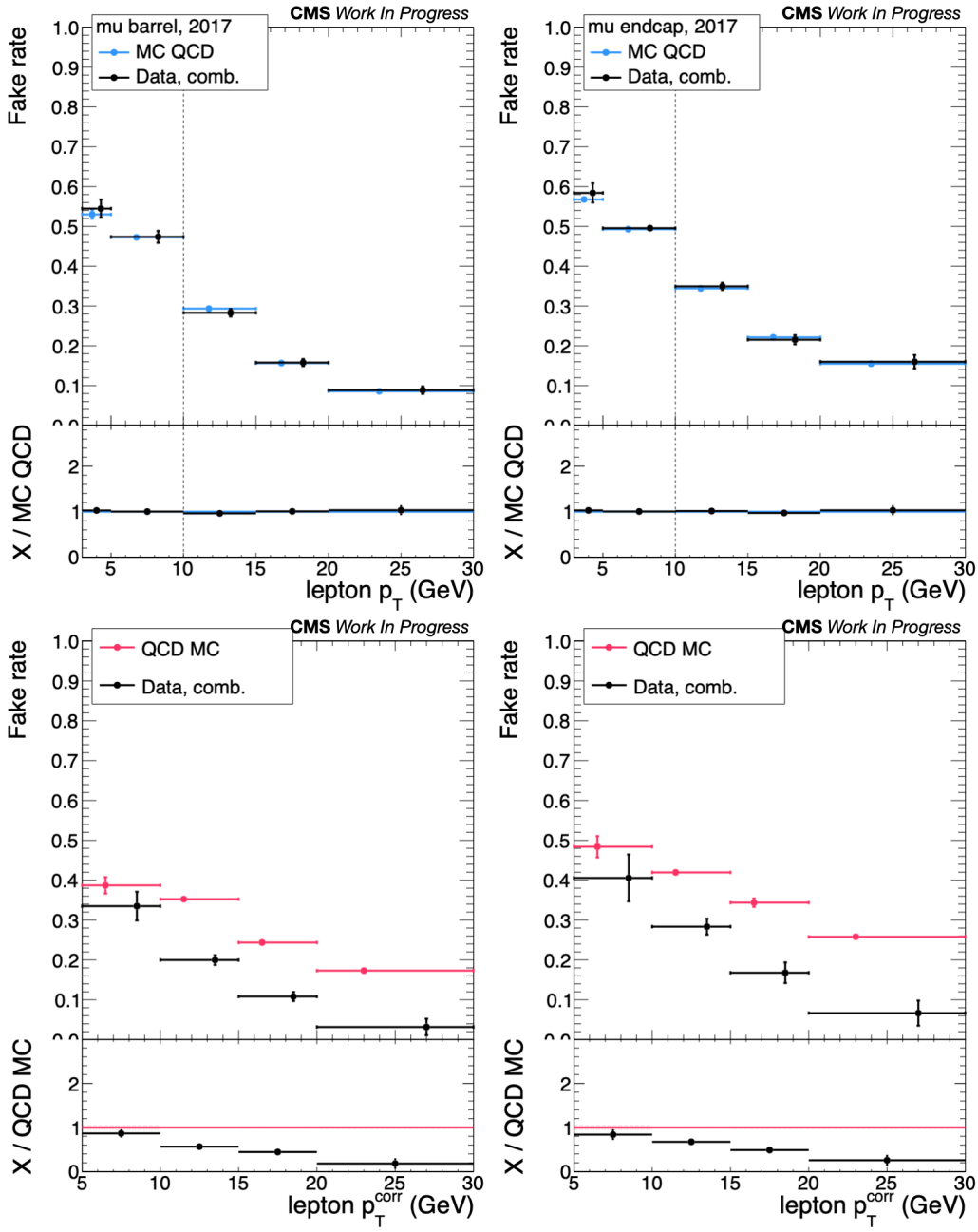


Figure 4.22: The FR of muons (upper row) and electrons (lower row) in the barrel (left) and endcap (right) for 2017, as a function of the lepton p_{T}.

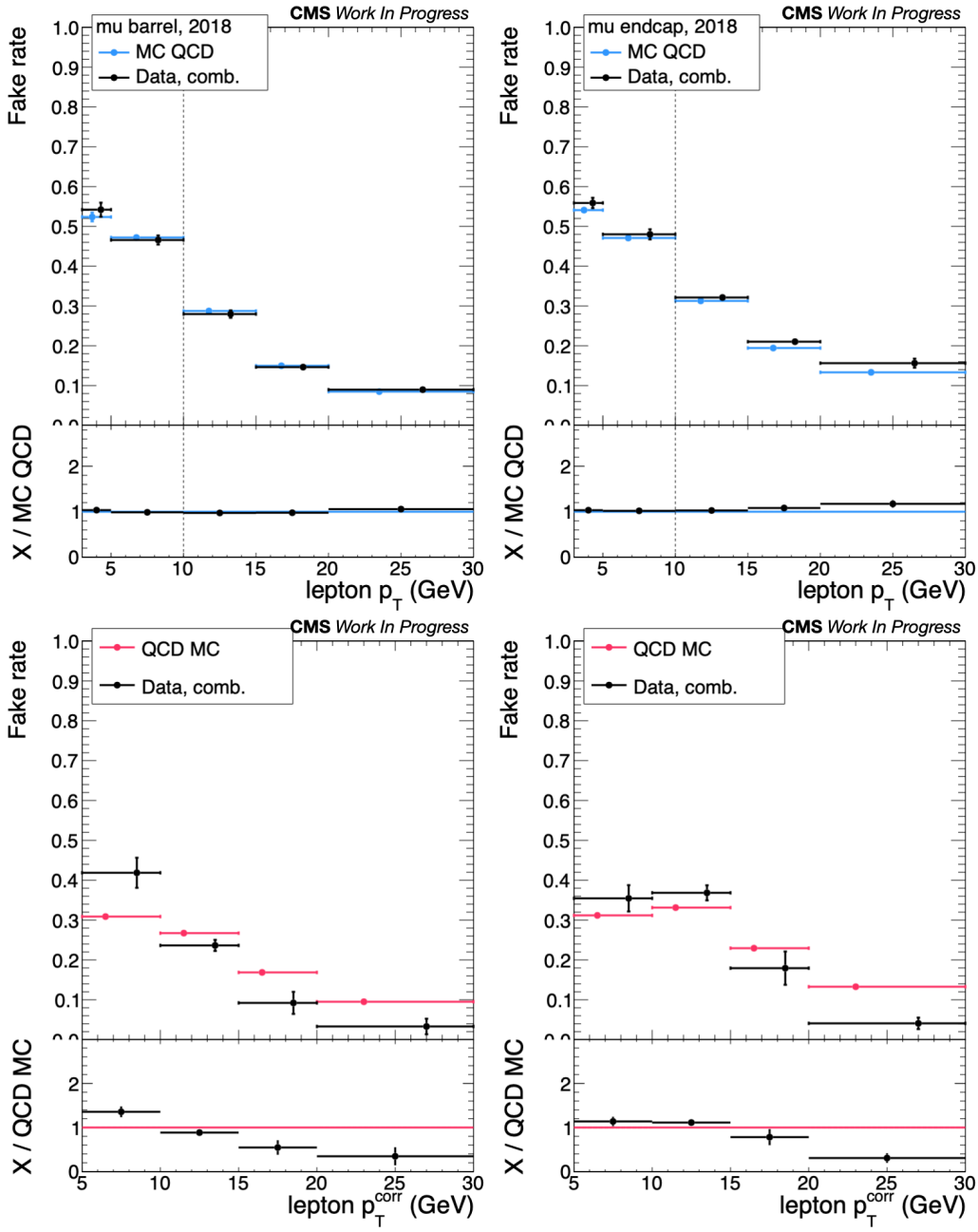


Figure 4.23: The FR of muons (upper row) and electrons (lower row) in the barrel (left) and endcap (right) for 2018, as a function of the lepton p_{T}.

year separately, Figs. 4.24, 4.25 and 4.26 show the year-inclusive ARs corresponding to the 2ℓ-EWK, the 3ℓ-EWK and the 2ℓ-Stop SRs respectively, in order to showcase their general features. A first observation is that a data-simulation disagreement is evident in the low-MET bin of the 2ℓ ARs. This disagreement, which progressively vanishes in the higher-MET bins, is attributed to the the absence of QCD simulation. The reason for the omission of the QCD simulation in the ARs is that it would require an extremely large number, $\mathcal{O}(10^9)$, of QCD simulation events to get the necessary yield in the phase space that this analysis probes. However, such a computational investment is not essential, since the DD tight-to-loose method only uses the data of the ARs and does not rely at all on information from the simulation.

The higher-MET bins of the 2ℓ SRs show good agreement between data and simulation, since the requirement for $p_{\mathrm{T}}^{\mathrm{miss,corr}} > 200\,\mathrm{GeV}$ significantly reduces the QCD contribution. However, especially in the high- and ultra-MET bins, the yield in the ARs is small and becomes even smaller when each AR is split in its sidebands for the application of the tight-to-loose method. As a consequence, the purely DD nonprompt background estimation becomes sensitive to statistical fluctuations in the data. To add to that, Eq. 4.9 implies that upward fluctuations in N_{LL}, alone or in combination with downward fluctuations in N_{TL} and N_{LT}, could lead to a negative yield estimation, since W_{LL} has a negative contribution.

To resolve the low-yield issue in the higher-MET SRs of the analysis, a modification of the tight-to-loose method is implemented. Since a good agreement between data and simulation is observed in those AR MET bins, information from the simulation can be exploited. More specifically, the modified tight-to-loose method, called semi-DD method, makes the most out of the simulation, which has better statistical accuracy than the data in those regions, by employing it to model the nonprompt background shape. The data are still used to fix the nonprompt background normalization and for the calculation of the FR. As a result, the semi-DD method combines the robustness and reliability of the data in essential parts of the nonprompt background estimation, while taking advantage of the simulation to reduce statistical fluctuations.

In the case of the semi-DD method, the extra information that is present in the simulation is utilized to simplify the Eq. 4.6 and, consequently, also Eqs. 4.8 and 4.9. The simulation provides details about the promptness of the leptons, hence the PR of the leptons can be set to unity, yielding:

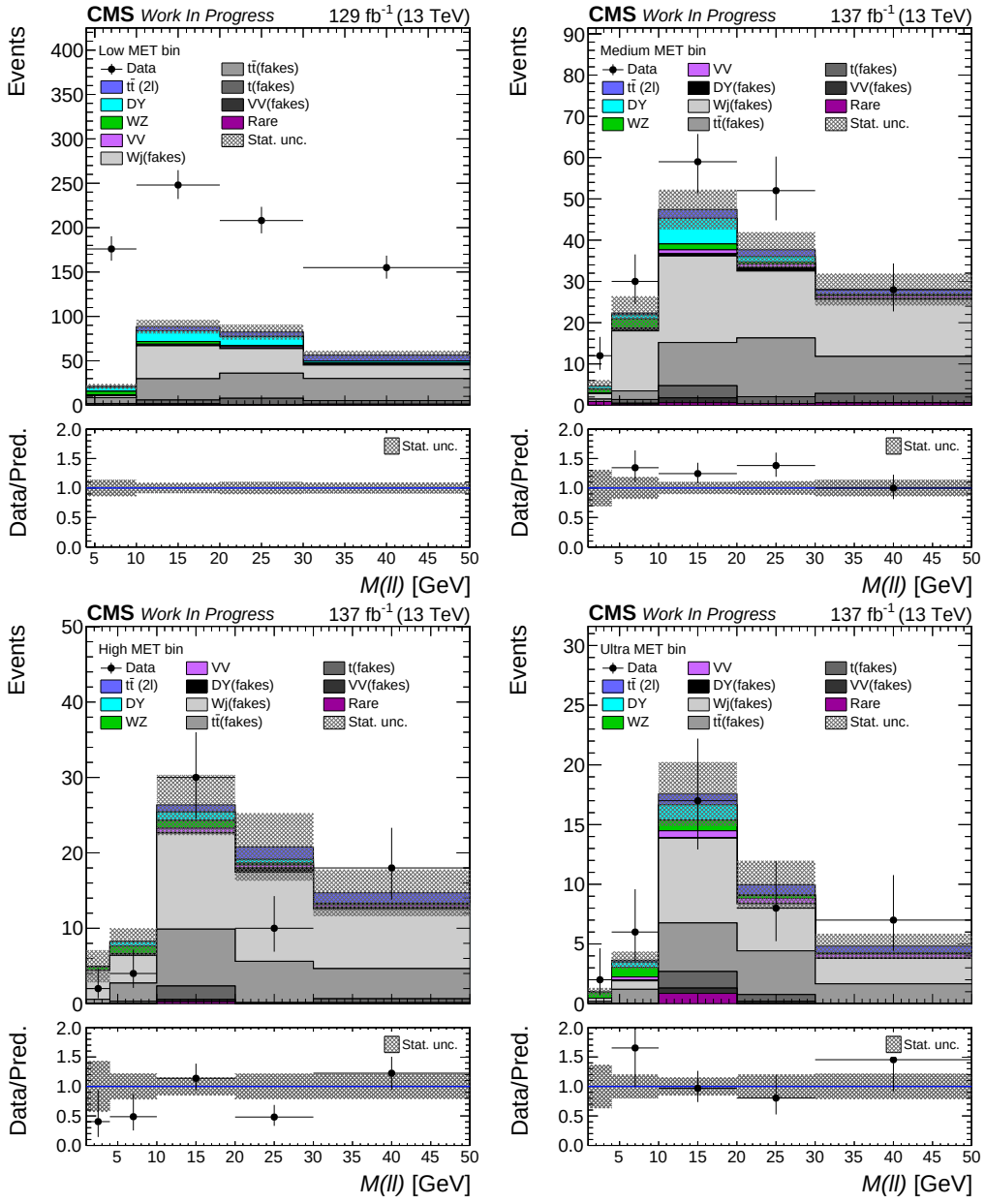


Figure 4.24: The uncorrected MC distribution of the $M(\ell\ell)$ variable is shown for the ARs corresponding to the 2ℓ-EWK SRs with the full Run 2 yields combined in the low-MET (upper left), medium-MET (upper right), high-MET (lower left) and ultra-MET (lower right) bin. The uncertainty band includes only the statistical component.

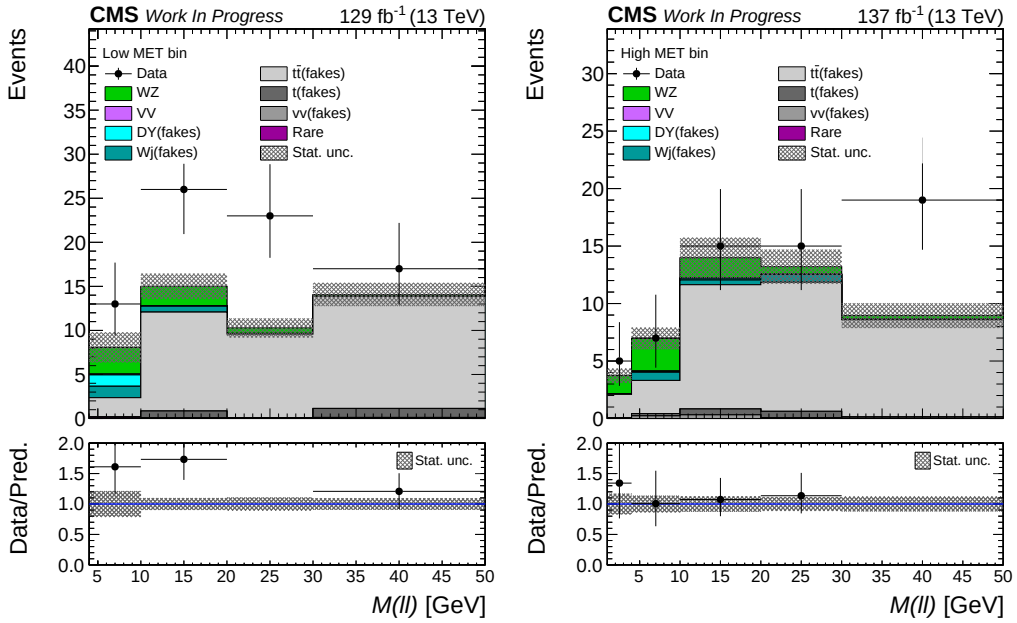


Figure 4.25: The uncorrected MC distribution of the $M_{\mathrm{SFOS}}^{\mathrm{min}}(\ell\ell)$ variable is shown for the ARs corresponding to the 3ℓ-EWK SRs with the full Run 2 yields combined in the low-MET (left) and high-MET (right) bin. The uncertainty band includes only the statistical component.

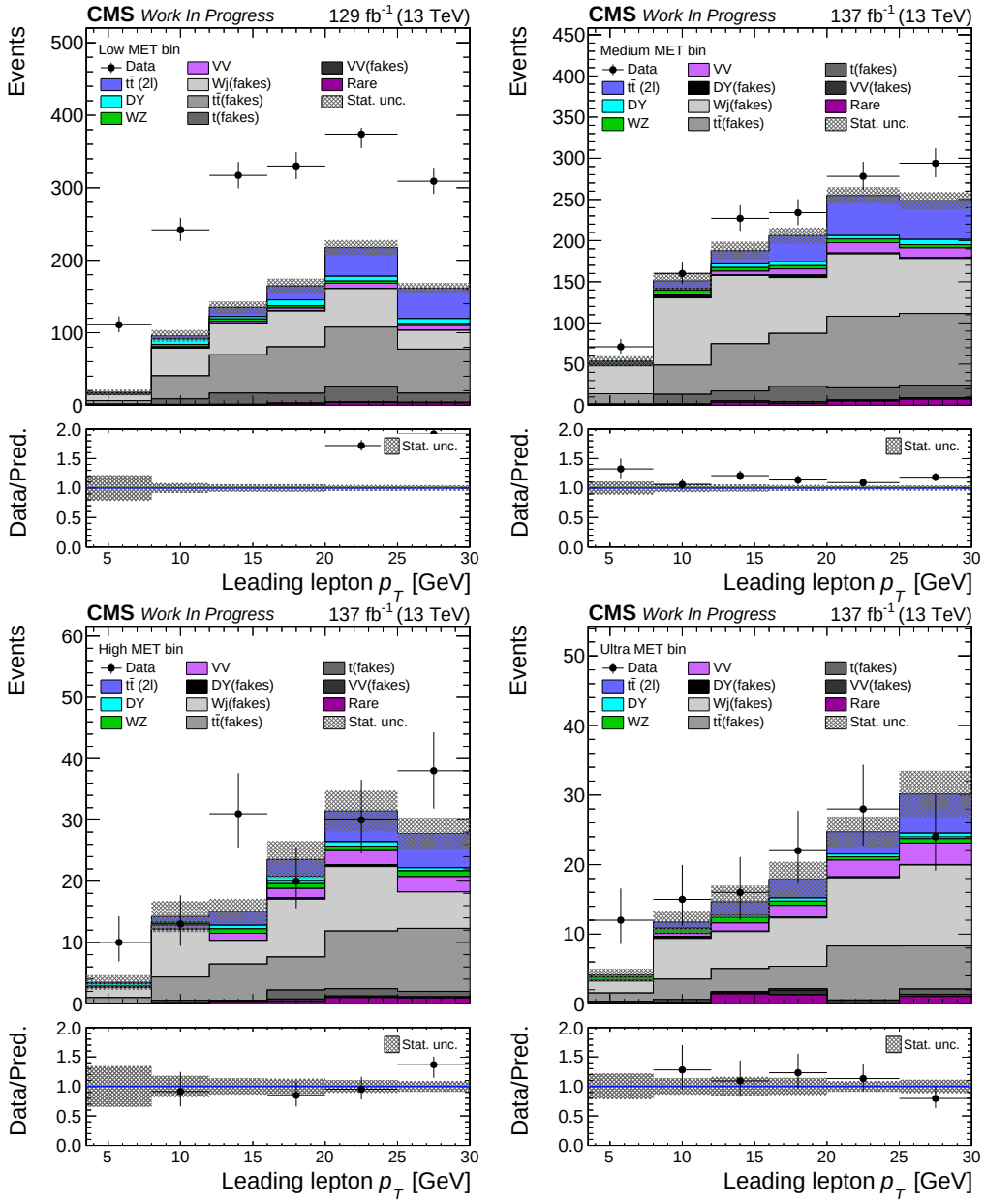


Figure 4.26: The uncorrected MC distribution of the $p_{\mathrm{T}}(\ell_1)$ variable is shown for the ARs corresponding to the 2ℓ-Stop SRs with the full Run 2 yields combined in the low-MET (upper left), medium-MET (upper right), high-MET (lower left) and ultra-MET (lower right) bin. The uncertainty band includes only the statistical component.

$$\begin{pmatrix} N_{TT} \\ N_{TL} \\ N_{LT} \\ N_{LL} \end{pmatrix} = \begin{pmatrix} 1 & \mathrm{FR}_2 & \mathrm{FR}_1 & \mathrm{FR}_1\mathrm{FR}_2 \\ 0 & (1-\mathrm{FR}_2) & 0 & \mathrm{FR}_1(1-\mathrm{FR}_2) \\ 0 & 0 & (1-\mathrm{FR}_1) & (1-\mathrm{FR}_1)\mathrm{FR}_2 \\ 0 & 0 & 0 & (1-\mathrm{FR}_1)(1-\mathrm{FR}_2) \end{pmatrix} \begin{pmatrix} N_{PP} \\ N_{PN} \\ N_{NP} \\ N_{NN} \end{pmatrix} \tag{4.14}$$

Eq. 4.8 still holds with:

$$\begin{aligned} W_{TT} &= 0 \\ W_{TL} &= \frac{\mathrm{FR}_2}{(1-\mathrm{FR}_2)} \\ W_{LT} &= \frac{\mathrm{FR}_1}{(1-\mathrm{FR}_1)} \\ W_{LL} &= \frac{-\mathrm{FR}_1\mathrm{FR}_2}{(1-\mathrm{FR}_1)(1-\mathrm{FR}_2)} \end{aligned} \tag{4.15}$$

As in the case of the DD tight-to-loose method, N_{TL}, N_{LT} and N_{LL} are computed from the AR sidebands. However, in the semi-DD method, the MC templates of the processes with nonprompt leptons of each AR sideband (1 loose-not-tight, 2 loose-not-tight or 3 loose-not-tight, in the 3ℓ case) are first normalized to the AR data with SFs, called "semi-DD SFs", that are computed as follows:

$$\text{Semi-DD SF} = \frac{\text{AR Data} - \text{AR Prompt MC}}{\text{AR Nonprompt MC}} \tag{4.16}$$

The events of normalized-to-data MC templates are then weighted with the proper FR combinations, as dictated by the weights of Eq. 4.15. The resulting nonprompt background estimation is free of large scale statistical fluctuations in most MET bins. Extra systematic uncertainties are applied to account potential mismodeling in the simulation shape.

Nevertheless, a low simulation yield is still observed for the medium-, high- and ultra-MET bins of the 2ℓ-EWK ARs. To mitigate this effect, for those regions only, the shape of the addition of the yields of those MET bins is used. After the MC shape of the merged medium+high+ultra MET bins has been extracted, the template is normalized to the yields of each respective MET bin, before the FR weights are applied. This is done by the so-called "semi-DD rate factors (RF)", which are calculated by:

$$\text{Semi-DD RF} = \frac{\text{AR Nonprompt MC in individual MET bin}}{\text{AR Nonprompt MC in merged MET bin}} \tag{4.17}$$

With this supplementary step, the statistical uncertainties in all three of medium-, high- and ultra-MET bins shrink significantly. Any possible differences in the MC shape among the different MET bins is considered by an additional systematic uncertainty for those bins.

To summarize, the DD tight-to-loose method is utilized for the nonprompt background estimation in the low-MET 2ℓ SRs and all of the CRs, except the SS CR, of the analysis. The semi-DD method, a modification to the original method, exploiting information from the simulation, is implemented in the 3ℓ and the higher-MET bins of 2ℓ SRs, as well as the SS CR.

As a check for the compatibility of the FR measured in the MR with the expected SR estimation of nonprompt events, a "closure test" is performed. For this test, the dominant sources of the nonprompt background in the SRs, W+jets and $\mathrm{t\bar{t}}$, are taken from simulation and are compared to the FR weighted W+jets and $\mathrm{t\bar{t}}$ simulation events of the ARs. The outcome of the closure is shown in Fig. 4.27, where the nonprompt simulation events of the SRs are labeled as "Fakes", while the FR weighted simulation events of the ARs are labeled as "FR QCD MC Fakes". The two estimates agree within their statistical uncertainties. Potential differences are taken into account in the systematic uncertainties of the analysis.

A specially designed CR is also used to evaluate the nonprompt background modeling. The SS CR follows the event selection of the 2ℓ-Stop SRs but requires that the two leptons are of the same charge, instead of opposite. Given that processes leading to dilepton final states with same-sign leptons are very rare in the SM, the SS CR is enriched in nonprompt background. It is defined only for $p_{\mathrm{T}}^{\mathrm{miss,corr}} > 200\,\mathrm{GeV}$, due to the double-$\mu$+MET trigger path requirement for opposite-sign muons. The uncorrected $M(\ell\ell)$ distribution of the SS CR, taken purely from simulation, is shown in Fig. 4.28. DD SF are extracted to correct the SS CR simulation through a simultaneous maximum likelihood fit of multiple regions of the analysis, which is described in great detail in Sec. 4.6.

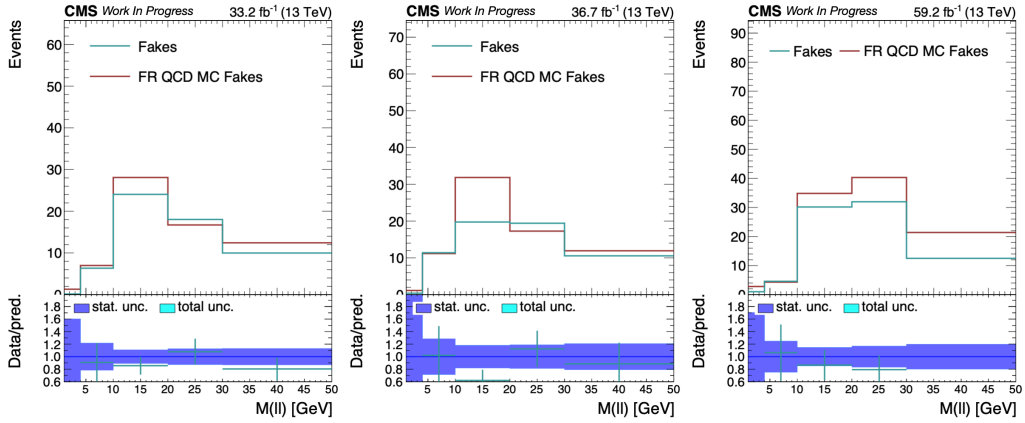


Figure 4.27: The result of the nonprompt background estimation method closure test in 2016 (left), 2017 (middle) and 2018 (right). The comparison between the nonprompt simulation events of the SRs ("Fakes") and the FR weighted simulation events of the ARs ("FR QCD MC Fakes") is shown, as a function of the $M(\ell\ell)$ variable.

4.4.2 $\mathrm{t\bar{t}}$ Process

The $\mathrm{t\bar{t}}$ events where the two W bosons, coming from $\mathrm{t} \to \mathrm{bW}$ process, decay leptonically can lead to final states that are very similar to those of the SUSY signal models probed in this search. However, due to the presence of two bottom quarks in the final state, the contribution of the $\mathrm{t\bar{t}}$ process in the SRs is effectively reduced by the b jet veto applied there. The $M_{\mathrm{T}}(\ell, p_{\mathrm{T}}^{\mathrm{miss}})$ upper bound also helps towards this direction. As a result of the SR selection, an orthogonal CR dominated by $\mathrm{t\bar{t}}$ events, denoted as "$\mathrm{t\bar{t}}$ CR", can be constructed by inverting the b jet veto, hence requiring $N_{\mathrm{b}}(p_{\mathrm{T}} > 25\,\mathrm{GeV}) \geq 1$. The $\mathrm{t\bar{t}}$ CR is further enriched by removing the $M_{\mathrm{T}}(\ell, p_{\mathrm{T}}^{\mathrm{miss}})$ requirement and the upper bound on lepton p_{T}. This selection leads to a high purity of $\mathrm{t\bar{t}}$ events in the $\mathrm{t\bar{t}}$ CR, reaching $\sim$90% and $\sim$78% in the low-MET and high-MET bin, respectively, for each year. The uncorrected $M(\ell\ell)$ distributions of the $\mathrm{t\bar{t}}$ CR, taken purely from simulation, is shown in Fig. 4.29 for the different MET bins. The normalization of the distribution in each MET bin and each year of the $\mathrm{t\bar{t}}$ CR is separately corrected by DD SF determined by the simultaneous maximum likelihood fit. These SFs range from 0.9–1.1 and 0.8–1.2 in the low- and high-MET bins for different years.

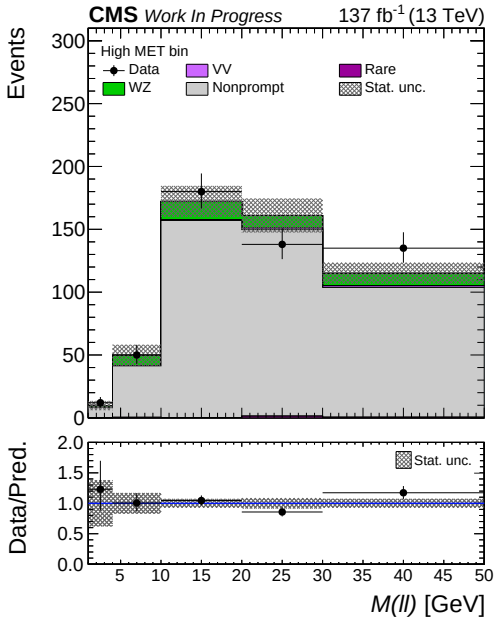


Figure 4.28: The uncorrected MC distribution of the $M(\ell\ell)$ variable is shown for the SS CR with the full Run 2 yields combined. The uncertainty band includes only the statistical component.

4.4.3 DY Process

The main handle for suppressing the DY process contributing to the analysis SM background via the $Z \to \tau\tau$ decay is the $M_{\tau\tau}$ variable, defined and explained in Sec. 4.3.4. The inversion of this SR requirement of this variable, imposing $0 < M_{\tau\tau} < 160\,\text{GeV}$, is sufficient to construct an orthogonal to the SRs and DY enriched region, denoted as "DY CR". To allow for more yields in the DY CRs, the upper bound on lepton p_T is also removed. The purity in DY events in these regions is $\sim 70\%$ in both low- and high-MET bins for each year. The uncorrected $M(\ell\ell)$ distributions of the DY CR, taken purely from simulation, is shown in Fig. 4.30 for the different MET bins. The DD normalization SFs for the DY CRs, computed in the same way as the ones for the $t\bar{t}$ CRs, are in the range 1.2–1.7 and 1.2–1.4 in the low- and high-MET bins for different years. The high SFs of the DY CR, especially in the low-MET bin, originate from the fact that the instrumental MET, which plays an important role in DY events, is not well-modeled in simulation and, hence, a more significant correction is needed to harmonize it with the data.

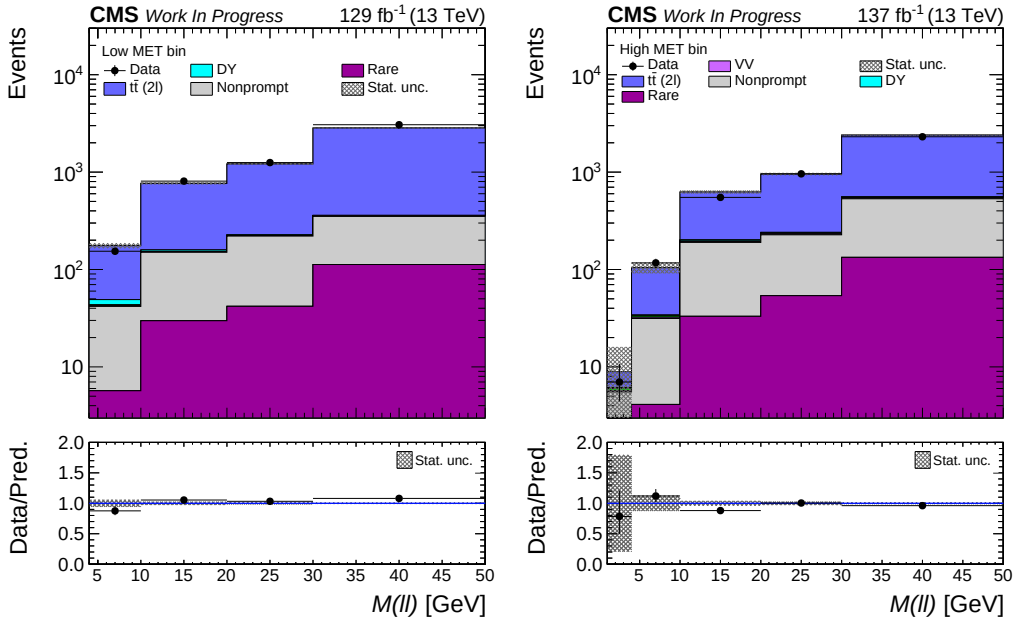


Figure 4.29: The uncorrected MC distribution of the $M(\ell\ell)$ variable is shown for the low-MET (left) and high-MET (right) bins of the $\mathrm{t\bar{t}}$ CR with the full Run 2 yields combined. The uncertainty band includes only the statistical component.

4.4.4 Fully Leptonic WZ Process

Given that the WZ process with a fully leptonic final state is the dominant prompt background for the 3ℓ SRs, it is critical that its contribution is kept under control by the exclusion of the Z boson mass window from the $M_{\mathrm{SF}}(\ell\ell)$ distribution. By removing the upper bounds on $M_{\mathrm{SFOS}}^{\mathrm{min}}(\ell\ell)$ and $M_{\mathrm{SFAS}}^{\mathrm{max}}(\ell\ell)$, a region enriched in fully leptonic WZ events can be obtained, denoted as "WZ ER". The orthogonality with the SRs is ensured by inverting the upper bound of $p_{\mathrm{T}}(\ell_1)$, requiring it to be greater than 30 GeV. The lepton bounds for the subleading and trailing leptons are also modified to increase the purity of the WZ ER, achieving purities as high as $\sim 80\%$ and $\sim 90\%$ in the low-MET and high-MET bins respectively. Due to the online requirement of the double-μ+MET trigger path for $M_{\mathrm{SF}}(\ell\ell) < 56$ GeV, the complimentary pure double-μ trigger paths, for which such a selection is not applied, are used in the low-MET bin of the WZ ER.

Due to the presence of nonnegligible signal yield in the low $M_{\mathrm{SFOS}}^{\mathrm{min}}(\ell\ell)$ bins

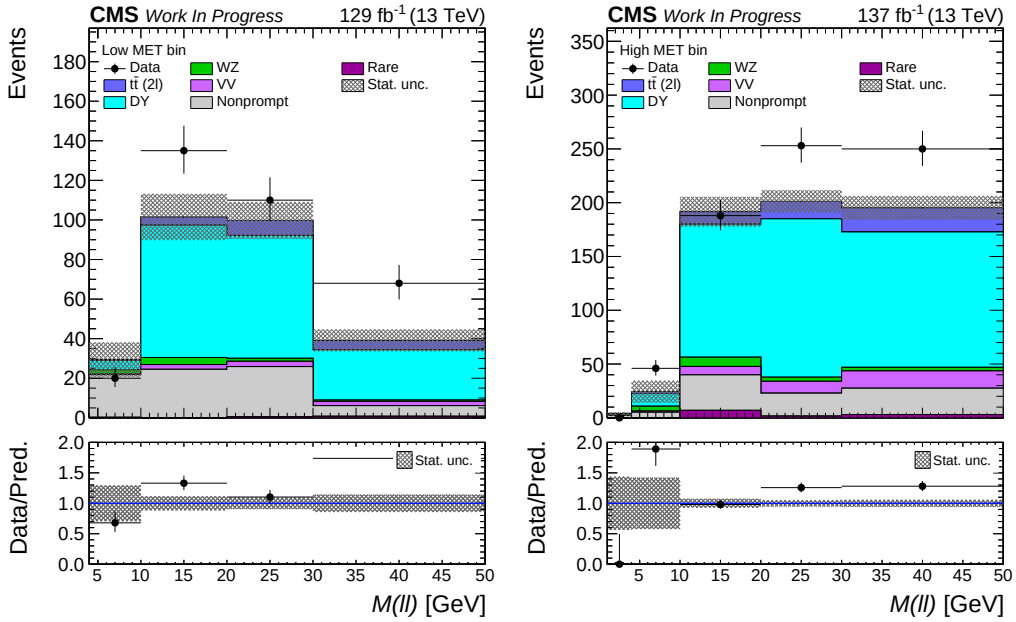


Figure 4.30: The uncorrected MC distribution of the $M(\ell\ell)$ variable is shown for the low-MET (left) and high-MET (right) bins of the DY CR with the full Run 2 yields combined. The uncertainty band includes only the statistical component.

of the WZ ER, especially for masspoints with $\Delta m > 30\,\text{GeV}$, the WZ ER is divided in two subregions: The "WZ-like selection SR", for $M^{\text{min}}_{\text{SFOS}}(\ell\ell) < 30\,\text{GeV}$ and the "WZ CR" for $M^{\text{min}}_{\text{SFOS}}(\ell\ell) > 30\,\text{GeV}$. The uncorrected $M(\ell\ell)$ distributions of the WZ ER, which are taken purely from simulation and are presented in Fig. 4.31 for the different MET bins, are corrected with DD SFs, similarly to the other prompt CRs. These SFs generally vary from 0.6 to 0.8, depending on the MET bin and the year, and are lower than unity due to overestimation of the WZ normalization by the simulation using the POWHEG generator with the $M^{\text{min}}_{\text{SFOS}}(\ell\ell)$ variable reaching down to 0.1 GeV.

4.4.5 Other Diboson Processes

As already mentioned, diboson processes other than the fully leptonic WZ one, that is the WW, the ZZ and the partially leptonic WZ processes, are collectively labeled as VV. These processes constitute a subleading prompt background for the analysis and their contribution is estimated from simu-

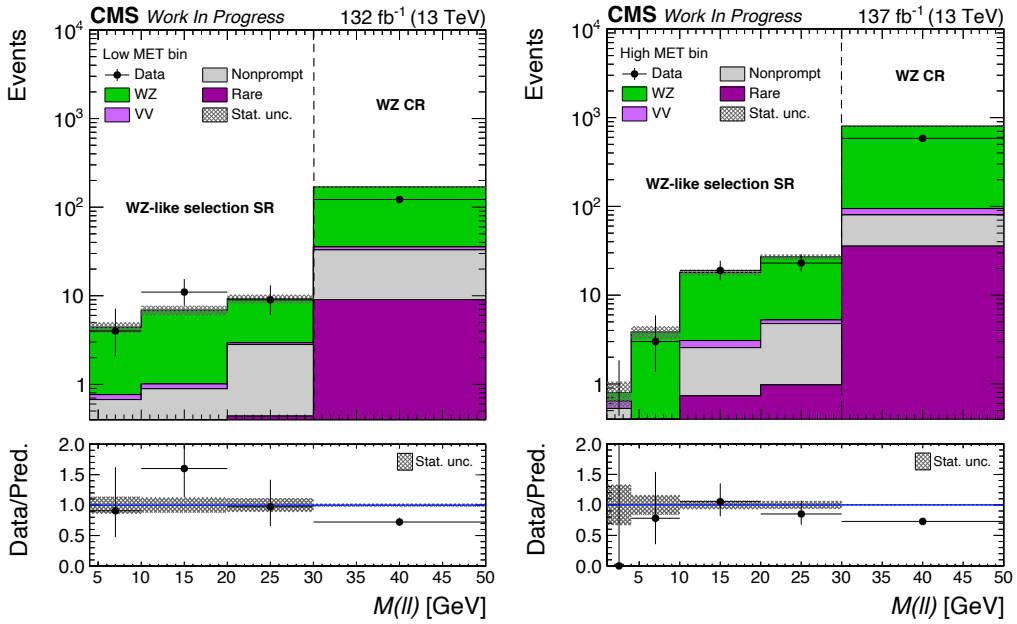


Figure 4.31: The uncorrected MC distribution of the $M(\ell\ell)$ variable is shown for the low-MET (left) and high-MET (right) bins of the fully leptonic WZ ER with the full Run 2 yields combined. The uncertainty band includes only the statistical component.

lation. To validate the proper modeling of the VV processes in the simulation, a specially designed VR, denoted as "VV VR", is used. This region is constructed orthogonal to the SRs by the inverting the leading lepton p_{T} and the $M_{\mathrm{T}}(\ell, p_{\mathrm{T}}^{\mathrm{miss}})$ requirements, i.e. imposing $p_{\mathrm{T}}(\ell_1) > 30\,\mathrm{GeV}$ and $M_{\mathrm{T}}(\ell, p_{\mathrm{T}}^{\mathrm{miss}}) > 90\,\mathrm{GeV}$ respectively. The purely simulated $M(\ell\ell)$ distributions of the VV VR are given in Fig. 4.32 for the different MET bins. As expected, the VV VR is dominated by the $\mathrm{t\bar{t}}$ process, due to the presence of two W bosons in its final state, and the VV processes. A very good agreement is observed between the data and the MC simulation.

4.4.6 Rare Processes

A small percentage of the background of the analysis is due SM processes with low production cross sections, leading to dilepton or trilepton final states. Examples of such processes are the tV, the ttV or the VVV processes, which are grouped together as a single background category, referred to as "Rare". Due

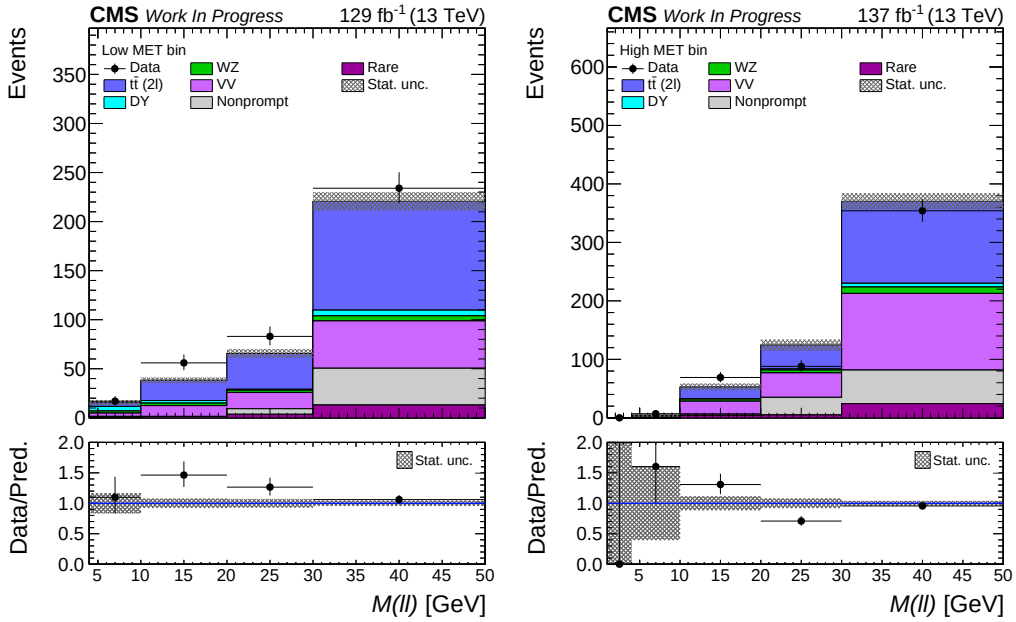


Figure 4.32: The uncorrected MC distribution of the $M(\ell\ell)$ variable is shown for the low-MET (left) and high-MET (right) bins of the VV VR with the full Run 2 yields combined. The uncertainty band includes only the statistical component.

to their minor contribution, the yield for these processes are taken directly from simulation.

4.5 Systematic Uncertainties

Every measurement performed and used in this search is accompanied by its respective uncertainty. These uncertainties are split into two components: The component related to the statistical precision of a measurement is referred to as "statistical uncertainty". The component associated to the experimental nature of the measurements, e.g. to detector effects, or to the precision of the signal and background simulation is denoted as "systematic uncertainty". Even though the statistical uncertainty is usually the most significant component in this analysis, due to the low yield in the majority of the SRs, systematic uncertainties need to be carefully taken into account, as they can also have an important impact on the final result in some cases.

This section is dedicated to explaining how the systematic uncertainties are estimated and to presenting their effect on each process, after they have been refined by the simultaneous maximum likelihood fit (post-fit effect), described in Sec. 4.6.

Possible mismodeling of the online selection in the simulation is taken into account by assigning a dedicated trigger uncertainty on the MC samples. This is done separately for the regions with different number of leptons (2ℓ or 3ℓ) and for each year, since the lepton reconstruction at the HLT as well as the PU profile that drive the trigger performance are different in different years. The uncertainty is also different for the FullSIM and FastSIM MC samples, because the former include the trigger emulation, while the latter do not. The trigger uncertainty arises from the statistical precision of simulated samples and the choice of functions used for the measurement of the trigger efficiencies. More specifically, the systematic uncertainty assigned for each muon in the low-MET bins of the analysis, where the double-μ+MET trigger path is used, is 2%. The low-MET bin of the WZ ER region is excluded from this per-muon trigger uncertainty, since it uses the pure double-μ trigger paths that have a negligible muon uncertainty. The online requirement on the MET is applied in all the MET bins of the analysis. In the range of $p_{\mathrm{T}}^{\mathrm{miss,corr}}$ where the efficiency of each trigger path has reached its plateau, i.e. $150 < p_{\mathrm{T}}^{\mathrm{miss,corr}} < 200\,\mathrm{GeV}$ for the double-μ+MET trigger path and $p_{\mathrm{T}}^{\mathrm{miss,corr}} > 250\,\mathrm{GeV}$ for the pure $p_{\mathrm{T}}^{\mathrm{miss,corr}}$ trigger paths, a 2% trigger uncertainty is assigned. Before reaching the plateau, the trigger efficiency is more susceptible to fluctuations of the background MC samples and the shape of the turn-on, therefore a higher uncertainty, 5%, is set for the ranges $125 < p_{\mathrm{T}}^{\mathrm{miss,corr}} < 150\,\mathrm{GeV}$ and $200 < p_{\mathrm{T}}^{\mathrm{miss,corr}} < 250\,\mathrm{GeV}$. After the simultaneous maximum likelihood fit, the trigger uncertainty has an effect on the final yield of the simulated samples that varies from 2% to 9%. The higher percentage affects the regions of the analysis on the MET trigger turn-on.

A systematic uncertainty, denoted as "leptonic", is used to consider data-to-simulation differences in the tight ID selection requirements for leptons. The leptonic uncertainties are applied to each lepton separately, as a function of lepton flavor (μ or electron), p_{T} and η (barrel or endcap). They are measured simultaneously with the lepton efficiencies presented in Secs. 4.3.1 and 4.3.2 and correspond to the error bars in Figs. 4.9, 4.11, 4.13 and 4.15. As in the case of the trigger uncertainties, the source of the leptonic uncertainties are the statistical uncertainty and the modeling functions of the simulated samples used in the lepton efficiency measurements. The post-fit effect of

the leptonic uncertainties is as low as 2% for high p_T leptons and reaching up 9% for low p_T leptons.

The JECs (Sec. 3.4.7) applied in the analysis come with their own systematic uncertainties that can classified in four general categories [65]: PU offset uncertainties are associated with the mismodeling of the PU in the simulation. The calibration uncertainties vs. p_T and η account for potential data-to-simulation differences of the JECs in different regions of the energy spectrum and different regions of the detector respectively. The jet-flavor response uncertainties are estimated in simulation to consider the dissimilarity between the samples that are used to compute the JECs and those that the JECs are applied on. All the JEC-related uncertainties are grouped together, considered correlated across the years and are applied to the simulated samples of the analysis, separately for FullSIM and FastSIM ones. The post-fit effect of the JEC uncertainty is between 1 and 10%.

The presence of either a b tag requirement or a b tag veto in all regions of the analysis introduces uncertainties connected to the b tagging efficiencies of the DeepCSV algorithm (Sec. 3.4.7). These uncertainties are related to the composition of the light-flavor rich and heavy-flavor rich regions used for the evaluation of the b tagging performance. The contamination of a region from jets of other flavors and the statistical fluctuations of those regions are taken into account. The uncertainties introduced in the b tagging efficiency due to the JEC uncertainties are also considered. These b tagging systematic uncertainties are applied separately for each source and, for FullSIM and FastSIM samples and, are taken to be correlated across the years. Their collective impact on the post-fit yields varies in the range 1–4%.

As already mentioned in Sec. 4.3.3, the simulated PU distribution is corrected with DD weights, calculated by using the measured total inelastic cross section. The uncertainty of this cross section is 4.6% [116] and it is propagated to the PU weights applied to the analysis MC samples. This results in a shape variation of the PU distribution, which is taken as a systematic uncertainty. The post-fit effect of this uncertainty amounts to 1–2%.

The luminosity value for each Run 2 data set is measured centrally by the CMS experiment, along with its associated uncertainty. The luminosity systematic uncertainty for the 2016 [119], 2017 [120] and 2018 [121] data sets are found to be 2.5%, 2.3% and 2.5% respectively and remain at these values also after the simultaneous maximum likelihood fit.

Each dominant prompt background process ($t\bar{t}$, DY, fully leptonic WZ) is associated to a CR, which drives the calculation of the DD SFs, used to

correct the normalization of each one of these processes, as mentioned in Secs. 4.4.2, 4.4.3 and 4.4.4. Each SF is calculated separately for different years and different CR MET bins, along with an uncertainty which is determined by the simultaneous maximum likelihood fit and is related to the data-to-simulation shape agreement and the statistical power of each CR. The post-fit values of these uncertainties are $\sim 15\%$ for the $\mathrm{t\bar{t}}$ process and range between 15–35% for the DY process and between 12–27% for the fully leptonic WZ process.

Each subleading prompt processes, i.e. the VV and the Rare processes, which are estimated from simulation, are each assigned a single systematic uncertainty correlated across all years and MET bins. Their systematic uncertainty is set to 50% and remains at the same level post-fit, because there is no high-purity and high-yield CR that can be constructed to further constrain the uncertainty of these processes.

There are multiple systematic uncertainties related to nonprompt background estimation. An overall normalization uncertainty, correlated across the years, is applied to the nonprompt background. Its pre-fit value is estimated by the maximum data-to-simulation disagreement seen in the closure tests for the separate years (Fig. 4.27) and is set to 40%. The inclusion of the nonprompt background-rich SS CR in the simultaneous maximum likelihood fit constrains this uncertainty post-fit down to $\sim 5\%$.

As a result of the usage of the MC information in the semi-DD tight-to-loose method, additional systematic uncertainties are considered. To account for potential mismodeling in the shape of the MC distribution used in the method, a shape uncertainty is applied to the $p_{\mathrm{T}}(\ell_1)$ distribution for the 2ℓ-Stop, separately for each year. In the case of the $p_{\mathrm{T}}(\ell_1)$ distribution, the pre-fit uncertainty is modeled as:

$$\mathrm{Unc}\left(p_{\mathrm{T}}(\ell_1)\right) = \begin{cases} 1.4^{\frac{p_{\mathrm{T}}(\ell_1)-15}{15}} & p_{\mathrm{T}}(\ell_1) \leq 15\,\mathrm{GeV} \\ \max\left(1.4^{\frac{p_{\mathrm{T}}(\ell_1)-15}{15}}, 1.2\right) & p_{\mathrm{T}}(\ell_1) > 15\,\mathrm{GeV} \end{cases} \tag{4.18}$$

for 2016 and 2017, while no such uncertainty is applied for 2018, since the shape disagreements between data and simulation were found to be negligible, i.e. within the statistical uncertainty. The functional form and the parameters were chosen such that this uncertainty is interpreted as a linear variation of the $p_{\mathrm{T}}(\ell_1)$ distribution with a maximum value of 40% by the simultaneous maximum likelihood fit. The max function is used to ensure

a continuous dependency on the $p_{\mathrm{T}}(\ell_1)$ while also constraining the maximum uncertainty value in the higher $p_{\mathrm{T}}(\ell_1)$, where the shape discrepancies are milder. The post-fit effect of this systematic uncertainty ends up being $\sim$8%.

The $M(\ell\ell)$ or $M^{\mathrm{min}}_{\mathrm{SFOS}}(\ell\ell)$ shape agreement is better for the 2ℓ- and 3ℓ-EWK SRs, generally dominated by statistical uncertainties, especially in the latter regions. However, a systematic uncertainty is used in this case to consider the $M(\ell\ell)$ shape differences between the 2ℓ-EWK MET bins when these are merged in the AR sidebands. This uncertainty is estimated as follows: The $M(\ell\ell)$ distribution in the medium-MET bin of the AR, which has the most yields, is divided by the average distribution of the high- and ultra-MET bins of the AR. The ratio is shown in Fig. 4.33 and is fitted with a linear function of the form:

$$f(M(\ell\ell)) = p_0 + p_1(\overline{M(\ell\ell)} - M(\ell\ell)) \tag{4.19}$$

where p_0 and p_1 are the fit parameters and $\overline{M(\ell\ell)}$ is the weighted average of the distribution, equal to 18.1 GeV. This linear function parametrization was specifically chosen to remove correlations between the fit parameters.

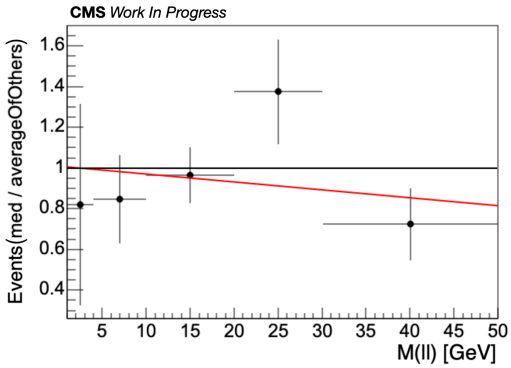


Figure 4.33: The ratio of the AR $M(\ell\ell)$ distribution of the medium-MET bin over the average of the high and ultra MET bins and a linear fit to the points, modeled as described in the main text.

The fit results in a constant term $c_0 = p_0 + p_1\overline{M(\ell\ell)} = 1.01 \pm 0.15$ and a slope $c_1 = p_1 = -0.004 \pm 0.007$. The compatibility of the constant term with unity and of the slope with zero imply the compatibility of the $M(\ell\ell)$ shape across the different MET bins. The uncertainty on the slope is used to estimate the maximum value of the uncertainty, which is found at the

edge of the distribution and amounts to $\sim 20\%$. Similarly to Eq. 4.18, the uncertainty:

$$\mathrm{Unc}\,(M(\ell\ell)) = 1.2^{\frac{M(\ell\ell)-25}{25}} \tag{4.20}$$

is applied to the nonprompt background high- and ultra-MET bins of the EWK SRs, separately for each year. The value of this systematic uncertainty is reduced to $\sim 5\%$ post-fit.

Additional uncertainties are applied on the signal samples: The uncertainty with which the parton density functions have been calculated affect the signal cross section and, as a result, the post-fit signal yields by 3.5–8.5%. The effect of the variation of the renormalization and factorization scales on the signal acceptance is also considered. These scales are varied up ($\times 2$) and down ($\times 0.5$) their original value and the change in the shape of the signal distributions leads to a post-fit uncertainty of $\sim 1\%$.

Furthermore, potential differences of the ISR modeling between data and simulation can have a nonnegligible impact on the signal distributions. Due to this, a dedicated set of p_{T}-dependent correction factors is derived for 2016 to harmonize the data with the simulation shapes. The data-to-simulation correction factors have been extracted from a region enriched with DY and $\mathrm{t\bar{t}}$ events, captured with single muon and electron trigger paths. At least two isolated leptons are required with $p_{\mathrm{T}}(\ell_1) > 40\,\mathrm{GeV}$ and $80 < M(\ell\ell) < 100\,\mathrm{GeV}$. At least one ISR jet tagged with the loose jet ID, as defined in [115], is selected with $p_{\mathrm{T}} > 30\,\mathrm{GeV}$. A jet cleaning procedure is implemented as well. These correction factors, with values up to 20%, are used to vary the signal distributions. This variation results in a post-fit uncertainty of 1–5%. For 2017 and 2018, the data to simulation differences were found to be within the statistical uncertainties. Therefore, a flat 1% uncertainty is set for those years, which remains at the same level after of the simultaneous maximum likelihood fit.

Normally, the FastSIM simulation agrees with the FullSIM simulation up to $\sim 1\%$ in the phase space probed by this analysis. To account for potential mismodeling, even after the correction of Eq. 4.5, the quantity

$$\mathrm{Unc} = \frac{1}{2}\left(p_{\mathrm{T}}^{\mathrm{miss}}(\mathrm{Reconstructed}) - p_{\mathrm{T}}^{\mathrm{miss}}(\mathrm{Generator})\right) \tag{4.21}$$

is used as a shape systematic uncertainty on FastSIM samples, applied for each year separately. The post-fit effect on this uncertainty on the yields of FastSIM samples varies between 1% and 10%.

Finally, the weights that are used to correct the L1T prefire issue, explained in Sec. 4.2.2, come with their own uncertainties. These uncertainties are propagated to the final yields of the analysis and have a post-fit effect of $\sim$1% in 2016 and 2017, where they are applied.

4.6 Results and Interpretations

For the extraction of the final results, a simultaneous maximum likelihood fit of the signal and background simulation yields to the data is performed. More precisely, the likelihood function L includes a single, positive, unconstrained parameter of interest, denoted as r, which scales the signal yields linearly, and incorporates the systematic uncertainties as nuisance parameters, the full set of which is denoted as $\vec{\theta} = (\theta_1, \theta_2, \ldots, \theta_i, \ldots)$. It is parametrized as follows:

$$L(r, \vec{\theta} \mid n_{\text{data}}) = \prod_{\text{bins}} \left[\text{Poisson} \left(n_{\text{data}}, r \times n_{\text{signal}}(\vec{\theta}) + \sum_{\text{bkg}} n_{\text{bkg}}(\vec{\theta}) \right) \times \prod_i P_{\theta_i} \left(\theta_i, \sigma_{\theta_i} \right) \right] \tag{4.22}$$

The likelihood function expresses the probability of finding a rate $r \times n_{\text{signal}} + \sum_{\text{bkg}} n_{\text{bkg}}$, where n_{signal} and n_{bkg} are the MC expectations for the signal and background yields of a specific process, respectively, given the data. The sum over all the background yields is performed to compute the total background yield, while the outer product takes into account each one of the analysis bins. These are comprised by the bins of the $M(\ell\ell)$ distribution of the 2ℓ-EWK SRs and 2ℓ CRs, the bins of the $M^{\text{min}}_{\text{SFOS}}(\ell\ell)$ distribution of the 3ℓ-EWK SRs and 3ℓ CRs and the bins of the $p_{\text{T}}(\ell_1)$ distribution of the 2ℓ-Stop SRs. The specific bins of these distribution are described in Sec. 4.3.4 and are illustrated explicitly in Fig. 4.18. The 2ℓ and 3ℓ CR binning exactly follows the 2ℓ- and 3ℓ-EWK SR binning. All the bins are included separately for each year.

The last term of the likelihood function is the product of the distributions of all the nuisance parameters, each one of which is related to a systematic uncertainty of the analysis. The distribution used for each systematic uncertainty depends on the nature of the uncertainty: The prompt background

normalization uncertainties are modeled with log-uniform distributions, so that each normalization can freely float, with no initial input to their pre-fit values. The rest of the uncertainties are modeled with log-normal distributions with its mean and standard deviation being set to the pre-fit estimations of the uncertainty, the full set of which is denoted as $\vec{\sigma}_\theta$ in Eq. 4.22 and is described in detail in Sec. 4.5.

The likelihood function is used to determine the best value of r given the data, with the construction of the test statistic

$$\tilde{q}_r = -2 \ln \left(\frac{L(r, \hat{\vec{\theta}}_r \mid n_{\text{data}})}{L(\hat{r}, \hat{\vec{\theta}} \mid n_{\text{data}})} \right) \tag{4.23}$$

The hat over some parameters signifies that these have been set to the values that maximize the likelihood function 4.22. The maximization is performed over all parameters, except those in the subscript. For example, $\hat{\theta}_r$ refers to the values of $\vec{\theta}$ that maximize L as a function of r, while $\hat{\theta}$ refers to the values of $\vec{\theta}$ that maximize L globally.

The test statistic $\tilde{q}_r$, combined with the observed values from data, is used to define the probability distribution functions of the *signal+background* hypothesis, $f\left(\tilde{q}_r \mid r, \hat{\vec{\theta}}_r^{\text{obs}}\right)$, and the *background-only* hypothesis, $f\left(\tilde{q}_r \mid 0, \hat{\vec{\theta}}_0^{\text{obs}}\right)$. These hypotheses establish the respective p-values, p_r and p_0, that eventually determine the upper limit on r at 95% confidence level, exploiting asymptotic formulae to extract the final result, which utilizes the CL_s criterion. An extensive discussion of the methods outlined above is presented in Refs. [122–125].

The background-only, post-fit distributions of the CR variables used as input in the simultaneous maximum likelihood fit are shown in Figs. 4.34–4.37. In contrast with the respective distributions in Secs. 4.4.1–4.4.4, these include the full set of DD corrections for all processes. The uncertainty bands represents the quadratic sum of statistical and systematic uncertainties, as determined by the fit to the data. An excellent agreement between simulation and data is observed in all of them.

Similarly, the same distributions, as determined by the background-only simultaneous maximum likelihood fit, are presented in Figs. 4.38–4.40 for the SRs. In the 2ℓ- and 3ℓ-EWK SRs (Fig. 4.38 and 4.39 respectively), the pre-fit distributions from the TChiWZ and the simplified higgsino models are overlaid in the scenario where the $\tilde{m}_{\tilde{\chi}_2^0} \tilde{m}_{\tilde{\chi}_1^0}$ is positive and negative respec-

tively. The numbers after the model name in the legend indicate the mass of the NLSP and the mass splitting between the NLSP and LSP, in GeV. In the 2ℓ-Stop SRs, the pre-fit distributions from the T2Bff$\widetilde{\chi}_1^0$ and T2BW models are overlaid. The significance of the numbers after the model name in the legend is the same as above, with the mass of the top squark replacing the mass of the NLSP. Generally, a good agreement between the background simulated processes and the data is observed.

The background-only, post-fit yields of all the SRs of the analysis are quoted in detail in Tables 4.10–4.13.

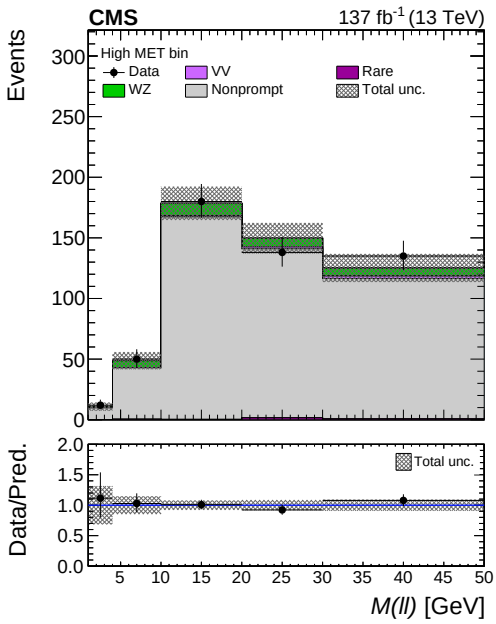


Figure 4.34: The post-fit distribution of the $M(\ell\ell)$ variable is shown for the SS CR with the full Run 2 yields combined. The uncertainty band includes both the statistical and systematic components.

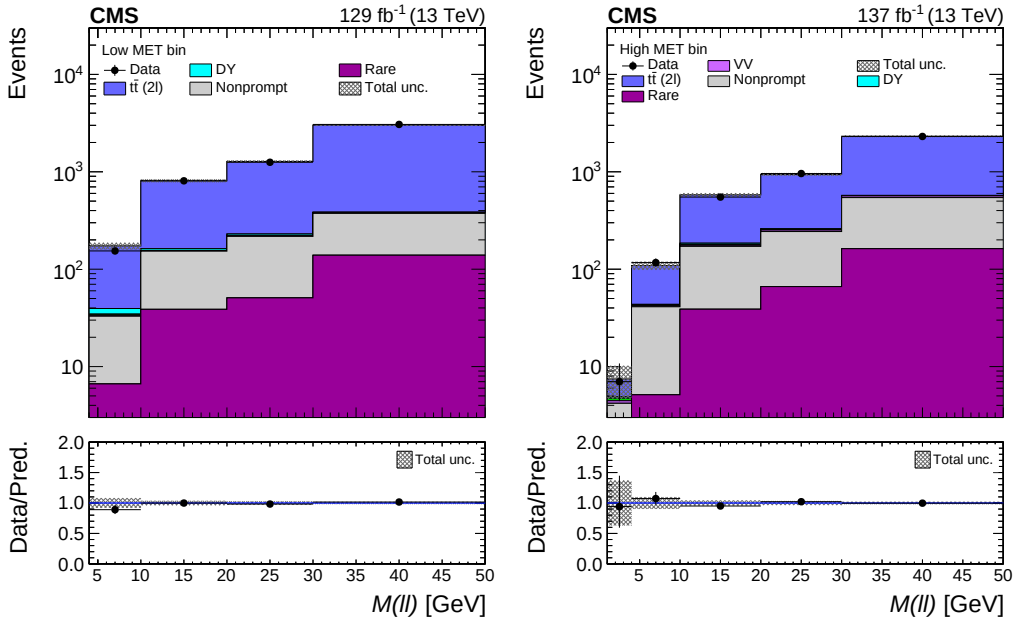


Figure 4.35: The post-fit distribution of the $M(\ell\ell)$ variable is shown for the low-MET (left) and high-MET (right) bins of the $t\bar{t}$ CR with the full Run 2 yields combined. The uncertainty band includes both the statistical and systematic components.

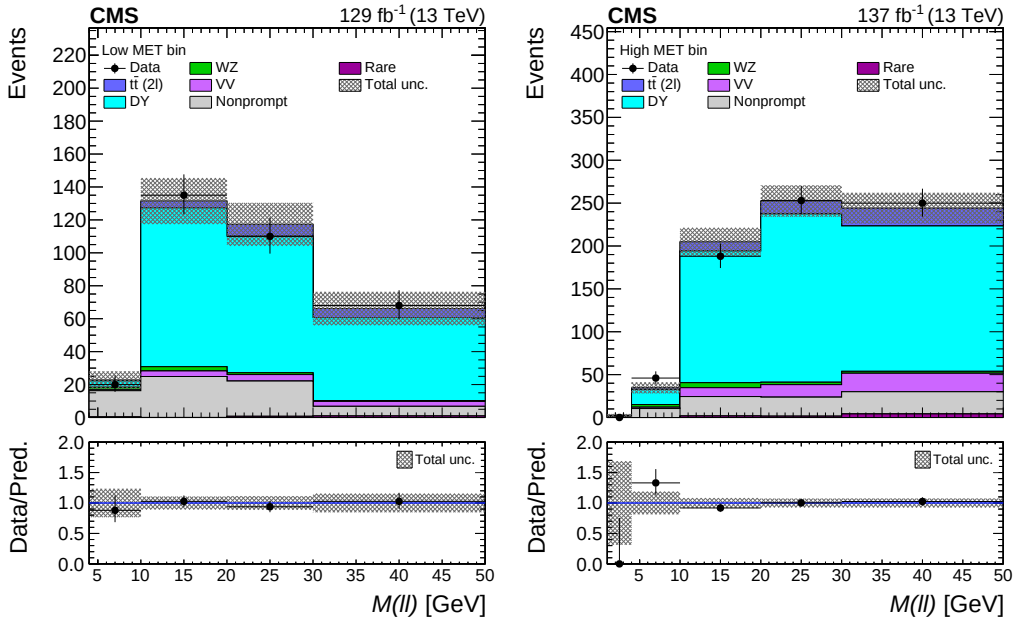


Figure 4.36: The post-fit distribution of the $M(\ell\ell)$ variable is shown for the low-MET (left) and high-MET (right) bins of the DY CR with the full Run 2 yields combined. The uncertainty band includes both the statistical and systematic components.

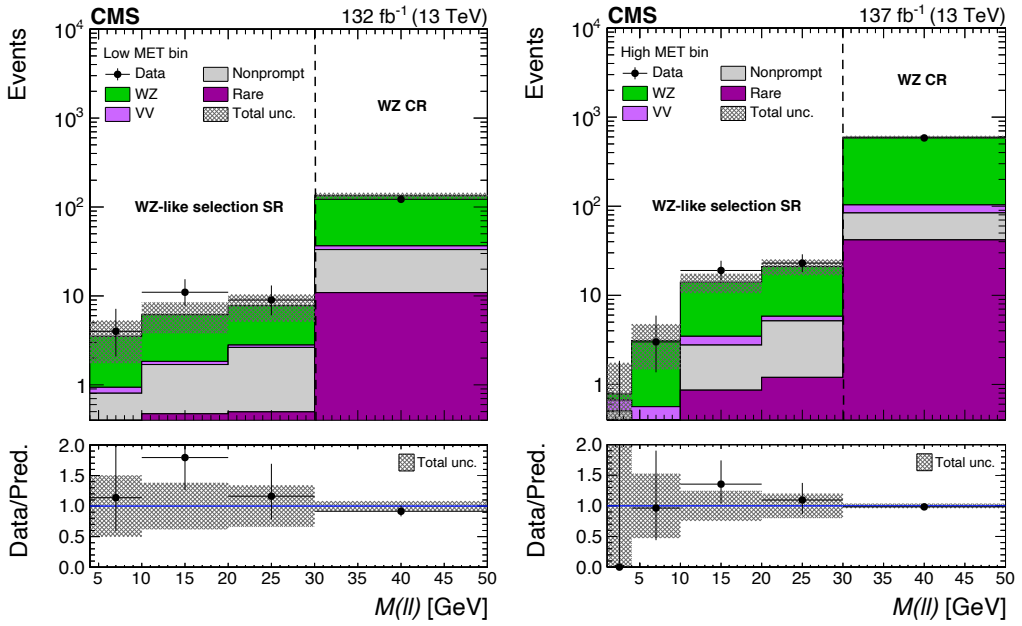


Figure 4.37: The post-fit distribution of the $M(\ell\ell)$ variable is shown for the low-MET (left) and high-MET (right) bins of the fully leptonic WZ ER with the full Run 2 yields combined. The uncertainty band includes both the statistical and systematic components.

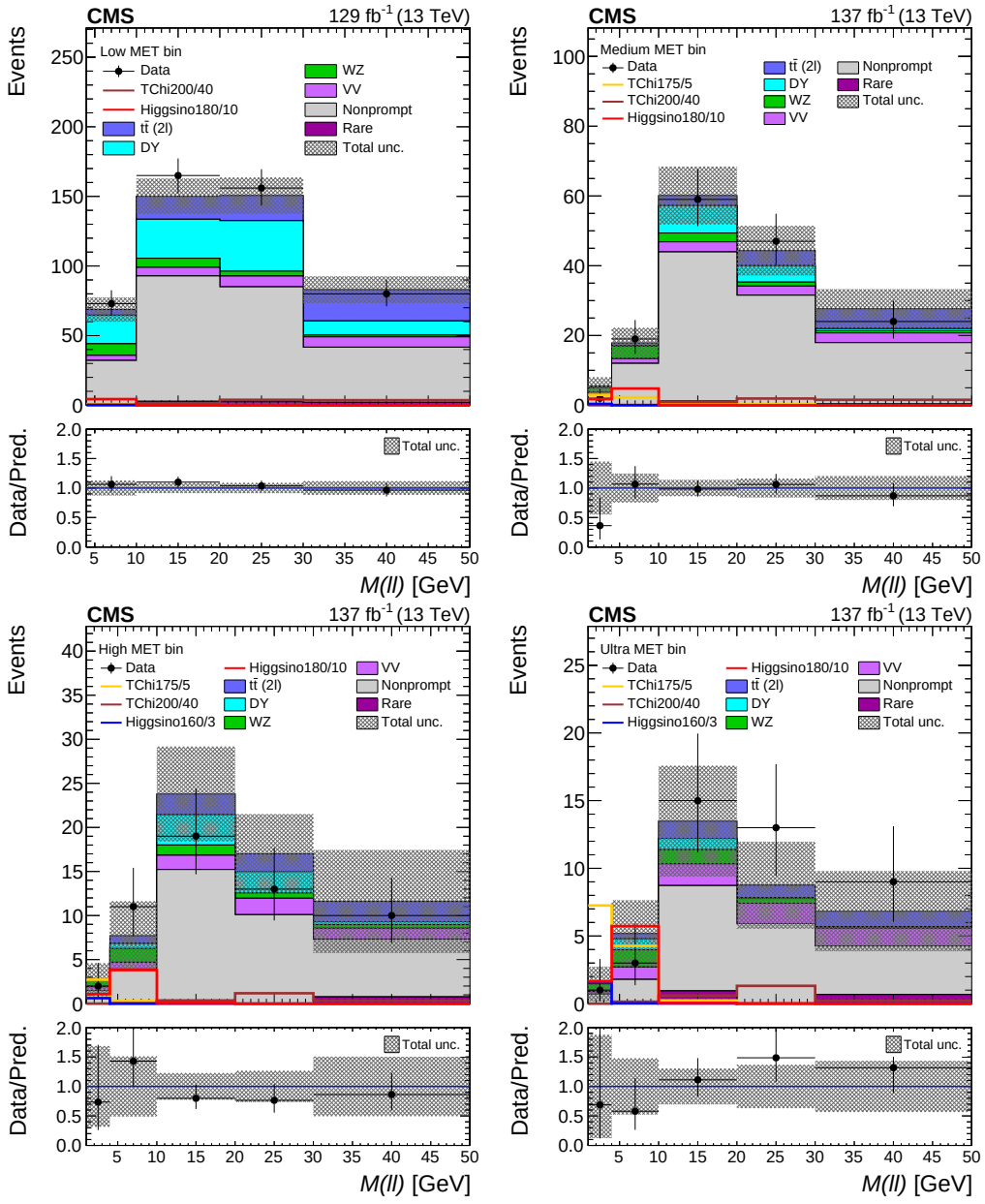


Figure 4.38: The post-fit distribution of the $M(\ell\ell)$ variable is shown for the low-MET (upper left), medium-MET (upper right), high-MET (lower left) and ultra-MET (lower right) bins of 2ℓ-EWK SR with the full Run 2 yields combined. The uncertainty band includes both the statistical and systematic components. More details are given in the main text.

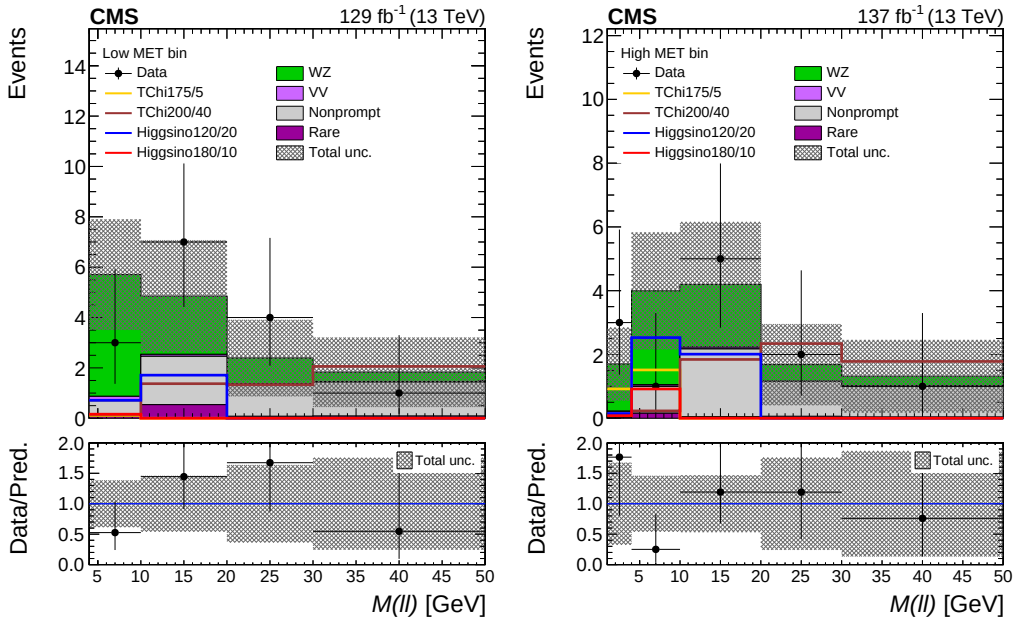


Figure 4.39: The post-fit distribution of the $M_{\mathrm{SFOS}}^{\mathrm{min}}(\ell\ell)$ variable is shown for the low-MET (left) and high-MET (right) bins of the 3ℓ-EWK SR with the full Run 2 yields combined. The uncertainty band includes both the statistical and systematic components. More details are given in the main text.

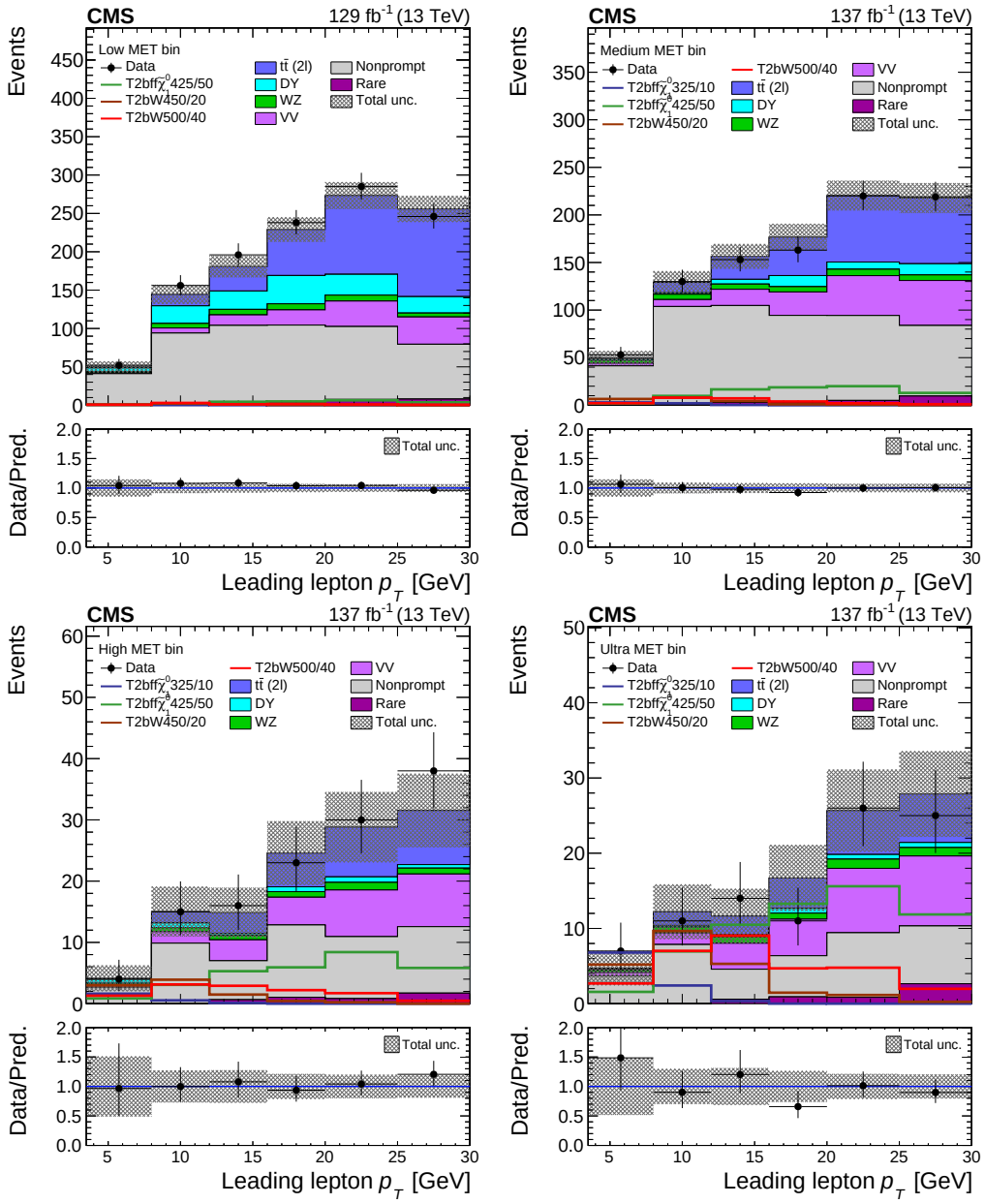


Figure 4.40: The post-fit distribution of the leading lepton p_{T} variable is shown for the low-MET (upper left), medium-MET (upper right), high-MET (lower left) and ultra-MET (lower right) bins of the 2ℓ-Stop SR with the full Run 2 yields combined. The uncertainty band includes both the statistical and systematic components. More details are given in the main text.

MET Bin	$M(\ell\ell)$ [GeV]	$t\bar{t}$	DY	VV	WZ	Rare	Nonprompt	Total bkg	Data
Low	4–10	4.0 ± 2.0	20.6 ± 5.2	3.7 ± 2.4	8.3 ± 2.6	$0.28^{+0.72}_{-0.27}$	31.9 ± 5.6	68.7 ± 8.7	73
	10–20	16.5 ± 4.2	28.0 ± 6.2	6.2 ± 3.2	6.5 ± 2.3	2.8 ± 2.1	90.1 ± 9.3	151 ± 13	165
	20–30	18.0 ± 4.4	36.3 ± 7.1	7.8 ± 3.5	3.5 ± 1.7	2.9 ± 2.1	82.1 ± 8.9	151 ± 13	156
	30–50	22.4 ± 4.9	10.2 ± 3.7	7.4 ± 3.5	1.3 ± 1.0	2.1 ± 1.8	39.6 ± 6.2	82.9 ± 9.6	80
Medium	1–4	$0.11^{+0.33}_{-0.10}$	$0.37^{+0.72}_{-0.36}$	$0.7^{+1.1}_{-0.7}$	1.3 ± 1.0	$0.04^{+0.23}_{-0.03}$	3.0 ± 2.0	5.5 ± 2.5	2
	4–10	$0.75^{+0.90}_{-0.74}$	$0.15^{+0.50}_{-0.14}$	$1.4^{+1.5}_{-1.4}$	3.5 ± 1.7	$0.14^{+0.39}_{-0.13}$	11.9 ± 3.6	17.8 ± 4.4	19
	10–20	2.9 ± 1.7	7.9 ± 3.4	2.9 ± 2.2	2.5 ± 1.4	1.2 ± 1.2	42.8 ± 6.8	60.1 ± 8.3	59
	20–30	4.3 ± 2.1	4.7 ± 2.6	2.6 ± 2.0	1.1 ± 1.0	$0.27^{+0.54}_{-0.26}$	31.3 ± 5.8	44.3 ± 7.1	47
	30–50	5.7 ± 2.4	$0.6^{+1.0}_{-0.6}$	2.8 ± 2.1	$0.63^{+0.70}_{-0.62}$	$0.35^{+0.65}_{-0.34}$	17.6 ± 4.4	27.7 ± 5.6	24
High	1–4	< 0.02	< 0.1	$0.43^{+0.88}_{-0.42}$	0.8 ± 0.8	< 0.07	1.5 ± 1.3	2.7 ± 1.9	2
	4–10	$0.9^{+1.2}_{-0.9}$	$0.57^{+0.97}_{-0.56}$	$0.8^{+1.1}_{-0.8}$	1.5 ± 1.1	$0.3^{+2.6}_{-0.3}$	3.7 ± 2.0	7.7 ± 3.9	11
	10–20	2.4 ± 1.6	3.4 ± 2.3	1.6 ± 1.6	1.2 ± 0.9	$0.3^{+1.3}_{-0.3}$	14.9 ± 4.0	23.8 ± 5.4	19
	20–30	2.0 ± 1.5	2.4 ± 1.9	1.9 ± 1.7	$0.61^{+0.67}_{-0.60}$	$0.03^{+0.45}_{-0.02}$	10.1 ± 3.3	17.0 ± 4.5	13
	30–50	2.3 ± 1.7	$0.32^{+0.73}_{-0.31}$	$1.2^{+1.4}_{-1.1}$	$0.40^{+0.53}_{-0.39}$	$0.8^{+4.6}_{-0.7}$	6.6 ± 2.7	11.6 ± 5.8	10
Ultra	1–4	< 0.02	< 0.1	$0.18^{+0.65}_{-0.17}$	$0.57^{+0.65}_{-0.56}$	< 0.01	$0.70^{+0.88}_{-0.69}$	1.5 ± 1.3	1
	4–10	$0.38^{+0.64}_{-0.37}$	$0.8^{+1.1}_{-0.8}$	$0.9^{+1.2}_{-0.9}$	1.3 ± 1.0	$0.12^{+0.44}_{-0.11}$	1.7 ± 1.3	5.2 ± 2.5	3
	10–20	1.3 ± 1.2	$0.8^{+1.2}_{-0.8}$	1.6 ± 1.6	1.05 ± 0.89	$0.9^{+1.4}_{-0.9}$	7.8 ± 2.9	13.5 ± 4.1	15
	20–30	$0.9^{+1.0}_{-0.9}$	$0.06^{+0.28}_{-0.05}$	$1.5^{+1.6}_{-1.5}$	$0.3^{+0.50}_{-0.34}$	< 0.09	5.9 ± 2.5	8.8 ± 3.2	13
	30–50	1.2 ± 1.1	< 0.1	$1.3^{+1.5}_{-1.3}$	$0.09^{+0.24}_{-0.08}$	$0.7^{+1.2}_{-0.7}$	3.6 ± 2.0	6.8 ± 3.0	9

Table 4.10: Observed and predicted yields as extracted from the background-only maximum likelihood fit, in the 2ℓ-EWK SRs. Uncertainties include both the statistical and systematic components.

MET Bin	$M^{\text{min}}_{\text{SFOS}}(\ell\ell)$ [GeV]	VV	WZ	Rare	Nonprompt	Total bkg	Data
Low	4–10	$0.18^{+0.54}_{-0.17}$	4.8 ± 1.9	$0.08^{+0.38}_{-0.07}$	$0.61^{+0.83}_{-0.60}$	5.7 ± 2.2	3
	10–20	$0.08^{+0.35}_{-0.07}$	2.3 ± 1.3	$0.5^{+1.0}_{-0.5}$	1.9 ± 1.4	4.9 ± 2.2	7
	20–30	$0.03^{+0.23}_{-0.02}$	1.0 ± 1.0	$0.07^{+0.35}_{-0.06}$	1.3 ± 1.2	2.4 ± 1.5	4
	30–50	$0.01^{+0.13}_{-0.01}$	$0.39^{+0.55}_{-0.38}$	$0.08^{+0.37}_{-0.07}$	1.4 ± 1.2	1.8 ± 1.4	1
High	1–4	$0.01^{+0.18}_{-0.01}$	1.5 ± 1.0	$0.03^{+0.20}_{-0.02}$	$0.18^{+0.44}_{-0.17}$	1.7 ± 1.2	3
	4–10	$0.05^{+0.34}_{-0.04}$	2.9 ± 1.4	$0.16^{+0.47}_{-0.15}$	$0.85^{+0.99}_{-0.84}$	4.0 ± 1.8	1
	10–20	$0.06^{+0.32}_{-0.05}$	2.0 ± 1.2	$0.05^{+0.26}_{-0.04}$	2.1 ± 1.5	4.2 ± 2.0	5
	20–30	< 0.002	$0.52^{+0.60}_{-0.51}$	$0.06^{+0.29}_{-0.05}$	1.1 ± 1.1	1.7 ± 1.3	2
	30–50	< 0.002	$0.31^{+0.46}_{-0.30}$	$0.03^{+0.23}_{-0.02}$	1.0 ± 1.0	1.3 ± 1.1	1

Table 4.11: Observed and predicted yields as extracted from the background-only maximum likelihood fit, in the 3ℓ-EWK SRs. Uncertainties include both the statistical and systematic components.

MET Bin	$M_{\text{SFOS}}^{\text{min}}(\ell\ell)$ [GeV]	VV	WZ	Rare	Nonprompt	Total bkg	Data
Low	4–10	$0.13^{+0.47}_{-0.12}$	2.6 ± 1.4	$0.31^{+0.67}_{-0.30}$	$0.49^{+0.70}_{-0.48}$	3.5 ± 1.8	4
	10–20	$0.14^{+0.47}_{-0.13}$	4.3 ± 1.8	$0.47^{+0.83}_{-0.46}$	1.2 ± 1.1	6.1 ± 2.3	11
	20–30	$0.17^{+0.51}_{-0.16}$	5.0 ± 2.0	$0.50^{+0.85}_{-0.49}$	2.1 ± 1.5	7.8 ± 2.6	9
High	1–4	$0.16^{+0.56}_{-0.15}$	$0.11^{+0.29}_{-0.10}$	$0.06^{+0.33}_{-0.05}$	$0.44^{+0.66}_{-0.43}$	$0.78^{+0.97}_{-0.77}$	0
	4–10	$0.22^{+0.60}_{-0.21}$	2.6 ± 1.4	$0.10^{+0.38}_{-0.09}$	$0.24^{+0.59}_{-0.23}$	3.1 ± 1.6	3
	10–20	$0.7^{+1.1}_{-0.7}$	10.6 ± 2.8	$0.9^{+1.1}_{-0.9}$	1.9 ± 1.4	14.0 ± 3.4	19
	20–30	$0.7^{+1.0}_{-0.7}$	15.2 ± 3.3	$1.2^{+1.3}_{-1.2}$	4.0 ± 2.0	21.0 ± 4.2	23

Table 4.12: Observed and predicted yields as extracted from the background-only maximum likelihood fit, in the WZ-like selection SRs. Uncertainties include both the statistical and systematic components.

MET Bin	$p_{\text{T}}(\ell_1)$ [GeV]	$t\bar{t}$	DY	VV	WZ	Rare	Nonprompt	Total bkg	Data
Low	3.5–8	1.2 ± 1.2	5.2 ± 3.1	$1.0^{+1.2}_{-1.0}$	1.4 ± 1.1	$0.06^{+0.27}_{-0.05}$	41.0 ± 6.3	49.9 ± 7.2	52
	8–12	15.0 ± 4.0	22.9 ± 5.9	6.6 ± 3.1	6.0 ± 2.1	$0.96^{+0.99}_{-0.95}$	93.1 ± 9.4	144 ± 12	156
	12–16	31.8 ± 5.9	24.0 ± 6.1	13.7 ± 4.5	7.2 ± 2.4	2.8 ± 1.7	101.3 ± 9.9	180 ± 14	196
	16–20	59.9 ± 8.0	36.9 ± 7.5	19.8 ± 5.5	7.9 ± 2.5	4.2 ± 2.1	100.2 ± 9.8	229 ± 16	238
	20–25	103 ± 11	27.2 ± 6.5	33.2 ± 7.1	7.7 ± 2.5	7.5 ± 2.8	95.0 ± 9.5	273 ± 18	285
	25–30	114 ± 11	21.4 ± 5.7	35.5 ± 7.3	5.1 ± 2.0	8.0 ± 2.8	71.5 ± 8.3	256 ± 17	246
Medium	3.5–8	1.1 ± 1.0	1.7 ± 1.5	2.8 ± 2.1	2.9 ± 1.4	$0.04^{+0.20}_{-0.03}$	41.3 ± 6.6	49.9 ± 7.3	53
	8–12	11.0 ± 3.3	1.6 ± 1.5	7.3 ± 3.3	5.6 ± 2.0	$0.43^{+0.65}_{-0.42}$	103 ± 10	129 ± 12	130
	12–16	24.1 ± 4.9	5.0 ± 2.6	17.1 ± 5.0	5.5 ± 2.0	2.9 ± 1.7	102 ± 10	156 ± 13	153
	16–20	40.3 ± 6.3	11.7 ± 4.2	24.7 ± 6.1	5.6 ± 2.0	2.4 ± 1.6	92.0 ± 9.8	177 ± 14	163
	20–25	69.9 ± 8.3	7.6 ± 3.4	41.9 ± 7.9	6.7 ± 2.2	5.0 ± 2.2	89.3 ± 9.7	220 ± 16	220
	25–30	69.0 ± 8.3	11.8 ± 4.1	47.3 ± 8.4	5.9 ± 2.0	9.6 ± 3.1	74.2 ± 8.9	218 ± 16	219
High	3.5–8	$0.15^{+0.35}_{-0.14}$	$0.67^{+0.90}_{-0.66}$	$0.34^{+0.72}_{-0.33}$	$0.29^{+0.44}_{-0.28}$	< 0.05	2.7 ± 1.7	4.1 ± 2.1	4
	8–12	1.9 ± 1.4	$0.8^{+1.1}_{-0.8}$	1.9 ± 1.7	$0.64^{+0.67}_{-0.63}$	$0.01^{+0.11}_{-0.01}$	9.9 ± 3.2	15.0 ± 4.1	15
	12–16	3.4 ± 1.8	$0.33^{+0.61}_{-0.32}$	3.4 ± 2.3	0.69 ± 0.69	$0.64^{+0.85}_{-0.63}$	6.4 ± 2.6	14.8 ± 4.1	16
	16–20	5.5 ± 2.3	$0.8^{+1.1}_{-0.8}$	4.5 ± 2.6	0.91 ± 0.80	1.0 ± 1.0	11.8 ± 3.5	24.6 ± 5.2	23
	20–25	8.1 ± 2.8	$0.9^{+1.2}_{-0.9}$	7.6 ± 3.4	1.24 ± 0.93	$0.82^{+0.89}_{-0.81}$	10.1 ± 3.2	28.8 ± 5.8	30
	25–30	8.8 ± 2.9	$0.58^{+0.97}_{-0.57}$	8.6 ± 3.6	0.96 ± 0.81	1.7 ± 1.3	10.8 ± 3.4	31.5 ± 6.0	38
Ultra	3.5–8	$0.12^{+0.37}_{-0.11}$	$0.14^{+0.51}_{-0.13}$	$0.48^{+0.86}_{-0.47}$	$0.29^{+0.46}_{-0.28}$	< 0.03	3.7 ± 2.0	4.7 ± 2.3	7
	8–12	1.8 ± 1.3	$0.22^{+0.59}_{-0.21}$	1.5 ± 1.5	0.78 ± 0.75	$0.02^{+0.12}_{-0.01}$	7.8 ± 2.9	12.2 ± 3.6	11
	12–16	2.4 ± 1.5	$0.31^{+0.63}_{-0.30}$	3.5 ± 2.3	0.87 ± 0.78	$0.60^{+0.79}_{-0.59}$	4.0 ± 2.0	11.6 ± 3.6	14
	16–20	4.0 ± 2.0	$0.64^{+0.89}_{-0.63}$	4.9 ± 2.7	0.80 ± 0.75	$0.90^{+0.93}_{-0.89}$	5.5 ± 2.5	16.7 ± 4.4	11
	20–25	5.8 ± 2.3	$0.62^{+0.95}_{-0.61}$	8.6 ± 3.6	1.22 ± 0.93	$0.84^{+0.92}_{-0.83}$	8.6 ± 3.0	25.7 ± 5.5	26
	25–30	6.5 ± 2.5	$0.7^{+1.0}_{-0.7}$	9.3 ± 3.7	1.12 ± 0.88	2.6 ± 1.6	7.7 ± 2.9	27.9 ± 5.7	25

Table 4.13: Observed and predicted yields as extracted from the background-only maximum likelihood fit, in the 2ℓ-Stop SRs. Uncertainties include both the statistical and systematic components.

The search results are interpreted in terms of all the signal models that were elaborated on in Sec. 4.2. Due to small differences in the final state between EWKino and stop pair production models, different SR regions are included in the simultaneous maximum likelihood fit in each case. For the EWK signal models (TCHIWZ, higgsino SMS and pMSSM), the 2ℓ-EWK SRs and the 3ℓ-EWK SRs are included, totalling 28 bins in each year. Only the 2ℓ-Stop SRs are utilized in stop pair production signal models (T2BFF$\widetilde{\chi}_1^0$, T2BW), contributing 24 bins for each year. In both cases, the full suite of CRs (SS, $\bar{\text{t}}\text{t}$, DY) and WZ ER are exploited, corresponding to 32 bins in each year, in total.

The value of r resulting from the maximum likelihood fit is determined separately for each masspoint of each signal model. Each value is subsequently multiplied by the corresponding cross section to extract the 95% CL upper limits on the SUSY production cross section of the specific masspoint of the the specific signal model. The results are plotted as points in the masspoint space of each model and intermediate values between them are interpolated using the Delauney triangulation method [126]. The uncertainty lines drawn correspond to the 68% CL cross section upper limits for the expected limit and to the uncertainty of the signal cross section, due to the variation of the parton density functions (mentioned in Sec. 4.5) for the observed limit. To iron out potential crude interpolations between masspoints, an extra smoothing procedure is applied, where the value of each masspoint is corrected by the values of the neighbouring masspoints, using the kernel:

$$\text{Smoothing kernel} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 0 \end{pmatrix} \tag{4.24}$$

The procedure described above results in the limit plots for each signal model, shown in Fig. 4.41 for the TCHIWZ model. In this case, the $M(\ell\ell)$ spectrum and the W/Z boson branching fraction corrections have been applied, as described at the end of Sec. 4.1. Due to the fact that both $M(\ell\ell)$ reweighting scenarios are viable for the TCHIWZ model, two versions of the limit plot have been produced, one with the hypothesis $\tilde{m}_{\widetilde{\chi}_2^0}\tilde{m}_{\widetilde{\chi}_1^0} > 0$ (upper) and one with the hypothesis $\tilde{m}_{\widetilde{\chi}_2^0}\tilde{m}_{\widetilde{\chi}_1^0} < 0$ (lower). The different reweighting scenarios affect the $M(\ell\ell)$ distribution and that directly impacts the limit plots: The expected limit line is generally better for the positive reweighting hypothesis, since, according to Fig. 4.6, it leads to a more narrow $M(\ell\ell)$ bump for the signal, resulting in better signal and background separation.

This is especially true for higher masspoints, $\Delta m \gtrsim 15\,\text{GeV}$. For lower masspoints, the shift of the negatively reweighted $M(\ell\ell)$ distribution balances the absence of a more pronounced bump and the expected sensitivity is comparable for the two reweighting scenarios. For both scenarios, the sensitivity reaches up to 290 GeV for $\Delta m \approx 10\,\text{GeV}$.

There is a noticeable mismatch between the expected and the observed limit lines for masspoints with $\Delta m \gtrsim 20\,\text{GeV}$, where the observed limit is more than 1σ weaker. The effect is more prominent in the negative reweighting scenario, owing to the shift of $M(\ell\ell)$ shape in that case. This mismatch is explained by higher than expected data yields in specific SR bins:

- In the $20 < M(\ell\ell) \leq 30\,\text{GeV}$ bin of the 2ℓ-Ewk SR ultra-MET bin (Table 4.10),
- In the $10 < M^{\text{min}}_{\text{SFOS}}(\ell\ell) \leq 30\,\text{GeV}$ bin of the 3ℓ-Ewk SR low-MET bin (Table 4.11),
- In the $10 < M^{\text{min}}_{\text{SFOS}}(\ell\ell) \leq 20\,\text{GeV}$ of the WZ-like selection SR low- and high-MET bin (Table 4.12).

The signal yields of intermediate and high Δm masspoints mostly populate these bins, resulting in a higher cross section upper limit for them. Quantifying a hypothetical excess, the maximum local significance reached is 2.4 for the masspoint with $m_{\widetilde{\chi}^0_2} = 125\,\text{GeV}$ and $\Delta m = 40\,\text{GeV}$.

Fig. 4.42 shows the limit plots for the higgsino models, the higgsino SMS model (upper) and the higgsino pMSSM model (lower). Since the $\tilde{m}_{\widetilde{\chi}^0_2}\tilde{m}_{\widetilde{\chi}^0_1}$ product can only be negative in the case where the LSP is mostly higgsino, only one version of the limit plots is given. The higgsino models use the same fitting strategy as the TChiWZ model, with exactly the same bins included. As a result, the data excesses summarized above also affect the higgsino models in a similar way. This is evident in the upper plot of Fig. 4.42 which exhibits the same difference between the observed and expected limit lines for intermediate and high Δm masspoints. The maximum excluded $m_{\widetilde{\chi}^0_2}$ for this model is 205 GeV for a $\widetilde{\chi}^0_2$-$\widetilde{\chi}^0_1$ mass difference of $\sim 7.5\,\text{GeV}$, while $\widetilde{\chi}^0_2$ masses are excluded up to 150 GeV down to mass difference of 3 GeV.

The connection of the pMSSM parameters, as plotted in the lower plot of Fig. 4.42, with the physical measurable quantities, such as the masses of the particles and the mass difference between them, is slightly more complicated. Up to first order corrections, the μ parameter is equal to the mass of the $\widetilde{\chi}^0_1$,

which is close to the $\widetilde{\chi}_2^0$ mass in the compressed mass spectrum that this search probes. Consequently, the horizontal axis of the higgsino pMSSM model roughly corresponds to the horizontal axis of the higgsino SMS model, i.e. $m_{\widetilde{\chi}_2^0}$. The mass difference between the LSP and the NLSP approximately scales as $\frac{m_W^2}{M_1}$ in this case, therefore the lower part of the vertical axis of the higgsino pMSSM model corresponds to the upper part of vertical axis of the higgsino SMS model and vice versa. Moreover, low values of Δm are mapped on a greater range of M_1 compared to high values. With these considerations, it is concluded that the higgsino pMSSM limit plot is very similar to the higgsino SMS limit plot, plotted on a different parameter space. Signal masspoints with μ up to 180 GeV are excluded for M_1 values of $\sim$ 800 GeV.

Finally, the T2BFF$\widetilde{\chi}_1^0$ and T2BW limit plots are shown in the upper and lower plots of Fig. 4.43 respectively. For these models, the $M(\ell\ell)$ reweighting procedure is no longer relevant, however the W boson branching fraction reweighting still applies. The maximum sensitivity is found in the Δm range 20–30 GeV, since the acceptance for masspoints with $\Delta m \lesssim 20$ GeV rapidly decreases due to the lower bound p_{T} requirements of the analysis. For a mass difference between the $\widetilde{\mathrm{t}}_1$ and the $\widetilde{\chi}_1^0$ of 30 GeV, $m_{\widetilde{\mathrm{t}}_1}$ is excluded up to 540 GeV for the T2BFF$\widetilde{\chi}_1^0$ model and up to 480 GeV for the T2BW model.

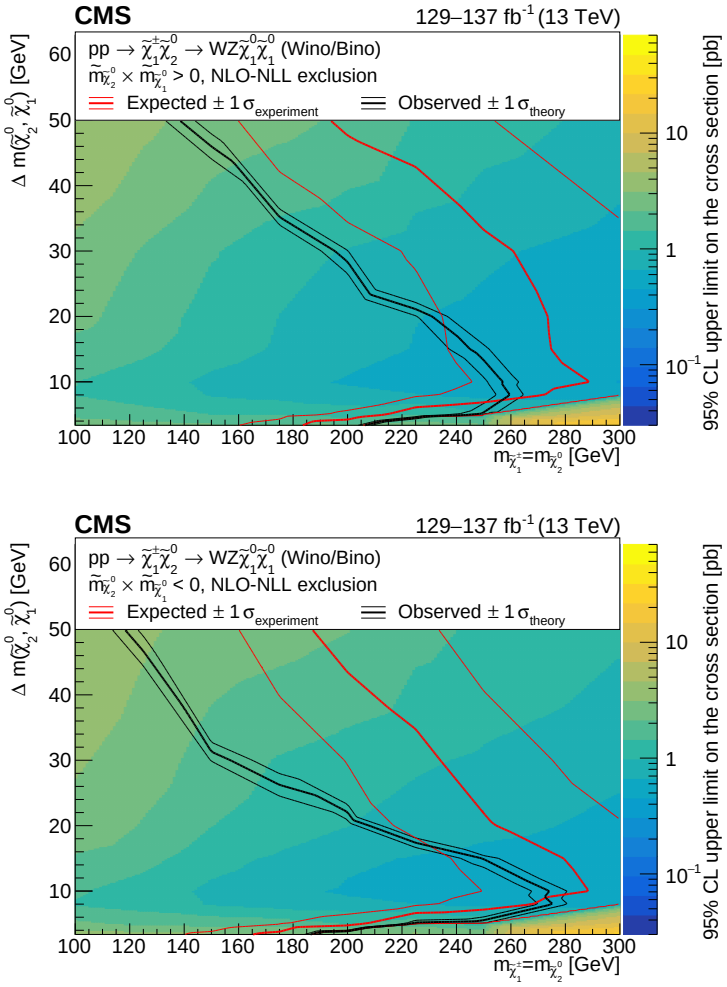


Figure 4.41: The observed 95% CL exclusion contours (black curves) assuming the NLO+NLL cross sections, with the variations (thin lines) corresponding to the uncertainty in the cross section for the TCHIWZ model. The red curves present the 95% CL expected limits with the band (thin lines) covering 68% of the limits in the absence of signal. The results are reported for the $\tilde{m}_{\widetilde{\chi}_2^0}\tilde{m}_{\widetilde{\chi}_1^0} > 0$ (< 0) $M(\ell\ell)$ spectrum reweighting scenario in the upper (lower) plot. The range of luminosities of the analysis regions included in the fit is indicated on the plot.

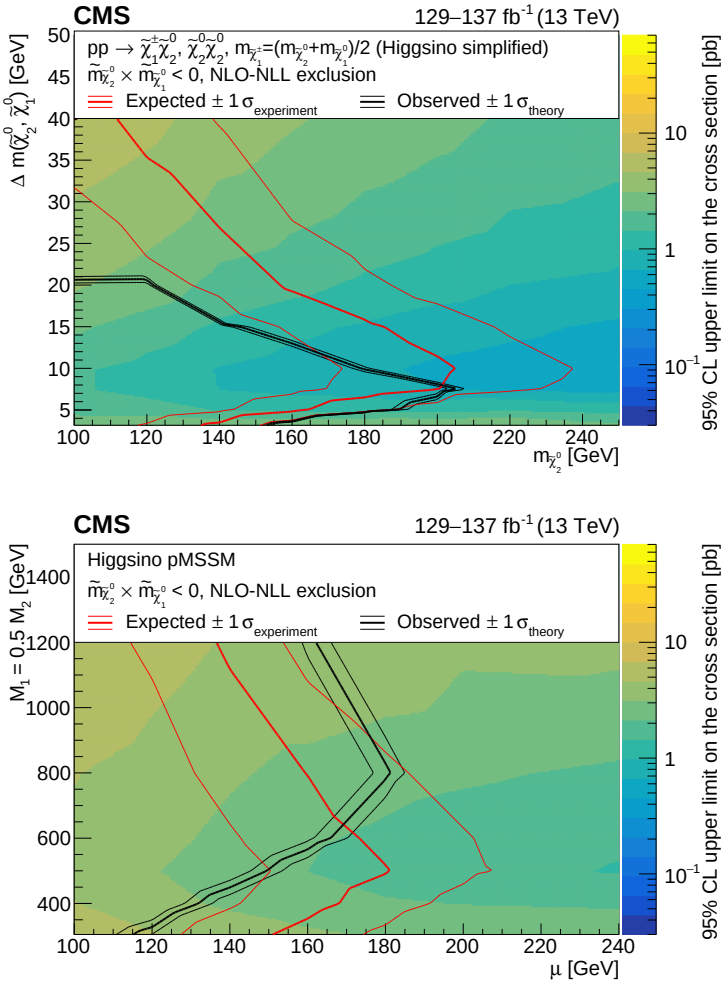


Figure 4.42: The observed 95% CL exclusion contours (black curves) assuming the NLO+NLL cross sections, with the variations (thin lines) corresponding to the uncertainty in the cross section for the simplified (upper) and the pMSSM (lower) higgsino models. The red curves present the 95% CL expected limits with the band (thin lines) covering 68% of the limits in the absence of signal. The results are reported for the $\tilde{m}_{\widetilde{\chi}_2^0}\tilde{m}_{\widetilde{\chi}_1^0} < 0$ $M(\ell\ell)$ spectrum reweighting scenario. The range of luminosities of the analysis regions included in the fit is indicated on the plot.

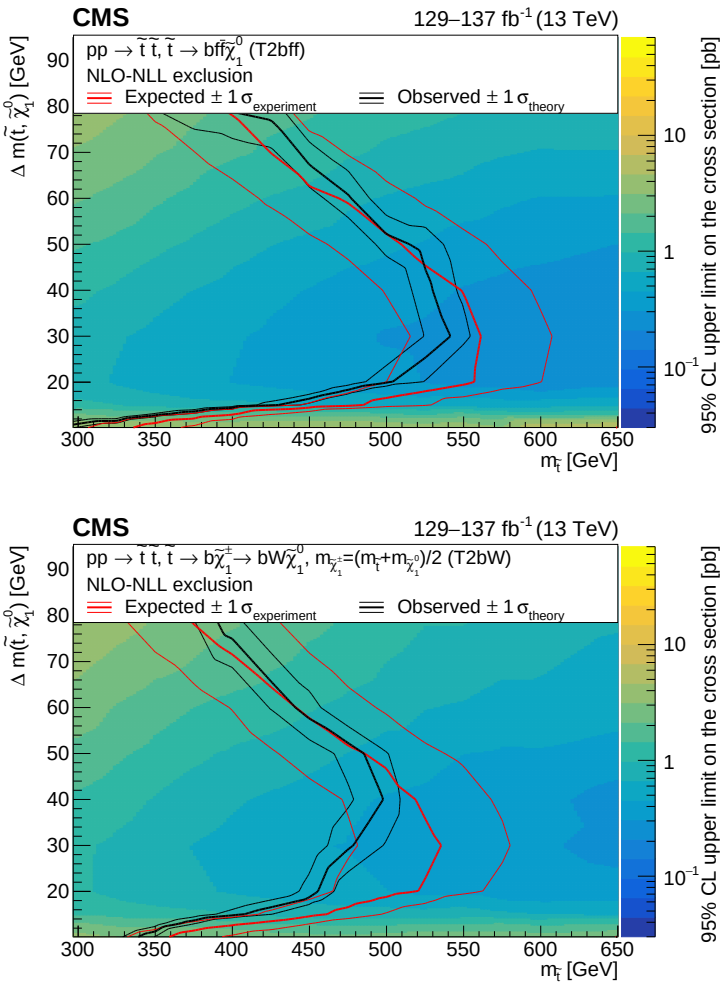


Figure 4.43: The observed 95% CL exclusion contours (black curves) assuming the NLO+NLL cross sections, with the variations (thin lines) corresponding to the uncertainty in the cross section for the T2Bff$\widetilde{\chi}_1^0$ (upper) and T2BW (lower) simplified models. The red curves present the 95% CL expected limits with the band (thin lines) covering 68% of the limits in the absence of signal. The range of luminosities of the analysis regions included in the fit is indicated on the plot.

Chapter 5

The CMS Phase 2 Trigger Upgrade

5.1 The High Luminosity LHC

The LHC is the largest accelerator complex in the world. Given the huge amount of work, infrastructure and, consequently, time for a next-generation project with similar goals and extended reach, the LHC will remain the central hub for the experimental high energy community in many years to come. As a result, there is a global effort to exploit its capabilities at the fullest. The direction of such an exploitation is given in physical terms by the fact that the timescale needed for the reduction the statistical uncertainty of a measurement after 2020 is approximately ten years. This leads to the conclusion that a significant enhancement of the LHC luminosity is required for its continuing utility. In 2016, a proposal towards this end was made by the CERN management: An upgrade to the LHC machine to increase its design luminosity by five times, producing, in the long run, ten times the design integrated luminosity. The High Luminosity LHC (HL-LHC) project, as it was called, was officially approved by the CERN Council on 16–17 June 2016, twenty years since the final approval of the original LHC. The LHC/HL-LHC baseline program, as it was shaped in 2015, while the HL-LHC was initially conceived is shown in Fig. 5.1. Note that this program is constantly under reconsideration, especially after the delays incurred by the COVID-19 pandemic.

The target of ten times the integrated Run 1 and Run 2 luminosity cor-

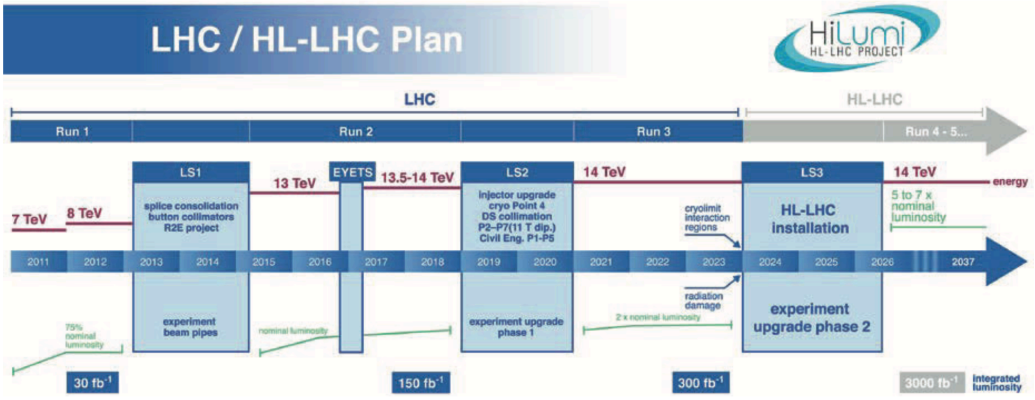


Figure 5.1: The (HL-)LHC program, as it was defined in 2015. Currently, the schedule has been shifted by one year later, due to the COVID-19 repercussions [127].

responds to a HL-LHC integrated luminosity of $3000\,\mathrm{fb}^{-1}$. Due to the phenomenon of luminosity burn-off, according to which the decrease of luminosity as a function of time is larger for higher luminosity values, the HL-LHC integrated luminosity target would require very high peak luminosity values ($> 10^{35}$ cm^{-2}s^{-1}). However, such peak luminosity values prove problematic when taking into account the energy deposition from debris of the collisions and the amount of PU produced in the experimental detectors. A proposed solution to this is the "luminosity levelling", i.e. the limitation of the peak luminosity to a value lower than the maximum one but for longer time, in such a way that the average luminosity is almost equal to that without levelling (Fig. 5.2). One method to achieve luminosity levelling is by varying the beta-star function but other possibilities are explored as well.

There are multiple modules of the current LHC machine that must be replaced or upgraded in order to attain the HL-LHC luminosity goals. Characteristic examples of such systems are: The inner triplet magnets, at $300\,\mathrm{fb}^{-1}$, will have received a dose of 30 MGy, increasing significantly the possibility of serious radiation damage and need to be replaced proactively. The cryogenic system can be improved by making independent different parts of the magnet cooling system, so that, in case of intervention, only specific, small parts of the magnet system need to warm-up, instead of entire sectors. The machine protection system needs to be upgraded to better account for the increase beam intensities; improvements to this system implies upgrades to

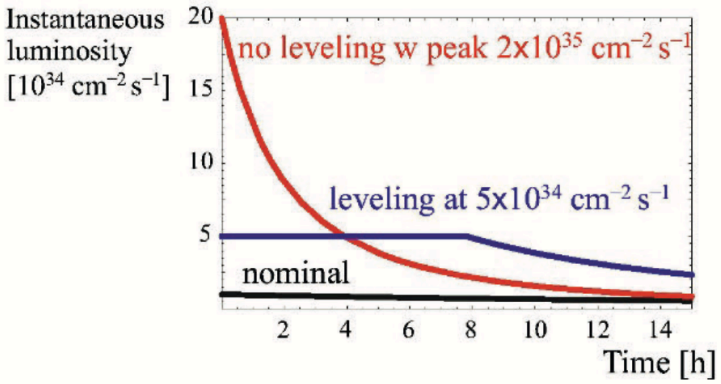


Figure 5.2: The luminosity profile with luminosity levelling (blue line) and without (red line), compared to the nominal LHC design profile (black line) [127].

hardware modules such as the collimation or the quench protection system. The electronic board system is radiation-sensitive and needs to be changed with radiation-hard cards or even be moved to underground locations where they receive small radiation doses.

The HL-LHC upgrade aims at the modification of key beam parameters that allow reaching the integrated luminosity target. The entirety of the hardware upgrades will enable to arrive at the beam current value needed for the HL-LHC performance, which exceeds the LHC nominal value by 30%. The increase in beam brightness is achieved by the raising the number of protons per bunch, while maintaining or even decreasing the emittance. A method of principal importance for increasing the instantaneous luminosity consists in minimizing the beta-star function with the use of the upgraded magnet triples. However, a small β^* usually comes with a large bunch crossing angle (θ_c). This reduces the geometric factor, therefore limiting the benefit of such a configuration. The curve on Fig. 5.3 shows the change in the geometric factor, here denoted as R, as a function of β^*. An efficient and also elegant solution to this issue is the usage of superconducting crab cavities (CC). Their purpose would be to rotate the incoming bunches by $\theta_c/2$, hence setting up the bunch for a head-on collision, as shown in the lower right corner of Fig. 5.3. The gain in the geometric factor induced by the utilization of CC for the same β^* value is also shown in the same figure. In addition to the above, the CC can be exploited to facilitate luminosity leveling, providing

another means to control the peak luminosity, and PU distribution. A list of parameters updated to achieve the HL-LHC configuration are summarized in Table 5.1.

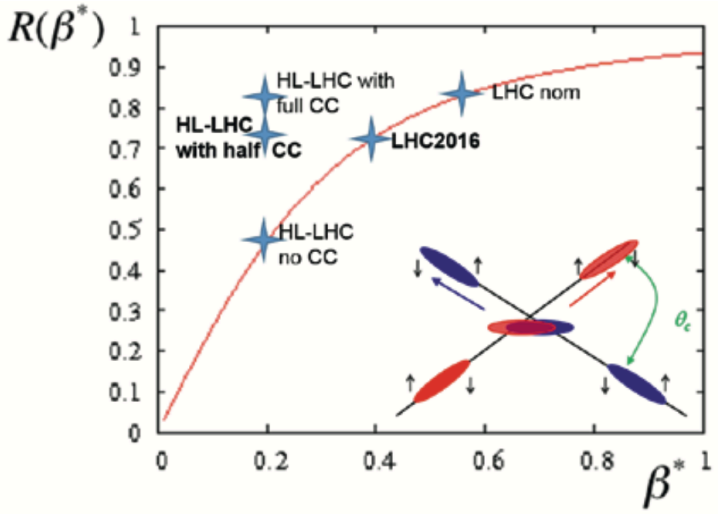


Figure 5.3: The geometric factor R as a function of β^*. The machine configuration without CC is shown as a curve and the scenarios with 4 CC (full CC) and 2 CC (half CC) per IP are also marked for $\beta^* = 0.2$ m. The function of a CC is also shown in the bottom right of the figure [127].

Parameter	Symbol	Nominal Value
Center-of-mass energy	$\sqrt{s}$	14 TeV
Peak instantaneous luminosity (w/ CC)	L_{inst}	$12.6 \times 10^{34} cm^{-2} s^{-1}$
Number of protons per bunch	N_p	2.2×10^{11}
Number of bunches per beam	n_b	2748
Transverse normalized emittance	ϵ_n	2.50 μm rad
Beta-star function (at the CMS IP)	β^*	0.2 m
Geometric factor at min. β^* (w/o CC)	F_0	0.369

Table 5.1: The updated values of crucial HL-LHC parameters in p-p collisions [127].

5.2 The CMS Phase-2 Level-1 Trigger

The HL-LHC machine upgrade is accompanied by plans from the LHC experiments to consolidate and improve their detectors. The CMS experiment has envisaged, sketched out and start realizing an ambitious Phase-2 upgrade strategy for benefiting from the HL-LHC running period with the maximum physics performance [128]. More specifically, the following major upgrades are planned:

- Complete replacement of the strip and pixel silicon trackers with brand new sensors. As part of this upgrade, the tracker acceptance will be extended up to $|\eta| = 3.8$, while also providing enhanced longitudinal and transverse resolution [129].

- Complete replacement of the barrel calorimeter readout electronics that permit capturing data with finer granularity as well as with timing information [130].

- Complete replacement of the endcap calorimeter with a brand new high granularity calorimeter (HGCal). The HGCal is a sampling calorimeter, composed of silicon sensors and scintillator tiles, that is designed to withstand the huge radiation doses sustained during the HL-LHC operation. It provides unprecedented longitudinal and transverse granularity for electromagnetic and hadronic showers, along with precise timing information [131].

- Complete replacement of the muon system electronics and extension of acceptance up to $|\eta| = 2.8$ with the addition of improved RPC (iRPC) and GEM detectors [132, 133].

- Installation of a new minimum ionizing particle (MIP) timing detector (MTD) at the inner surface of the barrel and endcap calorimeters with the aim of precising measuring the timing of charged tracks [134].

- Upgrade of the totality of the L1, HLT and data acquisition (DAQ) systems to accommodate the higher collision rate and PU of the HL-LHC operation conditions [135, 136]. The upgraded Phase-2 L1 system is the subject of the rest of this section.

Apart from the CMS detector upgrades described above, commercial technological developments also played an important role in allowing for a major expansion of the CMS L1 system capabilities. The availability of field programmable gate arrays (FPGAs) capable of housing data input/output (I/O) and logic operations of increasing speed, complexity and volume open up the possibility of implementing complex algorithms and machine learning techniques at the L1 system. To complement this, high-speed optical links are used that can transfer information between different detectors in the tight time limits of the L1 trigger, hence enabling the combination of data and giving a global view of the whole detector. The above components are incorporated in a modular architecture that allows for flexibility depending on the running conditions and extended redundancy and contingency.

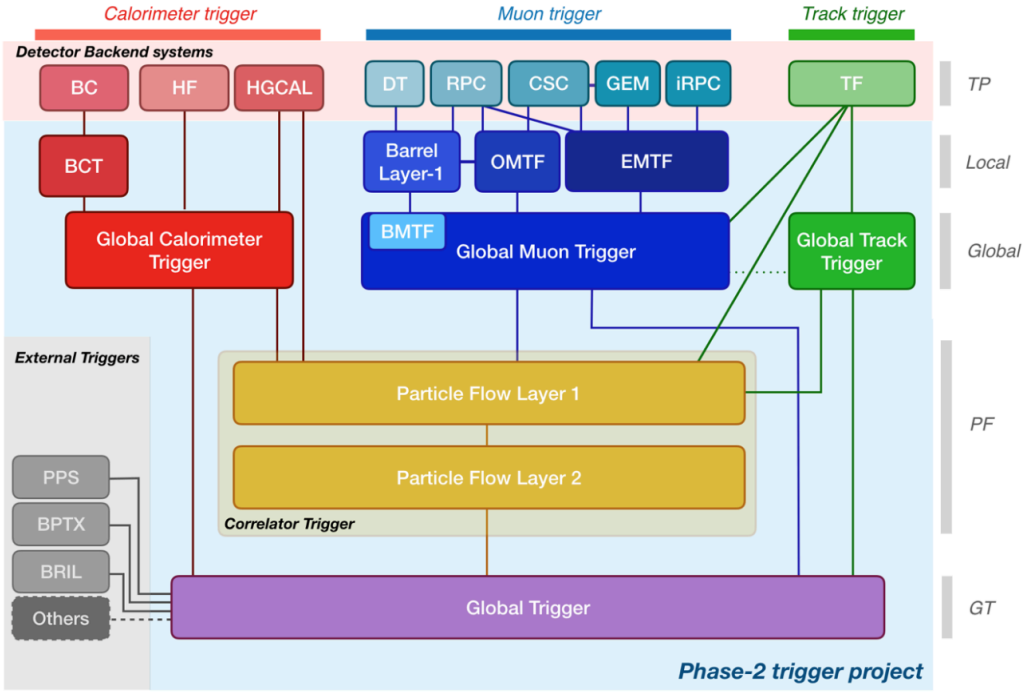


Figure 5.4: Functional diagram of the CMS Phase-2 L1 Trigger (blue box), displaying the different subdetectors and subsystems and the connections between them. Dashed lines represent other potential connections. The gray column on the right designates different levels of processing: Trigger primitives (TP), local and global reconstruction of standalone trigger subsystems, PF reconstruction and global trigger (GT) decision [135].

Fig. 5.4 shows an overview of the CMS Phase-2 L1 trigger system. The main changes with respect to the Phase-1 design is the presence of the new track trigger (L1TT) and correlator trigger (L1CorT), while the muon trigger (L1MT), the calorimeter trigger (L1CT) and the global trigger (L1GT) have been also consolidated. 12.5 μs latency, 750 kHz output rate.

One of the novelties of the new CMS L1 trigger and one of the main drivers of its significantly improved performance is the inclusion of tracks with $p_{\mathrm{T}} > 2\,\mathrm{GeV}$ and $|\eta| < 2.4$. This is possible by exploiting the architecture of the upgraded outer tracker: The closely-spaced silicon modules of each layer allow for an approximate measurement of the bending of each track. The bending is a proxy for the p_{T} of the tracks and, hence, hits from low-p_{T} tracks can be discarded early on, reducing the computational complexity of the track building algorithm. This leads to a manageable rate of tracks for the L1 system ($\sim 10^3$ tracks per PU 200 event). The on-detector electronics of each outer tracker layer generate "stubs" out of the hits that pass the p_{T} threshold and these stubs are later propagated to the trigger finder system. As a baseline, a Kalman filter fit with beamspot constraint is performed on associated stubs and the resulting track contains information on the full set of fitted parameters (p_{T}, η, ϕ, d_0, z_0 and track quality). As a extension to this, studies are ongoing for the inclusion of an "extended L1 tracking" that could reconstruct displaced tracks with high efficiency by utilizing different seeds for the track building. The efficiency for reconstructing tracks at the L1 trigger system as a function of η is shown in the left plot Fig. 5.5, while preliminary results on the benefits of the use of the extended L1 tracking is showcased in the right plot of the same figure.

The main upgrades of the barrel calorimeter L1 system comes from the replacement of the EB and HB electronics. The new electronics enable the transmission of single crystal information in the EB and depth and timing information in the HB, resulting in a 25- and 4-fold increase of granularity respectively. As mentioned above, the endcap calorimeter will be entirely replaced with the new HGCal detector, consisting of 28 electromagnetic and 22 hadronic layers and more than 6×10^6 channels in total, As a result, the granularity of the endcap calorimeter is improved by a factor of 500 with respect to the Phase-1 system. To fully exploit both the transverse and longitudinal information, a 3D clustering algorithm is implemented. The created clusters can potentially carry a large variety of information for the subsequent classification of their origin as electromagnetic, hadronic or PU. An algorithm for this purpose is discussed in Sec. 5.4.1 in the context of

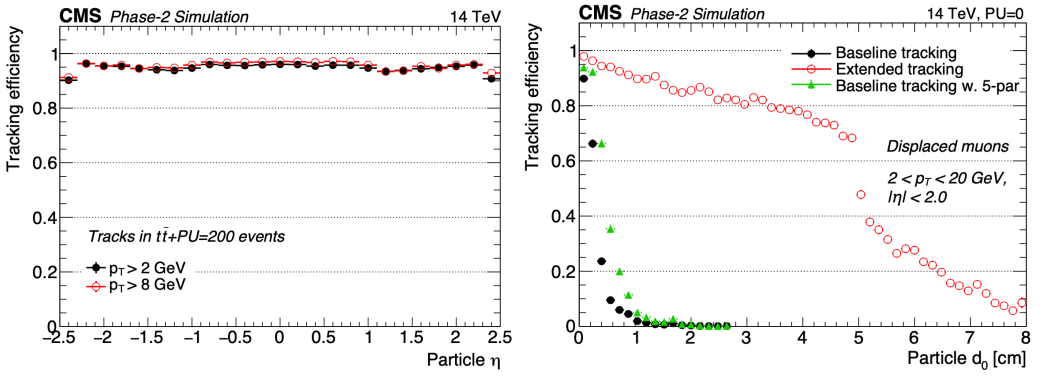


Figure 5.5: Left: Efficiency of L1 tracks as a function of η in a $t\bar{t}$ sample at PU 200. The black(red) markers indicate the efficiency of tracks with $p_{\mathrm{T}} > 2(8)\,\mathrm{GeV}$. Right: The L1 tracking efficiency as a function of d_0 in a displaced muon sample at PU 0. The black points correspond to the baseline track building algorithm with beamspot constraint ($d_0 = 0$), the green points to the same algorithm but without beamspot constraint, while the the red points to the extended tracking algorithm, i.e. without beamspot constraint and with alternative seeds [135].

the L1CorT. The number of clusters for different E_T thresholds is shown in Fig. 5.6 for the barrel (left) and the endcap (right) at PU 200. The very large number of clusters produced in the HGCal, even in the case of a moderate E_T threshold, is a major motivation for development of cluster ID algorithms in the downstream L1 systems (Sec. 5.4.1).

The replacement of the DT and RPC electronics result in finer spacial and timing granularity (1.5 μs), as well as better bunch crossing assignment for the barrel L1MT. In the endcap L1MT, the new CSC electronics can tolerate higher rates, anticipated in the HL-LHC running conditions. With the installation of the GEM detectors, an enhancement of the muon bending direction is also achieved for $1.6 < |\eta| < 2.8$. Finally, the addition of the iRPC completes the redundancy of the the muon system, which very important, especially at the L1 trigger system, in case of accidental or ageing-induced failure of one of the subdetectors.

The L1 trigger system is responsible for capturing all the data that are used in CMS analyses. A large scale failure in this system would deem the CMS detector largely inoperational and, as a result, its redundancy is of paramount importance. Apart from the redundancy in the L1 trigger

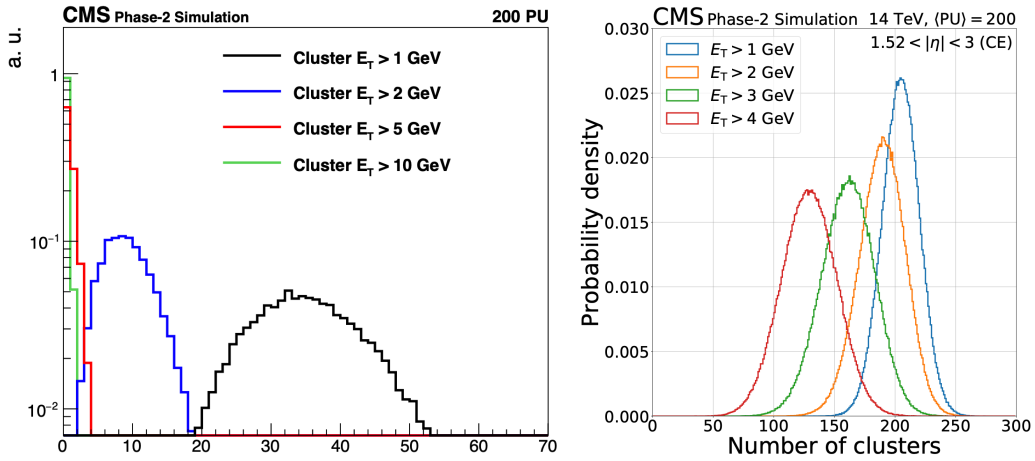


Figure 5.6: Distribution of the number of 3×5 clusters in the barrel (left) and HGCal clusters (right) at PU 200 for different E_T thresholds [135].

subsystems, mentioned above, the robustness of the L1 trigger algorithms is also provisioned. To this end, four different paths for data processing are planned, each one independent from the others, in order to provide maximum flexibility: The tracker, the calorimeter, the muon and the PF paths.

Starting off with the tracker path, tracker-only jets and $p_{\mathrm{T}}^{\mathrm{miss}}$ can be produced. Moreover, the tracker data processing path is responsible for the reconstruction of the primary vertex (one even multiple candidates of it) for use within the path and for its propagation to downstream subsystems. The calorimeter path uses information only from the EB, HB and HGCal subdetectors to create high-resolution clusters. These are used to either reconstruct calorimeter-only e/γ, hadronic τ, jet and energy sum objects or to be propagated downstream. The muon path involves the reconstruction of muon tracks in three separate regions, as in Phase-1: barrel, overlap and endcap. These muon tracks are subsequently passed on to the global muon trigger (GMT) for clearing duplicate and misreconstructed tracks. At the GMT, apart from the standalone muon tracks, track-matched and tracks matched to muon stubs ("tracks+μ-stubs") objects are formed, and all of them are sent to downstream subsystems. The PF data processing path takes place in the L1CorT. The L1CorT is a new L1 subsystem which operates as a hub of information from all upstream systems mentioned above and its goal is to perform global event reconstruction, producing higher-level trigger objects. For this purpose, simplified versions of the PF and the PUPPI algorithms

have been developed. A significant proportion of the work described in this thesis has been carried out on this subsystem and, hence, Sec. 5.4 is dedicated to a more detailed discussion on it.

The outputs from the L1TT, the L1CT, the L1MT and the L1CorT are transferred to the L1GT, where a set of selection algorithms, collectively called "L1 trigger menu", determines whether an event will be kept for offline processing or be discarded. Information from external triggers, such as the precision proton spectrometer (PPS), beam position and timing monitors (BPTX) and the luminosity and beam monitoring (BRIL) detectors, is also included. A positive decision by the L1GT initiates a full readout by the detector backend systems to the DAQ. A second, parallel readout system is also envisaged. This system constitutes a triggerless data capturing chain, which allows the acquisition of data at the (HL-)LHC clock frequency, i.e. at 40 MHz, albeit with lower resolution compared to the offline events. The system, called the "40 MHz scouting" system, is described in greater detail in Sec. 5.3, where emphasis is given on its demonstrator which has been implemented during the Run 2 running period.

Finally, a significant amount of flexibility is planned to be implemented in the architectural design of the L1 trigger system. More specifically, the choice of commercially available, general purpose processors, instead of custom-designed ones, allows the same set of boards to be used for any function in any subsystem, according to needs. In addition to that, regional and time-multiplexing (TMUX) techniques [137] are being employed to minimize the number of processing boards and interconnection links, while achieving low latency and increased system redundancy. The current L1 trigger architecture takes into account latency and bandwidth requirements of potential additions or improvements of the system, for example the inclusion of timing information from the MTD, that are up till now not in the baseline design.

With the set of upgrades summarized in the previous paragraphs, the physics reach of the CMS L1 trigger system is not only preserved with respect to Phase-1, despite the much harsher running condition, but it is extended as well. One of the main targets of the HL-LHC is the precision measurements on the characteristics of the Higgs Boson with the ultimate goal of measuring its self-coupling, mainly via the Higgs boson pair production. This effort is facilitated with the improvements of Phase-2 L1 trigger system regarding final states with leptons (HW/Z $\rightarrow$ leptons, HH $\rightarrow \tau_{\ell/\mathrm{h}}\tau_{\mathrm{h}}$bb - Fig. 5.7), photons (H $\rightarrow \gamma\gamma$, HH $\rightarrow \gamma\gamma$bb - Fig. 5.8, left), multiple jets (HH $\rightarrow$ bbbb - Fig. 5.8, right) or $p_{\mathrm{T}}^{\mathrm{miss}}$ (HH $\rightarrow \nu\nu$bb - Fig 5.9). SM precision measurements,

such as single top and top pair production (Fig 5.7, left), and BSM searches, such as SUSY models with compressed mass spectrum (Fig. 5.9) are also going to be benefited. The access to tracking information allows for the efficient selection of processes with light mesons, e.g. $B \to \phi\phi \to 4K$, or of lepton flavor violating processes, most importantly $\tau \to \mu\mu\mu$. The extended $|\eta|$ coverage significantly increases the acceptance for analyses measuring the SM processes, such as the $t\bar{t}$ differential cross section or double parton scattering. The enhanced capabilities of the L1 trigger system now include the reconstruction of displaced objects (muons or jets), probing exotic physics signatures, and the implementation of machine learning techniques that can potentially improve the selection efficiency of L1 trigger algorithms.

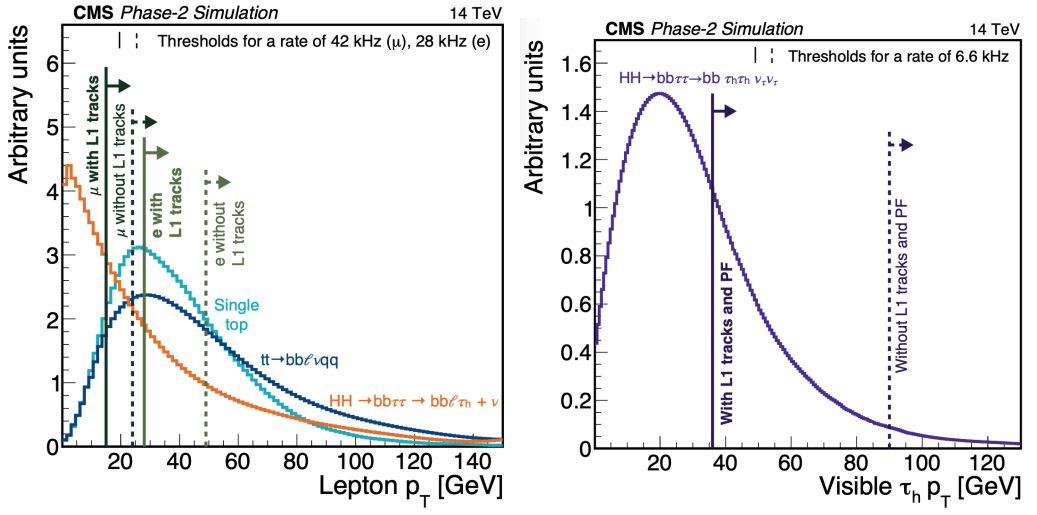


Figure 5.7: Left: The p_{T} distribution of electrons and muons in a HH $\to \tau_\ell\tau_{\mathrm{h}}$bb (orange), a semileptonic $t\bar{t}$ (blue) and a single top production (cyan) simulated sample at PU 200. The solid(dashed) vertical lines indicate the L1 trigger threshold (95% efficiency) for single lepton L1 seeds with(without) the use of L1 tracks. The rate is fixed at 28 kHz for electrons and at 42 kHz for muons. Right: The p_{T} distribution of the lowest-p_{T} τ_{h} in a HH $\to \tau_{\mathrm{h}}\tau_{\mathrm{h}}$bb simulated sample at PU 200. The solid(dashed) vertical lines indicate the L1 trigger threshold (50% efficiency) for single τ_{h} L1 seed with(without) the use of L1 tracks and L1 PF inputs. The rate is fixed at 6.6 kHz [135].

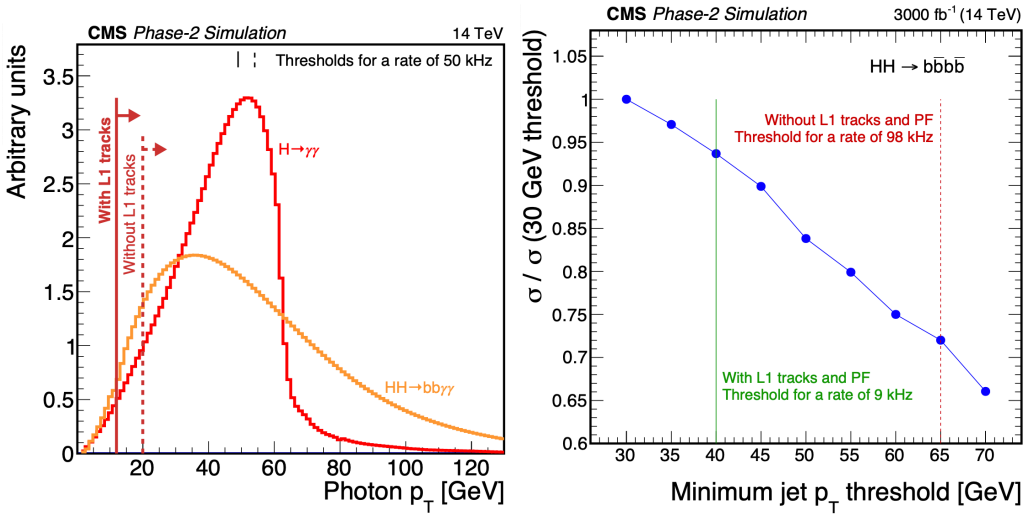


Figure 5.8: Left: The p_{T} distribution of the lowest-p_{T} photon in a H $\rightarrow$ $\gamma\gamma$ (red) and a HH $\rightarrow$ $\gamma\gamma$bb (orange) simulated sample at PU 200. The solid(dashed) vertical lines indicate the L1 trigger threshold (95% efficiency) for single photon L1 seed with(without) the use of L1 tracks. The rate is fixed at 50 kHz. Right: The expected loss of signal significance, compared to the one at 30 GeV, as a function of minimum jet p_{T} threshold in a HH $\rightarrow$ bbbb simulated sample at PU 200. The solid(dashed) vertical lines indicate the L1 trigger threshold for a multi-jet L1 seed with(without) the use of L1 tracks and L1 PF inputs for a rate of 9(98) kHz [135].

5.3 40 MHz Muon Scouting

The implementation of the 40 MHz scouting system, a second, independent readout chain that is capable of capturing data from every HL-LHC collision, can have important benefits, both for physics studies and system validation. Starting from the physics potential, there is a number of rare Higgs rare decays, such as H $\rightarrow$ J/$\psi\gamma$, H $\rightarrow$ $\phi\gamma$ or H $\rightarrow$ $\rho\gamma$, that exhibit limited acceptance due to the threshold of the single photon L1 seeds. The 40 MHz scouting could help in increasing the acceptance of a physics analysis targeting these decays by providing data with a very low or even no threshold on the photon p_{T}, nonetheless with the complication of lower resolution. Possibly affected less by the decreased resolution but substantially benefited by the high statistics that the 40 MHz scouting can provide, lepton flavor violating analyses

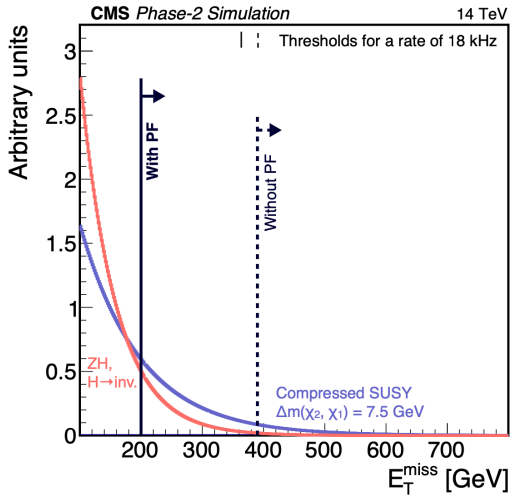


Figure 5.9: The $p_{\mathrm{T}}^{\mathrm{miss}}$ distribution in a ZH $\rightarrow \nu\nu$bb (red) and a $\widetilde{\chi}_1^{\pm}\widetilde{\chi}_2^0 \rightarrow \mathrm{W}^*\mathrm{Z}^*\widetilde{\chi}_1^0\widetilde{\chi}_1^0$ ($m(\widetilde{\chi}_1^{\pm}) = m(\widetilde{\chi}_2^0) = 300\,\mathrm{GeV}$ and $m(\widetilde{\chi}_1^0) = 292.5\,\mathrm{GeV}$) simulated sample at PU 200. The solid(dashed) vertical lines indicate the L1 trigger threshold (95% efficiency) for $p_{\mathrm{T}}^{\mathrm{miss}}$ L1 seed with(without) the use of L1 PF inputs. The rate is fixed at 18 kHz [135].

can take advantage of the scouted data. This is especially true for signal processes with τ leptons in the final state that are also generally limited by the threshold of the single object L1 seeds. Furthermore, the 40 MHz scouting opens up the possibility of extensive displaced muon studies, even at low p_{T}, and allows for studying longlived signals by exploiting cross-bunch crossing correlations. Finally, QCD measurements with very high statistics in regions of the phase inaccessible to the standard L1 trigger seeds due to intolerable rate could be possible with the data captured with the 40 MHz scouting system.

Data collected by 40 MHz scouting system can have applications in monitoring and testing critical detector systems. Given that it is going to receive the output of all L1 trigger subsystems, hence the full input and output of the L1GT, the 40 MHz scouting can verify the correct L1GT operation and also help test and prototype its algorithms, without actually affecting the main readout chain of CMS. The continuous, real-time acquisition of data can be used for for promptly identifying potential subdetector issues by creating live heat maps of their I/O data flow during runtime. Since this data will be stored as well, the 40 MHz scouting allows for the offline examination

of problematic incidents with the goal of designing improved solutions and precautions. Finally, an application of the 40 MHz scouting system that was even showcased in its the Run 2 demonstrator is that it provides an alternative to standard luminosity measurements. Fig. 5.10 shows how the number of muons captured by the Run 2 40 MHz scouting demonstrator follows an emittance scan, in agreement with other luminometers.

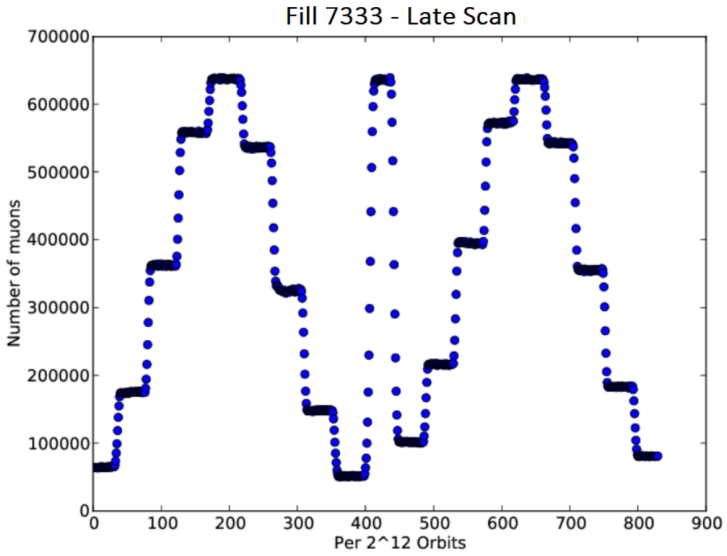


Figure 5.10: The distribution of the number of muons as a function of the number of orbit during an LHC emittance scan. The distribution follows the distribution of luminosity, as measured by other luminometers (not shown in the plot) [135].

The inclusion of the 40 MHz scouting system does not modify the Phase-2 L1 trigger system architecture. Instead, the data from the different subsystem will be transmitted to the 40 MHz scouting system with spare links from each subsystem processing board. However, dedicated boards receiving the input, firmware and software development for its preprocessing and a S/T storage for saving the data will be needed. Fig. 5.11 shows the Phase-2 L1 trigger design on the left side with the addition of the 40 MHz scouting system on the left, along with the links connecting them. Beginning from bottom to top, the scouting decision system (sDS) receives the output of the L1GT system, along with other information it contains, such the information of the prescales of L1 seeds and the inputs of external triggers, to which the

40 MHz scouting system is not connected. Moving further up, the scouting global system (sGS) captures with output of the L1CT, L1CorT, L1TT and L1MT. This output is the same as the input to the L1GT, hence the 40 MHz scouting system gets the full I/O data flow of the L1GT. For each subdetector, there is also a corresponding scouting local system (sLS): sLS(mu) for the L1 BMTF, the OMTF and the EMTF output and sLS(calo) for the regional barrel calorimeter trigger (RCT), the HGCal and HF output. The sLS receiving the track finder outputs is called scouting track system (sTS). Another scouting system capturing the primitives from the calorimeter backends (sPS) has been considered and is included in dashed lines in Fig. 5.11, however further studies are required for the feasibility and the usefulness of its implementation. The sDS and the sGS are collectively referred to as the stage1 scouting system, while all the SLSs together form the stage2 scouting system.

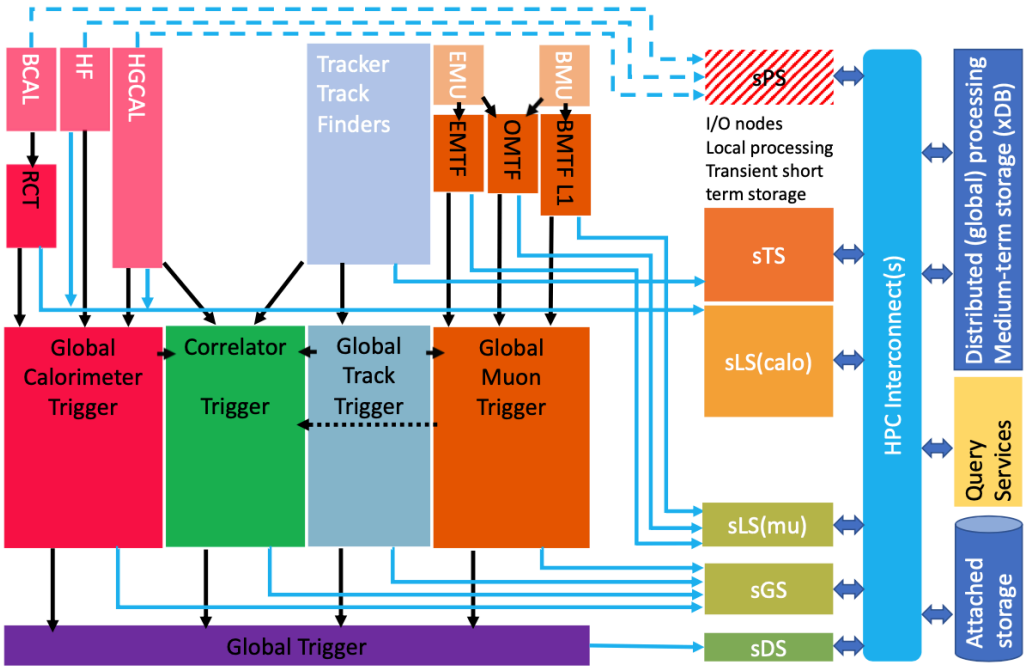


Figure 5.11: Overview of the architecture of the CMS Phase-2 L1 trigger system with the inclusion of the 40 MHz scouting system. Dashed lines(boxes) indicate potential connections(subsystems) that are not in the current baseline [135].

The stage1 and stage2 scouting systems input boards perform zero-suppression and possibly a rudimentary selection, correction and/or calibration of the data, with the aim of reducing the huge amount of data by discarding useful information. These boards are then connected to an HPC interconnect (I/C) that distributes the data for storing: A S/T, local storage is envisaged as a buffer to collect the preprocessed scouting data until it is organized into higher-level structures. Following that, the data is saved in a medium-term, distributed-processing storage, from where they can be retrieved for offline analysis.

The different scouting stages have different hardware requirements, depending on their complexity. All the input links to the 40 MHz scouting system are assumed to be of 16 or 20 Gbps speed, TMUXed according to the TMUX factor of the subsystem connected to. For this reason, for the proper combination of events from different subdetectors, each input frame needs to carry information about the orbit number and each one of the bunch crossings. The zero-suppression performed in the input boards can then remove empty bunch crossings or even empty input objects, leading to a variable-size format.

For stage1, a readily-available board, like the Xilinx KCU1500 FPGA Acceleration Development Kit, is needed in terms of processing resources but supporting at least 8 25 Gbps links. Such devices exist, so the stage1 scouting system can be built with present day technology. For stage2, where the data will mostly be transferred via 16 Gbps links, boards with 16 or even 24 links would be optimal, so that the resources of each board are exploited to the fullest. Such devices could become commercially-available within the timescale of Run 4 but could be custom-designed, if needed. For some systems, e.g. the sLS(mu), the input links are under-utilized, so the existence of intermediate, "concentrator" boards could help in making the most out of the hardware used.

Table 5.2 summarizes the hardware resources (boards, links and storage) for the different stages of the 40 MHz scouting system, broken down to separate contributions from each subsystem. The requirements are small for the stage1 scouting system but get progressively larger for stage2 and sPS. As a result, the 40 MHz scouting system design goes as follows: Stage1 is included in the baseline. Stage2 is planned to be implemented with limited resources, so that only one of its subsystems can be supported at any time. For example, when the output of the sTS needs to be captured, then the data of any other sLS are discarded. This design is subject to change, so that data from

multiple sLS can be captured, if a compelling physics study case is found. Finally, the sPS requires an order of magnitude higher resources, so its even its partial implementation is still under discussion.

System	Links [Gb/s]	I/O boards	S/T Storage [TB]	Links to I/C
sDS(GT+TCDS)	3+1 (25/10)	1	small	1
sGS(GMT)	8 (25)	1	0.75	1
sGS(GCT)	8 (25)	1	0.75	1
sGS(GTT)	16 (25)	2	1.5	2
sGS(PF)	32 (25)	4	3	4
Total stage1	67(25)+1(10)	9	6	9
sLS(RCT+HF)	144+36 (16)	22	43	15
sLS(HGCal)	432 (16)	54	50	34
sLS(xMTF)	18+6+12 (16)	6	2.5	3
sTS	324 (25)	41	120	21
Total stage2	648(16)+324(25)	123	230	73
sPS	$\mathcal{O}(10^3)$	$\mathcal{O}(10^2)$	2–5 k	$\mathcal{O}(10^2)$

Table 5.2: Summary of hardware requirements of the 40 MHz scouting system: Number of links from L1 subsystems to the I/O boards, number of I/O boards, S/T storage capacity and number of 200 Gbps links from the I/O boards to the HPC I/C [135].

The feasibility of the 40 MHz scouting system was demonstrated in Run 2 by a scaled-down version of the full system. This demonstrator system captured data only from the muon subsystems and, as a result, is called "40 MHz muon scouting system". The implementation of this system was possible due to the presence of eight spare output links from the Phase-1 global muon trigger (uGMT). These 10 Gbps optical links were used to transfer up to eight highest-ranking final muon candidates (the same as what the GT receives) and up to eight intermediate muon candidates (only used for the uGMT internal debugging purposes) to Xilinx KCU1500 FPGA Acceleration Development Kit, serving as the scouting board. The incoming muon candidates were zero-suppressed by removing empty bunch crossings and counters for data synchronization, e.g. the LHC orbit number, were added. Other auxiliary information, such as the link ID, were also included.

The muon candidates were received by the scouting board, where they underwent a more sophisticated zero-suppression that removed empty muon objects, resulting in a variable-sized format. A 10 GbE connection was used to transfer the data to a 8 TB RAID storage, where they were compressed.

A control machine was also connected to the scouting board with the purpose of automating the data capture procedure to a significant degree, by automatically initiating data-taking or recovery routines when needed.

The 40 MHz muon scouting system was ready for deployment at the end of Run 2 and functioned for the last two weeks of p-p collisions and for the whole HI run of 2018. After compression, the collected data amounted to 1.1 TB/24 h for the p-p run and to 25 MB/μb^{-1} for the HI run. The overall efficiency in capturing the data was approximately 50%, mainly because of the debugging of some system components, such as the link alignment, which had to be done online. It is worth mentioning that the traditional data-taking procedures of the L1 trigger were completely unaffected by the operation of the 40 MHz muon scouting system.

The next step in demonstrating the ability to implement a full-scale 40 MHz scouting system during Phase-2 operation is to capture data from multiple L1 subsystems. For Run 3, it is envisaged that the output of the Layer 2 Calorimeter Trigger will be captured as well. Hence, the entirety of the GT inputs will recorded, similarly to the Phase 2 system. Fig. 5.12 shows a schematic of the Run 2 and Run 3 40 MHz scouting system implementation.

The Run 2 40 MHz muon scouting system was useful not only as a proof-of-concept for the hardware implementation of the Phase 2 40 MHz scouting system but also as a testing ground for the usefulness of scouting data and ways to make the most of them. As previously mentioned, the L1 data captured by the scouting system have very limited resolution, especially during Phase 1. Ideally, corrections and calibrations are to be applied during the preprocessing of the scouting data, in order to improve their quality. Promising avenues to be explored in this respect include the application of multivariate and machine learning techniques that become more and more available at the L1 trigger with the increasing capabilities of hardware devices.

More concretely, in terms of the 40 MHz muon scouting system, there are two possible scouted muon attributes that could benefit from improving their resolution, by applying data-driven calibrations: The muon p_{T} scale and the muon ϕ angle extrapolation. The calibration factors can be extracted by using the offline reconstructed data of a "ZeroBias" sample, i.e. a sample that consists of random events, even if they have not been selected by any L1 seed. The offline reconstruction greatly improves the resolution of the muon attributes and, consequently, the offline data can be used as a reference for the calibration. The choice of the ZeroBias sample is made, so that the

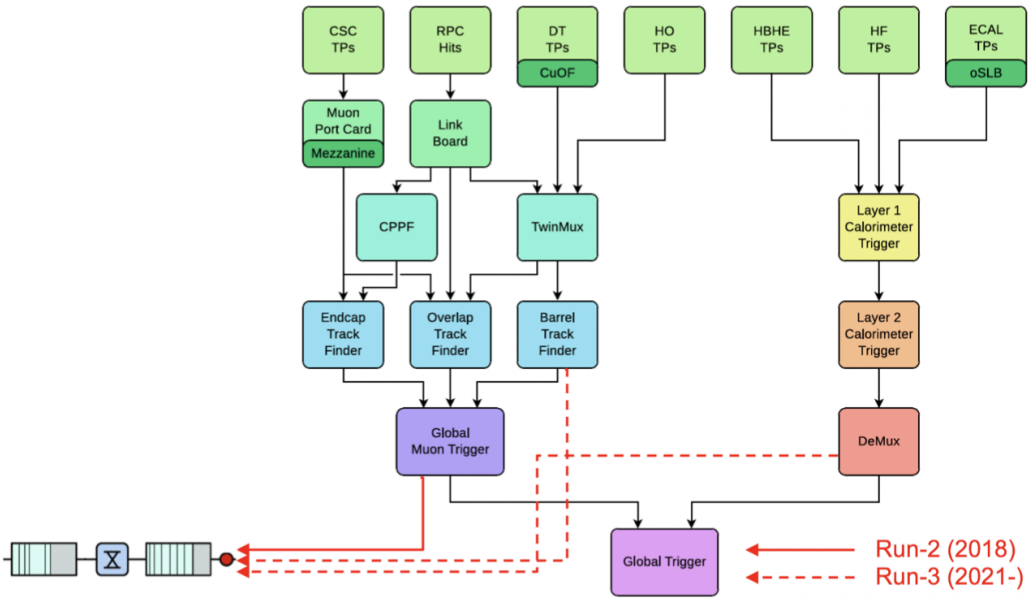


Figure 5.12: Overview of the architecture of the CMS Phase-1 L1 trigger system with the inclusion of the 40 MHz scouting system for Run 2 (solid red lines) and Run 3 (dashed red lines). The 40 MHz scouting system is indicated in the lower left part of the figure [135].

reference sample has the same composition as the scouted data, which are essentially events without any selection. The ZeroBias sample of 2017 is used to extract the corrections, while the ZeroBias sample of 2018 is used to test their performance. The "Charmonium" sample of 2017, mainly consisting of events that have been selected with HLT paths targeting heavy flavor resonances, is also employed for testing the calibrations, in the case that trigger biases are present.

The muon p_{T} scale is generally shifted towards higher p_{T} values. The logic is as follows: Due to the L1 p_{T} resolution, the p_{T} value measured by the L1 trigger is approximately half the times greater than the true value and half the times less than that, given an approximately symmetrical distribution. Therefore, if this L1 trigger p_{T} measurement is used as the L1 seed threshold, then the efficiency at the threshold value is $\sim$50%. However, considering that the events discarded by the L1 trigger are irretrievable, it is preferable that a higher efficiency is achieved at the L1 trigger threshold. As a result, the L1 trigger p_{T} is artificial increased, so that setting the L1 trigger threshold

at a specific value keeps ∼90% of the muons with the given true p_{T} value.

The L1 muon p_{T} can be calibrated back to its nominal scale by correcting with respect to the offline value. To this end, the average ratio of the p_{T} value measured at the L1 trigger (p_{T}(L1)) over the p_{T} value after the reconstruction (p_{T}(reco)) is measured in the ZeroBias 2017 sample in p_{T} bins, separately for each muon track finder. The choice of the binning is driven by the number of yields in different p_{T} ranges. Because the falling p_{T} spectrum of muons, the binning is finer at low p_{T} and gets progressively coarser at higher values. The p_{T}(L1)/p_{T}(reco) value at the highest p_{T} bin available is extended all to higher p_{T} values, where there are no yields. Fig. 5.13 shows the p_{T} scale correction factors, as they were extracted for the different muon track finders. Although the features of distributions are very complicated, owing to the many and intricate factors that play a role in determining them, the general conclusion confirms the principle of increased L1 muon p_{T}. For some very low p_{T} cases, the correction factors are below unity, in contrast with the logic outlined above. However, this usually happens for p_{T} values so low that the muons barely reach the muon subdetectors, hence the probability of a mismeasurement is much more significant than the resolution effects.

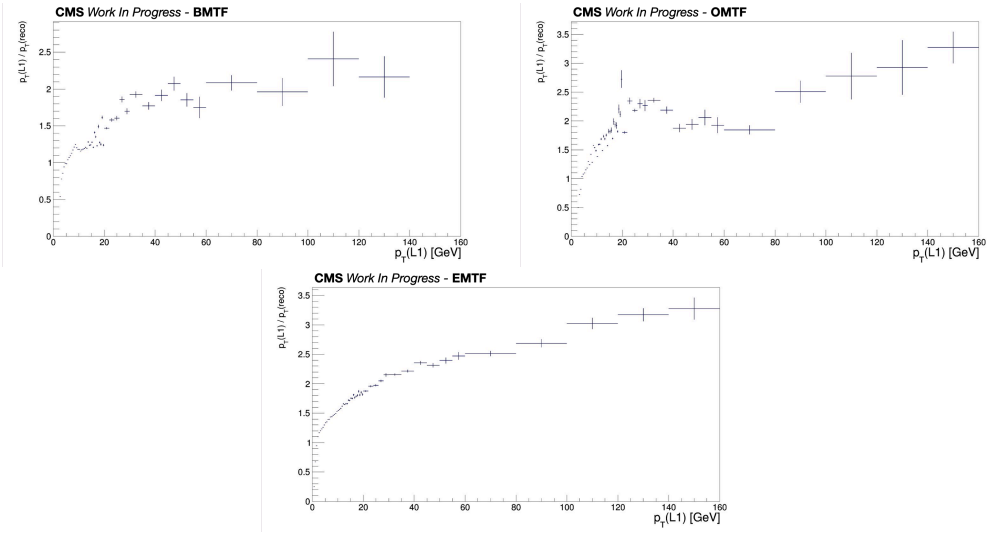


Figure 5.13: Distribution of the average p_{T}(L1)/p_{T}(reco) as a function of p_{T}(L1), measured in the ZeroBias sample of 2017 for the BMTF (upper left), the OMTF (upper right) and the EMTF (lower).

The correction factors, defined as $\alpha = \left(\frac{p_{\mathrm{T}}(\mathrm{L1})}{p_{\mathrm{T}}(\mathrm{reco})}\right)$ are applied as a function of $p_{\mathrm{T}}(\mathrm{L1})$ on the L1 muons following the formula:

$$p_{\mathrm{T}}(\mathrm{L1})^{\mathrm{corr}} = \frac{p_{\mathrm{T}}(\mathrm{L1})}{\alpha} \tag{5.1}$$

As mentioned above, the correction is applied on the ZeroBias sample of 2018 and the Charmonium sample of 2017 to cross check its performance. Fig. 5.14 shows the resolution of the L1 muon p_{T} before the corrections (red line) and its improvement after the corrections (blue line). The results show not only the removal of the bias in the mean value, now centred very close to zero, but also an improvement of the width of the distribution. Table 5.3 summarizes the changes in the mean and the width of the resolution.

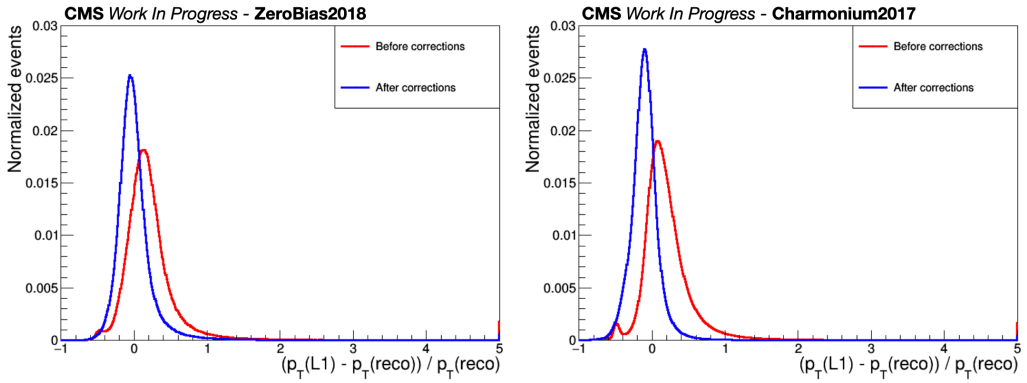


Figure 5.14: Resolution of the L1 p_{T}, defined as $\frac{p_{\mathrm{T}}(\mathrm{L1})-p_{\mathrm{T}}(\mathrm{reco})}{p_{\mathrm{T}}(\mathrm{reco})}$, applied in the ZeroBias sample of 2018 (left) and the Charmonium sample of 2017 (right). The red line shows the resolutions before any correction, while the blue line shows the resolution after the p_{T} scale correction.

	ZeroBias2018		Charmonium2017	
	Mean	Width	Mean	Width
Before corrections	0.219	0.436	0.214	0.425
After p_{T} correction	0.012	0.303	-0.094	0.245

Table 5.3: Summary of the resolution of the L1 p_{T} muons in the two testing samples before and after the application of the p_{T} scale corrections

A small bump can be noticed at $\frac{p_{\mathrm{T}}(\mathrm{L1})-p_{\mathrm{T}}(\mathrm{reco})}{p_{\mathrm{T}}(\mathrm{reco})} \approx -0.5$, especially in the Charmonium sample. Studies on the origin of this bump show that it is formed by muons that have $p_{\mathrm{T}}(\mathrm{L1}) \approx 3\,\mathrm{GeV}$ and $p_{\mathrm{T}}(\mathrm{L1}) \approx 5\,\mathrm{GeV}$. In general, muons of low p_{T} tend to lose a significant portion of their energy before reaching the muon chambers, for reasons like multiple scattering in the detector material. At the same time, due to the L1 p_{T} binning, the p_{T} values measured as 3 GeV or below are all assigned to the first L1 p_{T} bin, leading to a saturation at that bin. Since the offline reconstruction and identification performance reaches its maximum at $\sim$5 GeV, the misreconstructed muons of lower p_{T} have a higher probability to be reconstructed with this p_{T} value, leading to the configuration above, hence the bump.

The ϕ coordinate extrapolation is another aspect of the L1 trigger reconstruction that can be improved. The measurement of the ϕ variable at the L1 trigger is performed at the second muon station. However, the muon track is bent due to the CMS magnetic field, primarily in the ϕ direction. As a result, the measured ϕ value is not the same as the one at the primary vertex, which is the useful physics quantity. Generally, the limited L1 capabilities do not allow for a precise extrapolation of the muon track back to the primary vertex to recover the ϕ variable at the muon production point. Because of this, the scouted L1 data can profit by a correction of the ϕ coordinate. It is worth noting that the inaccurate track extrapolation has minimal effect on the η variable and, hence, no correction is applied for that.

As a consequence of the above, a complication arises to the matching of the reconstructed objects to the L1 objects. The procedure is implemented by matching each reconstructed muon to the closest, in terms of $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}$, L1 muon, where the variables are calculated at the primary vertex. The inaccurate L1 extrapolation back to the primary leads to a nonnegligible number of wrongly matched reconstructed-L1 pairs and, consequently, wrong correction factors. To mitigate this issue, instead of calculating the distance between muons at the primary vertex, the distance was calculated by using the ϕ variable at the second muon station. At that point, no track extrapolation is required at the L1 trigger, while the extrapolation of the reconstructed track can be performed with much better accuracy. In all of the cases where the ϕ variable is mentioned in the context of the scouted data corrections, the ϕ variable as measured at the second muon station is used.

For the correction of the ϕ coordinate of the scouted data, a regression

multivariate algorithm, more specifically a Multilayer Perceptron (MLP), was chosen. An MLP is a artificial neural network in which the "neurons", i.e. the points in the information flow where a mathematical operation takes place, is organized in multiple layers. The first layer is the input layer, where the variables, as available from physics processes, are inserted in the MLP and following layers, called hidden layers, include neurons with, in principle, nonlinear response. The last layer is the output layer, whose results are the target of the MLP. The MLP outputs are compared to the desired values with the aim of minimizing the mean square error. The results are also back-propagated through the network and the neuron responses are optimized to achieve the above minimization via this iterative procedure. Fig. 5.15 shows the diagram of an MLP with 4 input variables, one hidden layer with 5 neurons and 1 output variable.

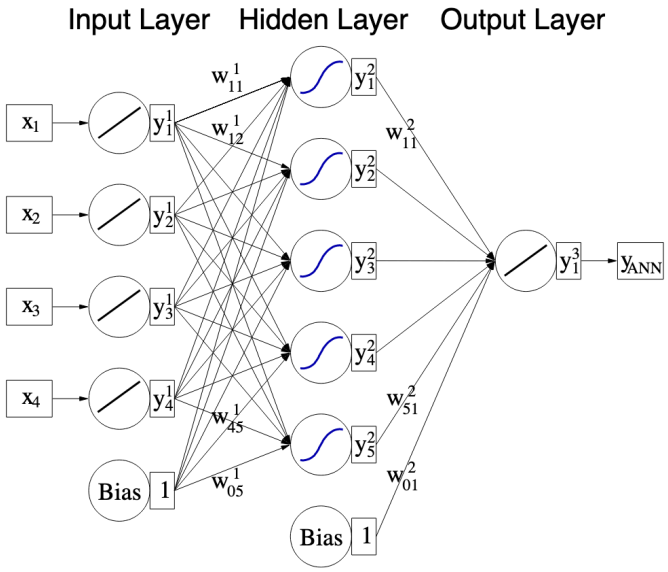


Figure 5.15: Schematic of a simple MLP with one hidden layer. The MLP has 4 input variables and 1 output variable. The connections between neurons are also shown [138].

The MLP training is done on the data of the ZeroBias sample of 2017. The data set is split into the different muon tracker finders and the training is done separately for each one of them. Even though the scouted data contain explicit information on the muon tracker the muon was detected by,

the η variable is exploited for the splitting, as it was giving better results (Fig. 5.16).

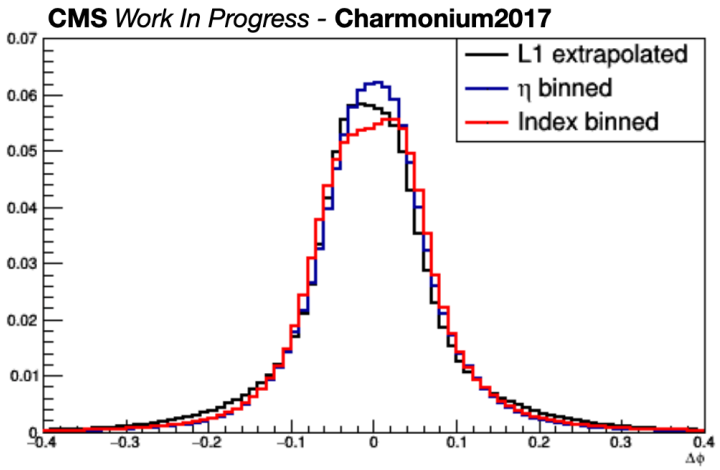


Figure 5.16: Difference between the ϕ coordinate as calculated at the L1 trigger (black line) and by the MLP when it uses the η variable (blue line) and the muon track finder index (red line) for the data set splitting.

Almost every variable available in the scouted data is included as an input variable in the MLP. These are:

- Corrected (with the method described above) p_{T}
- η
- ϕ (calculated at the second muon station)
- Charge

Before the MLP training, the range of all input variables is linearly scaled to $[-1, 1]$. This is done so that direct comparison between all the different variable weights is possible, irrespective of the variable range. At the same weight, the minimization of weights in the constrained range can be more effective. Additional information on the technical implementation of the MLP are given in Appx. C.

The target of the MLP is the correction factor that needs to be added to the ϕ variable, so that it is calibrated to the offline reconstructed value:

$$\Delta\phi^{\mathrm{MLP}} = \phi - \phi^{\mathrm{reco}} \tag{5.2}$$

The input variables, as well as the MLP target, are shown in Fig. 5.17 for the different muon track finders. The correlation of the input variables is also given in the same figure.

According to Eq. 5.2, the corrected ϕ coordinate is given by:

$$\phi^{\mathrm{corr}} = \phi - \Delta\phi^{\mathrm{MLP}} \tag{5.3}$$

The correction derived above are applied to the ZeroBias sample of 2018 and the Charmonium sample of 2017 to evaluate their performance. Fig. 5.18 shows the difference between the ϕ variable at the L1 trigger and the ϕ after offline reconstruction, which is a proxy for the ϕ resolution. The different lines on the plots indicate whether corrections have been applied (blue line) or not (red line). The plots are given for the different muon track finders separately and showcase the distinct behavior in each one of them, which motivates splitting the training data set. Focusing on the red lines, which use the ϕ variable as calculated by the L1 trigger extrapolation, the ϕ resolution deteriorates going from the BMTF to EMTF. This is expected, as the CMS magnetic field gets more and more nonuniform with increasing η, makes the extrapolation of the muon track back to the primary vertex harder. To add to that, the occupancy of the muon chambers is higher in the more forward regions, hence the probability of misreconstructed L1 muons is higher. The blue lines, which use the ϕ variable after the MLP correction, display a much improved resolution, which is of similar shape for all the muon track finder. The improvement is more evident in the ZeroBias sample of 2018 (upper set of plots), which tends to include muons of potentially lower p_{T} and quality, since they are not selected by any algorithm. The resolution enhancement is marginal for the BMTF and OMTF muons of the Charmonium sample of 2017 but it is still substantial for the EMTF muons.

Another test that can be performed to exhibit the improvement brought about by the p_{T} and ϕ calibration is the calculation of the dimuon invariant mass and the resolution of specific dimuon resonances, such as the J/ψ. The left plots of Fig. 5.19 show the dimuon mass spectrum in the ZeroBias and Charmonium samples of 2018 and 2017 respectively. The green line represents the dimuon invariant mass using the offline reconstructed variables. As a result, the different resonances are shown clearly: In the ZeroBias sample, the ω, ρ and ϕ resonances are visible at $M(\ell\ell) \approx 1\,\mathrm{GeV}$ and there is the huge

and very narrow J/ψ peak at $M(\ell\ell) \approx 3.1\,\text{GeV}$. In the Charmonium sample, which includes triggers tuned to select the c and b quark mesons with high efficiency, the J/ψ and its excitations are evident. Even though the B mesons are not separated, due to their low branching fraction to muons, the range in which they are found in the dimuon mass spectrum is also selected. The blue and red lines correspond to the dimuon invariant mass calculated with the L1 trigger variables with and without the corrections derived above, respectively. The resolution of the red line is not exceptionally good, leading to the smearing out of the peaks and bumps present in the reconstructed spectrum. On the other hand, the blue line shows signs of the underlying dimuon mass spectrum, both in the ZeroBias sample, where the resonances seen in the offline distribution are formed as small peaks or bumps, and in the Charmonium sample, where the J/ψ peak and the B meson range become evident. The right plots of Fig. 5.19 show the resolution of the J/ψ peak ($\frac{M(\ell\ell)-M(\text{J}/\psi)}{M(\text{J}/\psi)}$) with the corrected (blue line) and uncorrected (red line) dimuon invariant mass. As expected, both the mean and most importantly the width of the distributions are improved. A small underestimation of the J/ψ mass is seen in the Charmonium sample.

The studies presented above are a first demonstration that the scouted data, with the proper corrections, can prove useful of offline analysis of specific physics cases, which are not affected by the lower resolution of the measured quantities but benefit a lot by huge yields. The corrections described above and additional, such as the calibration of the η variable, are continuously studied and more advanced tools, such as deep NNs, are being utilized. These further improvements are beyond the scope of this thesis.

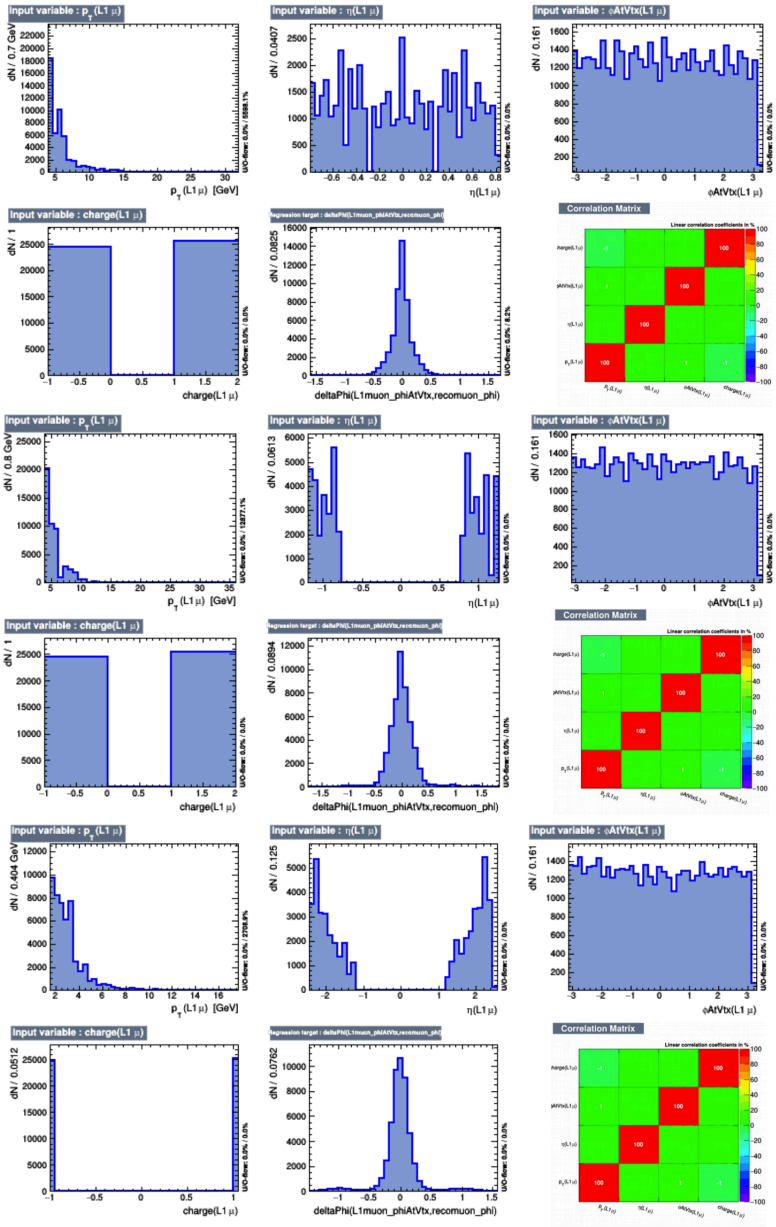


Figure 5.17: The input variables of the MLP (p_T: upper left, η: upper center, ϕ: upper right, charge: lower left), the MLP target (lower center) and the input variable correlation matrix (lower right). The BMTF, the OMTF and the EMTF is shown in the upper, middle and lower set of plots respectively.

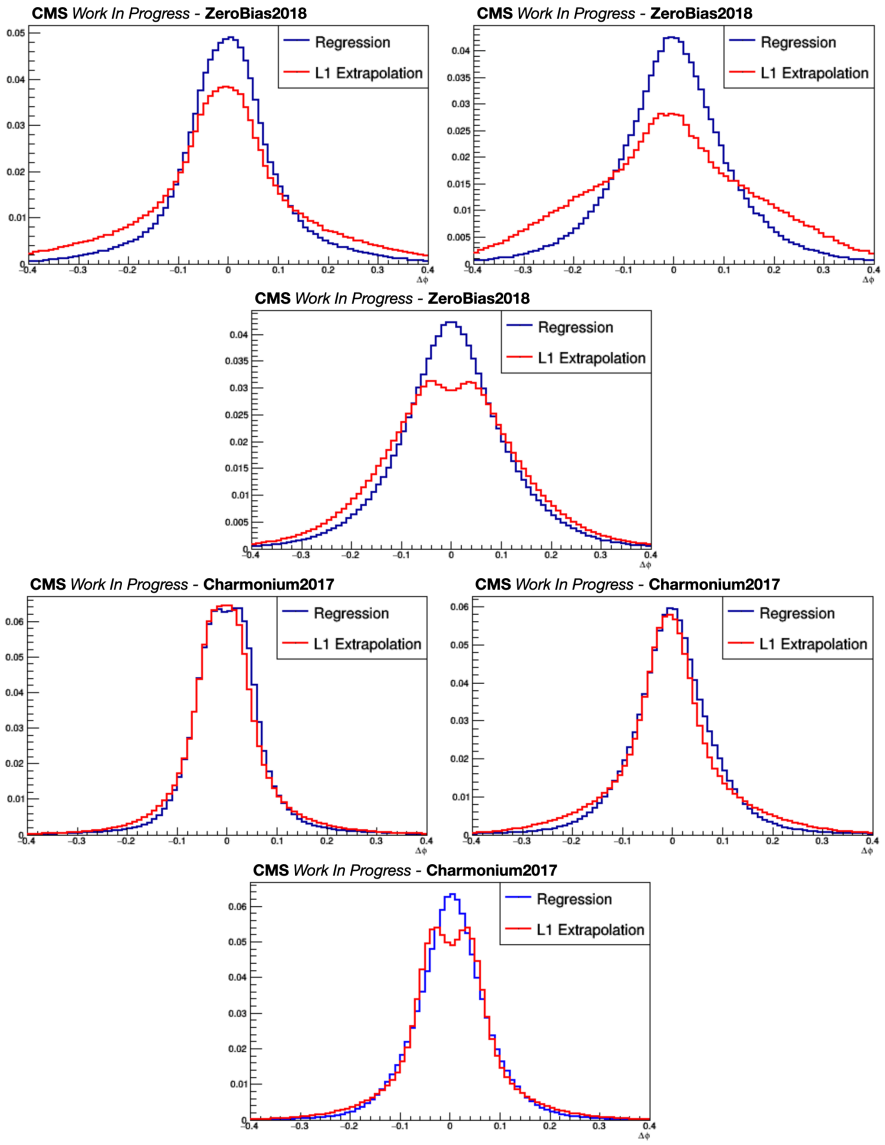


Figure 5.18: Distributions of the difference between the ϕ, as measured at the L1 trigger (red line), or the ϕ^{corr}, as defined in Eq. 5.3 (blue line), and the reconstructed ϕ value. The BMTF, the OMTF and the EMTF is shown in the upper left, upper right and lower plot respectively. The evaluation was performed in the ZeroBias(Charmonium) sample of 2018(2017) in the upper(lower) set of plots.

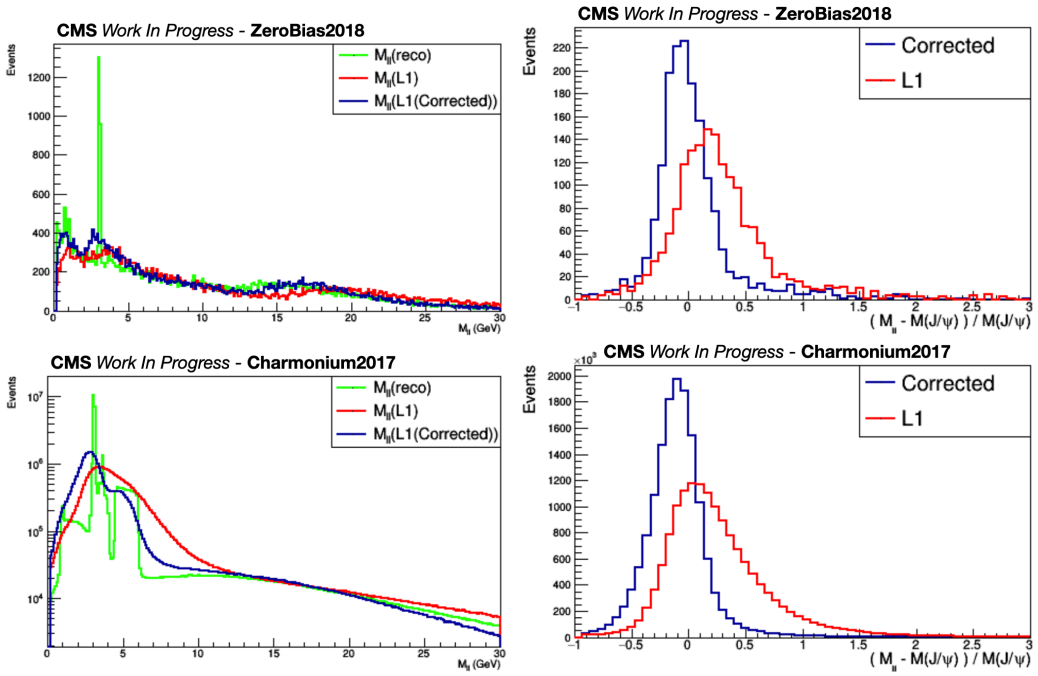


Figure 5.19: Left: The dimuon mass spectrum calculated with the reconstructed variables (green line) and the L1 trigger variables before (red line) and after the corrections mentioned in this section (blue line). Right: The resolution of the J/ψ peak, defined as $\frac{M(\ell\ell)-M(\mathrm{J}/\psi)}{M(\mathrm{J}/\psi)}$, with the same color code as above. The results are shown for the ZeroBias(Charmonium) sample of 2018(2017) in the upper(lower) set of plots.

5.4 The Level-1 Correlator Trigger

The PF algorithm has been applied to the offline and HLT reconstruction of the CMS data since Run 1 and has proven extremely valuable for the enhancement of the resolution of physics objects. Given its success in the higher levels of CMS reconstruction, it follows that the L1 trigger could also benefit from the implementation of such an algorithm. There are multiple Phase-2 upgrades in the CMS L1 trigger system that make this possible: The increased latency of the L1 trigger system and the advances in the capabilities of the hardware boards provide the time and the resources to run more complicated algorithms, such as the PF. Apart from that, the availability of tracks at L1 trigger and the high granularity of the upgraded CMS calorimeters allow for the fine separation of (charged) objects, which is very crucial for the efficiency of the algorithm.

As the running conditions of Phase-2 will be much harsher than those of Phase-1 in terms of PU, there is significant gain from the application of PU mitigation techniques, even at the L1 trigger. With the creation of PF candidates at the L1 trigger, the PUPPI algorithm can be applied, taking them as inputs, in order to reduce the L1 trigger rate by removing a large proportion of the PU early on. Following that, the PUPPI candidates can be used for the production of higher level objects, such as jets or hadronic tau leptons, with more sophisticated algorithms.

The implementation of the algorithms mentioned above is performed in the new Phase-2 L1CorT subsystem. The L1CorT is split in two layers, i.e. distinct groups of processing boards, to achieve an efficient management of the hardware resources and data flow. The L1CorT Layer-1 is responsible for hosting the PF and PUPPI algorithms. Its output is streamed to the L1CorT Layer-2, where the higher-level object algorithms are run. The rest of this section will focus on the specifics of the PF and PUPPI implementation at the L1 trigger as well as the input and output for these algorithms. Secs. 5.4.1–5.4.3 are dedicated to the detailed discussion of particular parts of the implementation, developed in the context of this thesis.

The PF Algorithm at the L1 Trigger The full PF algorithm applied to the higher levels of CMS reconstruction is very complex, making its implementation at the L1 trigger impossible as is. The L1 trigger PF algorithm is streamlined to be run at a short, fixed latency and is optimized for parallel processing of all the input objects. This is possible thanks to the CMS

detector geometry, which can be split in 7 η-regions for the purposes of the L1 trigger system, as shown in the upper plot of Fig. 5.20:

1. The negative endcap with tracker region: $-2.5 < \eta < 1.5$
2. The negative barrel region: $-1.5 < \eta < -0.5$
3. The central barrel region: $-0.5 < \eta < 0.5$
4. The positive barrel region: $0.5 < \eta < 1.5$
5. The positive endcap with tracker region: $1.5 < \eta < 2.5$
6. The endcap without tracker region: $2.5 < |\eta| < 3.0$
7. The forward region: $3.0 < |\eta| < 5.0$

Given its inherent local nature, the PF algorithm can be executed in small $\eta-\phi$ subregions, processing only the objects in their vicinity. The subregions are supplemented by an extra, small overlap region around them to account for input objects that must be linked to an object within the subregion but have ended up marginally out of it. The lower plot of Fig. 5.20 shows some examples of the PF subregions (in solid black) and their corresponding overlap region (in dashed black). The rest of the colored boxes represent the PF *input*-regions that are described below. The PF subregions can be processed in parallel. The choice of their sizes has been made with the aim of truncating less than 5% of the events in a $t\bar{t}$ sample at 200 PU, which is considered as a typical, reference process for studies, while keeping a relatively small number of objects ("input capacity") in each one of them to reduce the algorithmic complexity. Table 5.4 summarizes the number and size of the PF subregions.

Region	Number of subregions $(\eta \times \phi)$	Subregion size $(\eta \times \phi)$
Barrel (Regions 2–4)	$3 \times 9 = 27$	1.5×1.2
Endcap w/ tracker (Regions 1, 5)	$2 \times 9 = 18$	1.5×1.2
Endcap w/o tracker (Region 6)	$2 \times 9 = 18$	1.0×1.2
Forward (Region 7)	$4 \times 9 = 36$	1.5×1.2

Table 5.4: Summary of the number and size of PF subregions. Each subregion size includes additional 0.25 in each side for the overlap region [135].

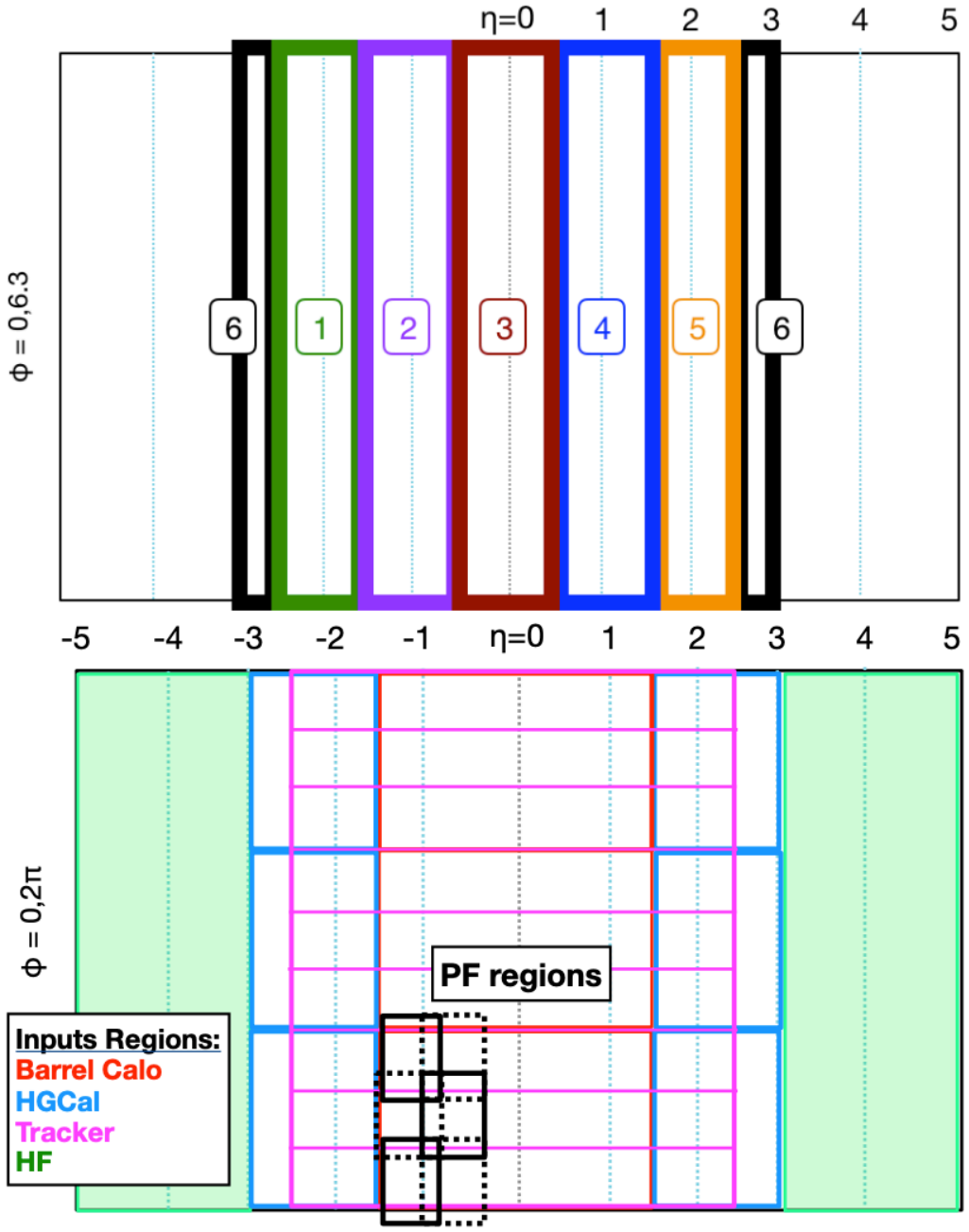


Figure 5.20: Upper: The PF η-regions. Lower: The PF subregions (in black) and PF *input*-regions from the different subdetectors (in color, as indicated in the legend) [135].

Starting with the calorimeter input objects for the PF algorithm, these are handled differently, depending on which subdetector they originate from. In the barrel, input ECAL clusters come already calibrated and are used in the PF algorithm as long as they have $p_{\mathrm{T}} > 0.5\,\mathrm{GeV}$. The energy in the HCAL is combined with the ECAL energy not used in the ECAL clusters to form the input calorimeter towers. Local maxima of tower energy are then used as seeds to create 3×3 tower clusters. These tower clusters are subsequently linked to the ECAL clusters, creating composite ECAL+HCAL clusters. The composite clusters are calibrated as a function of their p_{T}, η and $E_{\mathrm{EM}}/E_{\mathrm{cluster}}$ using $\pi^{\pm}$ mesons and are considered in the PF algorithm if they have $p_{\mathrm{T}} > 1\,\mathrm{GeV}$. The input capacity for the ECAL and HCAL clusters is set to 15 and 19 respectively.

In the endcaps, the input HGCal clusters contain both the electromagnetic and hadronic energy and are calibrated by $\pi^{\pm}$ and π^{0} mesons as a function of p_{T}, η and $E_{\mathrm{EM}}/E_{\mathrm{cluster}}$. The threshold for the inclusion of an HGCal cluster in the PF algorithm is set to 1 GeV. The input capacity for the HGCal is 15 clusters for the $1.5 < |\eta| < 2.5$ subregions and 10 for the $2.5 < |\eta| < 3.0$ subregions. It is worth noting that this low input capacity is only achieved after the application of a machine learning identification algorithm, which is described in detail in Sec. 5.4.1. Finally, the HF clusters are calibrated in a similar way but not as a function of $E_{\mathrm{EM}}/E_{\mathrm{cluster}}$, since there is no electromagnetic and hadronic energy separation in the HF. Due to the high occupancy of the very forward region and the absence of additional information, an acceptable input capacity of 14 clusters for the HF can only be realized by rejecting HF clusters of $p_{\mathrm{T}} < 15\,\mathrm{GeV}$ as inputs to the PF algorithm.

The L1 tracks are accepted as input in the PF algorithm if they fulfill a loose preselection: $p_{\mathrm{T}} > 2\,\mathrm{GeV}$, $|\eta| < 2.5$ and $N_{\mathrm{stubs}} > 4$ with $\chi^2 < 15$. This is looser than the preselection applied for the creation of the tracker-only objects, since the linking of objects from different subdetectors provides an inherent rejection of non-genuine tracks. For similar reasons, only the p_{T} and $|\eta|$ restrictions mentioned above are used for the preselection of the input standalone muons.

The first step in the execution of the PF algorithm is the linking between muons and tracks. Given that the muons are generally easily identified objects, their linking to tracks can be done without ambiguity in most cases. It is important that this is performed at the start of the algorithm, so that, after their association to muons, the L1 tracks used in this step can be dis-

carded from following computation, hence reducing the complexity for the rest of the algorithm. This is exceptionally important in the case of muons within jets, the wrong identification of which could result in the underestimation of the energy of jets. The association of a standalone muon to a track is successful if the track is contained in a η-ϕ square with sides of 0.2. The *max-over-min* p_{T} *ratio*, defined as $\frac{\max(p_{\mathrm{T}}(\mu),p_{\mathrm{T}}(\mathrm{track}))}{\min(p_{\mathrm{T}}(\mu),p_{\mathrm{T}}(\mathrm{track}))}$, also needs to be less than 4. Any ambiguities are resolved by requiring that the max-over-min p_{T} ratio is closer to unity. The PF muon, produced by this linking, acquires the four momentum of the track, which typically has better resolution.

Even though the general logic of the linking of calorimeter clusters to tracks is similar across the whole tracker acceptance, differences exist between the barrel and endcap implementation. In the barrel, all tracks within a ΔR cone of radius 0.04 around an EM cluster are linked to it and their scalar sum of p_{T}, $p_{\mathrm{T}}^{\mathrm{tracks}}$, is computed. After this linking, there are four possibilities for the associated objects:

- If an EM cluster is not linked to a track, then the EM cluster is promoted to a photon object.
- If an EM cluster is linked to tracks and its energy fulfils the relation $p_{\mathrm{T}}^{\mathrm{cluster}} < p_{\mathrm{T}}^{\mathrm{tracks}} - 2\sigma$, then the EM cluster is no longer considered, as it probably arises from a hadronic shower that happened to begin developing in the ECAL.
- If an EM cluster is linked to tracks and its energy fulfils the relation $p_{\mathrm{T}}^{\mathrm{tracks}} - 2\sigma \leq p_{\mathrm{T}}^{\mathrm{cluster}} < p_{\mathrm{T}}^{\mathrm{tracks}} + \sigma$, then the tracks are promoted to an electron object.
- If an EM cluster is linked to tracks and its energy fulfils the relation $p_{\mathrm{T}}^{\mathrm{cluster}} \geq p_{\mathrm{T}}^{\mathrm{tracks}} + \sigma$, then the tracks are promoted to an electron object with $p_{\mathrm{T}}^{\mathrm{e}} = p_{\mathrm{T}}^{\mathrm{tracks}}$ and the EM cluster is promoted to a photon object with $p_{\mathrm{T}}^{\gamma} = p_{\mathrm{T}}^{\mathrm{cluster}} - p_{\mathrm{T}}^{\mathrm{tracks}}$.

where the resolution $\sigma = \max(\sigma^{\mathrm{cluster}}, \sigma^{\mathrm{tracks}})$.

The next step of the calorimeter cluster linking involves the association of EM clusters with hadronic clusters. This is needed, since, as mentioned before, the hadronic clusters include also the EM energy of the ECAL. The EM cluster that is closest to a hadronic cluster and within a ΔR cone of radius 0.1 gets matched to the latter and its energy is subtracted from it. If

the remaining energy is below a configurable threshold, set to 10%, then the hadronic cluster is discarded.

The final step is the linking of the EM energy subtracted hadronic clusters to the remaining tracks. As in the case of the EM clusters, multiple tracks can be associated to a single hadronic cluster, under the conditions that $\Delta R < 0.15$ and $p_{\mathrm{T}}^{\mathrm{calo}} \geq p_{\mathrm{T}}^{\mathrm{track}} - 2\sigma^{\mathrm{calo}}(p_{\mathrm{T}}^{\mathrm{track}})$, where $\sigma^{\mathrm{calo}}(p_{\mathrm{T}}^{\mathrm{track}})$ is the resolution of a $\pi^{\pm}$ with a p_{T} equal to that of the track. The second condition is applied to reject non-genuine tracks or tracks with overestimated momentum. The matching ΔR is larger due to the coarser resolution of the HCAL and the larger Molière radius of hadronic showers. For the same reasons, there is the possibility that the multiple clusters can be matched to a single track. To resolve the ambiguity, the track is associated to the cluster that minimizes the quantity:

$$\frac{\Delta R^2}{0.4^2} + \frac{\max(p_{\mathrm{T}}^{\mathrm{track}} - p_{\mathrm{T}}^{\mathrm{calo}}, 0)}{\sigma^{\mathrm{calo}}(p_{\mathrm{T}}^{\mathrm{track}})}$$

where the max function is used to not penalize the cases where the hadronic cluster is the result of the merging of deposits from multiple particles. After this linking, there are three possibilities for the associated objects:

- If a hadronic cluster is not linked to a track, then the hadronic cluster is promoted to a photon or neutral hadron, depending on whether its energy is primarily electromagnetic or hadronic.

- If a hadronic cluster is linked to tracks and its energy fulfils the relation $p_{\mathrm{T}}^{\mathrm{cluster}} < p_{\mathrm{T}}^{\mathrm{tracks}} + \sigma$, then the tracks are promoted to a charged hadron object.

- If a hadronic cluster is linked to tracks and its energy fulfils the relation $(p_{\mathrm{T}}^{\mathrm{cluster}})^2 \geq (p_{\mathrm{T}}^{\mathrm{tracks}})^2 + \sum \left(\sigma^{\mathrm{calo}}(p_{\mathrm{T}}^{\mathrm{track}})\right)^2$, then the tracks are promoted to a charged hadron object with $p_{\mathrm{T}}^{\mathrm{ch}} = p_{\mathrm{T}}^{\mathrm{tracks}}$ and the hadronic cluster is promoted to a photon or neutral hadron with $p_{\mathrm{T}} = p_{\mathrm{T}}^{\mathrm{cluster}} - p_{\mathrm{T}}^{\mathrm{tracks}}$, depending on whether the excess energy is primarily electromagnetic or hadronic.

After all the linking steps have been completed, there is the possibility that some tracks are left unlinked. Unlinked tracks are used to create charged hadron objects if their p_{T} is less than 10 GeV and fulfils loose quality criteria or if their p_{T} is less than 20 GeV and fulfils tight quality criteria. The

configurable thresholds defined above basically reject as non-genuine tracks with $p_{\mathrm{T}} > 20\,\mathrm{GeV}$ than have no deposit in any subdetector other than the tracker.

In the endcaps, the linking procedure differs in that the track to calorimeter cluster association happens in one step. The matching of multiple tracks to a cluster is successful if a $\Delta R < 0.1$ requirement and the same max-to-min p_{T} ratio requirement as in the barrel are met, under the additional condition that $p_{\mathrm{T}}^{\mathrm{calo}} \geq p_{\mathrm{T}}^{\mathrm{track}} - 2\sigma^{\mathrm{calo}}(p_{\mathrm{T}}^{\mathrm{calo}})$. The promotion of the tracks to charged hadrons or electrons and of the clusters to neutral hadrons or photons is determined by the classification of the HGCal clusters as hadronic or electromagnetic, respectively, by the multivariate algorithm described in Sec 5.4.1. The same conditions as in the barrel hold for the remaining, unlinked tracks. Clusters outside of the tracker acceptance are trivially promoted to either photons or neutral hadrons, based on their classification.

The PUPPI Algorithm at the L1 Trigger The PUPPI algorithm has been successfully used for the rejection of PU in the CMS offline reconstruction. Similarly to the PF algorithm, its local nature makes it an appealing solution for the PU removal in the L1 trigger system. For its application, the PUPPI algorithm requires the knowledge of the PV of each event. Vertex finding algorithms have been implemented in firmware, within the latency constraints of the L1 trigger, making a simplified version of the PUPPI algorithm applicable to the L1CorT.

The goal of the PUPPI algorithm is to downweight or even discard objects, based on the probability that they do not originate from the PV. For charged particles, the presence of tracks makes their discrimination from PU easy: Only charged particles associated with the PV are kept and the rest are discarded. The absence of deposits in the tracker for the neutral particles makes their discrimination from PU harder: A metric that attempts to determine whether a neutral particle comes from the PV is constructed. In the central region, where the vertexing information is available, this metric depends on the p_{T} and the distance of tracks, associated with the PV, around the neutral particle. The logic behind this is that neutral particles coming from the PV will be in the close proximity of higher p_{T} tracks from the PV. Mathematically, the metric is written:

$$\alpha_{\mathrm{C}} = \log \sum_{i \in \mathrm{PV}} \frac{\min\left(p_{\mathrm{T}}^{i}, p_{\mathrm{T}}^{\mathrm{max}}\right)^2}{\max\left(\Delta R, \Delta R^{\mathrm{min}}\right)^2} \tag{5.4}$$

where the sum is taken over all the tracks from the PV within a cone of radius 0.3 around the neutral particle. The usage of tracks, instead of PF candidates, removes small inefficiencies that could be caused due to boundary effects on the edges of the calorimeters. The $p_{\mathrm{T}}^{\mathrm{max}}$ is set to 50 GeV and ΔR^{min} is set to 0.07 and 0.04 in the barrel and the endcap respectively. These thresholds have been implemented to prevent individual contributions to the sum to drive the calculation of the metric. When no tracks are found around a neutral particle, $\alpha_{\mathrm{C}} \equiv 0$.

Outside the tracker acceptance, the vertexing information is not available. As a result, in the forward region the metric α_{C} is replaced by the metric α_{F}, which has the same definition as in 5.4 but with the sum taken over all the PF candidates in the cone and not only those originating from the PV, since this information is not known. In view of the coarser granularity of the HF, the $p_{\mathrm{T}}^{\mathrm{max}}$ and ΔR^{min} parameters are set to 100 GeV and 0.1 respectively.

The metrics α are then combined with information on the particle p_{T} and type to compute the particle PUPPI weight, w. In contrast with the offline PUPPI implementation, where global event quantities are used, the only such parameter in the L1 PUPPI implementation is the number of PU, N_{PU}. The full expression of the particle weight is:

$$\begin{aligned}
w &= \frac{1}{1 + e^{-x_{\mathrm{tot}}}}, \text{ where:} \\
x_{\mathrm{tot}} &= x_{\alpha} + x_{p_{\mathrm{T}}} - x_{\mathrm{PU}} \\
x_{\alpha} &= \min\left(\max\left(c_{\alpha}(\alpha - \alpha_0), -x_{\mathrm{max}}\right), +x_{\mathrm{max}}\right) \\
x_{p_{\mathrm{T}}} &= c_{p_{\mathrm{T}}}(p_{\mathrm{T}} - p_{\mathrm{T0}}) \\
x_{\mathrm{PU}} &= \log\left(\frac{N_{\mathrm{PU}}}{200}\right) + c_0
\end{aligned} \tag{5.5}$$

The p_{T} dependence is exceptionally useful in the case of very energetic photons, which would, otherwise, be downweighted, because of the absence of tracks in their vicinity. The parameters c_{α}, α_0, x_{max}, $c_{p_{\mathrm{T}}}$, p_{T0} and c_0 have been tuned to achieve the best response and resolution of jets and $p_{\mathrm{T}}^{\mathrm{miss}}$. They have different values for different $|\eta|$ ranges and for electromagnetic and hadronic objects, hence there is an indirect dependence of the PUPPI

weight on the particle type. Part of their optimization was performed in the context of the development of the e/γ ID for HGCal clusters, described in Sec. 5.4.1. Finally, the p_{T} of each particle is reweighted to $p'_{\mathrm{T}} = wp_{\mathrm{T}}$. Particles with p'_{T} less than a threshold dependent on $|\eta|$ and particle type are discarded as PU particles. It is worth noting that, in the calculations mentioned above, there is the possibility of selecting multiple candidates for the primary vertex, hence recovering efficiency for particles.

Particle Isolation at the L1 Trigger With the creation of charged and neutral PF particles, the implementation of particle isolation is possible in the L1 trigger. The isolation is defined as:

$$\mathrm{Iso} = \sum_i p_{\mathrm{T}} \tag{5.6}$$

where the sum is over all the PF or PUPPI candidates within a cone of radius 0.4, defining PF or PUPPI isolation respectively.

Two working points (WP) having defined, tuned on a sample of Z bosons decaying to two muons, each one of which has $p_{\mathrm{T}}^{\mathrm{gen}} > 20\,\mathrm{GeV}$ and $|\eta^{\mathrm{gen}}| < 2.5$. The *loose* WP reaches 99% efficiency and the *tight* WP achieves 95%. Further studies are performed on the isolation implementation and WPs definition.

L1CorT Architecture As discussed previously, the L1CorT is the system of the Phase-2 L1 trigger that collects data from all subdetectors with the aim of combining it to create higher level objects, as described in the previous paragraphs. As a result, its architecture is complicated, connecting with multiple systems, each with its own data format, geometry and TMUX period. A diagram of all the connections of the L1CorT to its upstream and downstream systems is shown in Fig. 5.21 and a summary of the different characteristics of each input system is included in Table 5.5. A lot of thought and careful planning has been devoted to the organization of the L1CorT architecture, which has led to a two-layer design: L1CorT Layer-1 is dedicated to the creation of the PF and PUPPI candidates that are later transmitted to L1CorT Layer-2, where physics objects, such as jets or the $p_{\mathrm{T}}^{\mathrm{miss}}$, are reconstructed.

In L1CorT Layer-1, the PF and PUPPI algorithms are run in the PF subregions, shown in black in the lower plot of Fig. 5.20 and summarized in Table 5.4. The choice of these subregions is beneficial not only for the

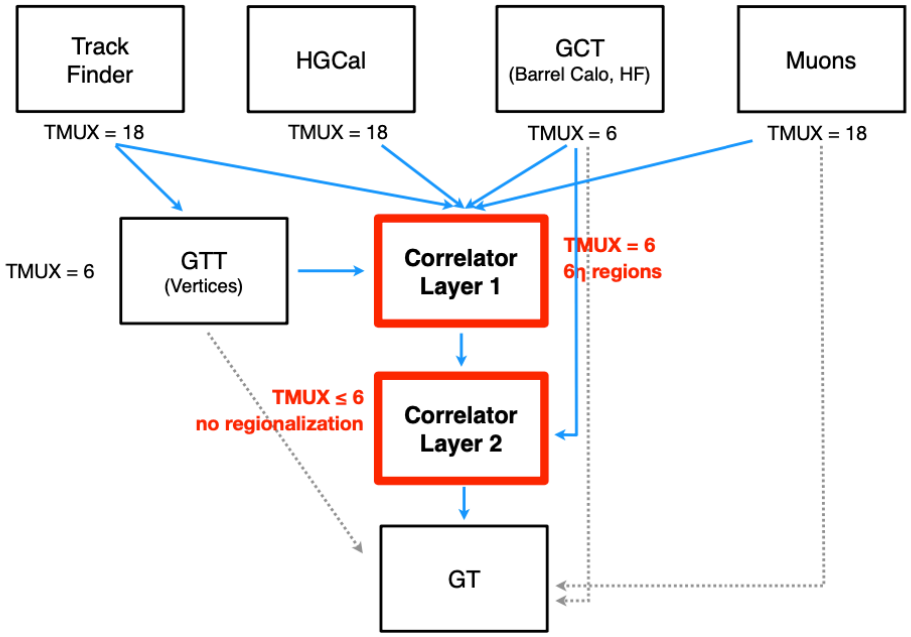


Figure 5.21: Sketch of the L1CorT architecture, showing its two layers in red, its upstream and downstream systems. The I/O data flow of each system is marked in blue or grey (if it is not associated with the L1CorT). The TMUX of the different systems is also indicated on the sketch [135].

System	TMUX (BX)	bits/event	$\eta \times \phi$ regions	Links/FPGA
Muon	18	6k	1×1	1
Tracker	18	81k	1×9	2
GTT (Vertices)	6	1k	1×1	1
Barrel Calorimeter	6	60k	1×3	6
HGCal	18	120k	2×3	4
HF	6	14k	2×1	1

Table 5.5: Summary of the input data flow in the L1CorT [135].

input capacity and truncation prevention of typical process but also for the hardware resource management. The subregions are not too small, so that the overlap regions would cover them entirely, rendering the regionalization pointless, and not too large, so that resources are wasted in attempting to associate objects too far away. It is also worth noting that, since there is no track information in the HF, the PF and PUPPI algorithms are trivial in

that region (Region 7) and they are performed in the L1CT to reduce the workload on the L1CorT. It is worth noting that the HF η-region is included in the overview of PF η-regions in the upper plot of Fig. 5.20, while it comes separately from Global Correlator Trigger (GCT) in Fig. 5.21 to complement the 6 other η-regions coming from L1CorT Layer-1.

To facilitate the firmware implementation and for contingency reasons, the target usage for FPGA resources of each board is set to $\leq 50\%$. Considering the high bandwidth and processing needs of the L1CorT algorithms, the best in terms of performance FPGAs, VU9P, are chosen. Regarding the data organization, a special firmware is designed, called *regionizer*, that receives the data from the different input systems and organizes the input objects into the proper PF subregions. Moreover, it accounts for the different input TMUX factors, transforming them in a TMUX factor of 6, which was chosen for the L1CorT Layer-1 as a balance between data organization and latency. With this setup, i.e. 6 η-regions (excluding Region 7, the HF) and 6 BX as the TMUX factor, L1CorT requires in total 36 boards.

In terms of the I/O links to and from L1CorT Layer-1, these are well accommodated within a VU9P FPGA. The input links from different upstream systems range from 49 (Regions 2, 4) to 76 (Regions 3, 6). On the output side, the links from the L1CorT Layer-1 to L1CorT Layer-2 are 3 per FPGA. A VU9P FPGA has 96 links available.

The L1CorT Layer-2 architecture is designed based on different principles. The L1CorT Layer-2 needs to access all the detector information, so that the creation of large-scale or global objects is possible. As a result, in contrast with the L1CorT Layer-1, Layer-2 includes a firmware module, called *deregionizer*, that collects objects from different PF subregions into a single buffer. Details on the function and implementation of the deregionizer will be given in Sec. 5.4.2. The Layer-2 TMUX factor is 6 BX, in sync with Layer-1. However, in Layer-2, the processing boards are not split over different PF η-regions but by algorithm. Five algorithms are envisaged: Jets, energy sums, τ leptons, electrons & photons (with and without isolation) and PF muons (with and without isolation). Given this, 30 boards are needed for the implementation of L1CorT Layer-2, with a VU9P FPGA each.

Apart from the 18 links from L1CorT Layer-1 (3 links $\times$ 6 different PF regions), each Layer-2 board receives 3 more links to get the PF and PUPPI HF objects from the L1CT. One 25 Gbps link from each Layer-2 board connects to the L1GT, transferring 27 objects. A copy of those objects is also sent to the sGS.

5.4.1 Development of L1 PF e/γ ID for HGCal Clusters Using Boosted Decision Trees

As discovered in early studies of PF objects, the implementation of the PF algorithm at the L1 trigger led to the creation of a large amount of electromagnetic objects. Fig. 5.22 shows the η distribution of PF electrons (left), PF photons (center) and PF neutral hadrons in a $t\bar{t}$ sample at PU 200. Even though an increase in yields is expected with increasing η, the excess of e/γ PF objects with respect to the neutral hadrons is significant. The reason for this is that the L1 PF implementation gives priority to the reconstruction of PF electrons and photons over charged and neutral hadronic PF objects (cf. the PF algorithm steps in Sec. 5.4). However, this observation alone does not imply a flaw in the implementation of the algorithm. On the contrary, it points to the availability of too many input calorimeter clusters, possibly from PU or wrongly labeled as "mostly-electromagnetic", that fall to the default case of unlinked clusters in the PF algorithm.

For the same sample, in the endcap region, where the issue was detected, Table 5.6 summarizes the number of input objects that would be required to avoid truncations in 95% of the events and the respective output PF and PUPPI objects. It is clear that the large excess of HGCal clusters (labeled as "calo") with respect to the tracks available in the same region leads to the creation of a lot of PF neutral objects. The significant reduction in the multiplicity of photons and neutral hadrons after the application of the PUPPI algorithm implies that the majority of those come from the PU. Combined with the priority given to the creation of e/γ objects mentioned above, the large multiplicity of input HGCal clusters is translated to a large multiplicity of PF electrons and photons.

To summarize, the excess of e/γ PF objects in the region $1.5 < |\eta| < 3.0$ was tracked down to the presence of a large amount of input HGCal clusters. Given that the selection on these clusters is minimal, based only on the p_{T}, and that the discrimination of EM and hadronic clusters is performed based on a simple $E_{\mathrm{had}}/E_{\mathrm{EM}}$ requirement, the aforementioned issue calls for the development of improved e/γ ID algorithms for the HGCal clusters. The purpose of these algorithms is two-fold: First, they need to discriminate genuine clusters from the hard scattering from those from PU, so as to reduce the number of input HGCal clusters overall. Second, they need to take advantage of the large variety of variables provided by the HGCal subdetector to improve the labelling of clusters as electromagnetic or hadronic. Since

the HGCal clusters are accepted as input to the PF algorithm if they have $p_{\mathrm{T}} > 1\,\mathrm{GeV}$, the ID algorithms are required to achieve sufficient efficiency from very low p_{T}, reaching a very high plateau for higher values.

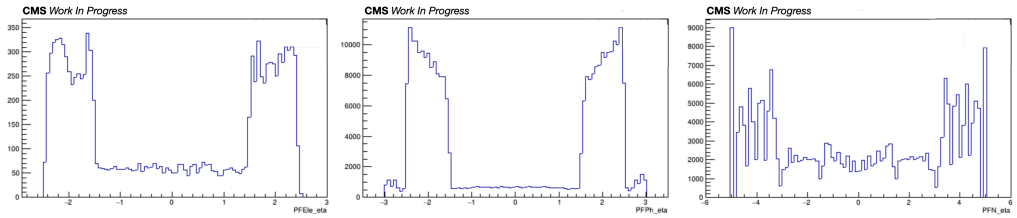


Figure 5.22: Distribution of the η coordinate for PF objects in a $\mathrm{t\bar{t}}$ at PU 200 sample: Electrons (left), photons (center) and neutral hadrons (right).

	Inputs			PF					PUPPI				
	Mu	Tk	Calo	Mu	Ele	Ph	CHad	NHad	Mu	Ele	Ph	CHad	NHad
N_{obj}	2	19	69	1	7	25	8	16	1	5	5	6	2

Table 5.6: Multiplicity of input objects as well as PF and PUPPI output objects in a $\mathrm{t\bar{t}}$ at PU 200 sample. The HGCal cluster multiplicity (Calo) significantly exceeds the input capacity set to 25 objects.

Taking the above requirements into consideration, two HGCal ID algorithms are developed. The first algorithm discriminates the genuine, hard-process e/γ and hadronic clusters versus the clusters originating from the PU. This is called *PU ID*. The second algorithm is responsible for tagging the genuine clusters, which have passed the PU ID, as EM or hadronic, a label that determines whether these will be promoted to EM or hadronic objects in the PF workflow. This is called *Pion ID*. It is worth noting that, based on the definitions above, the PU ID effectively discards input HGCal clusters, while the Pion ID dictates their nature and can only cause a change in the type of PF candidate created.

In view of the upgraded capabilities of the L1 trigger to run multivariate (MVA) algorithms, several linear and non-linear MVA methods were tried for the HGCal PU and Pion ID to achieve the best performance while also considering the hardware resource usage. More specifically, using the TMVA framework, a linear discriminant (LD), a BDT and an MLP method were implemented. Using the same input variables and the best parameters for each method, their performance for the Pion ID is compared in the left plot

of Fig. 5.23. The LD is clearly inferior with respect to the BDT and MLP methods. However, the MLP method is only comparable to the BDT one when the input variables are preprocessed to be normalized in the range $[-1, 1]$ (right plot of Fig. 5.23). However, such an operation takes extra, nonnegligible resources in firmware. As a result, the development of two separate BDT algorithms was chosen. Similar results were obtained for the training of the PU ID.

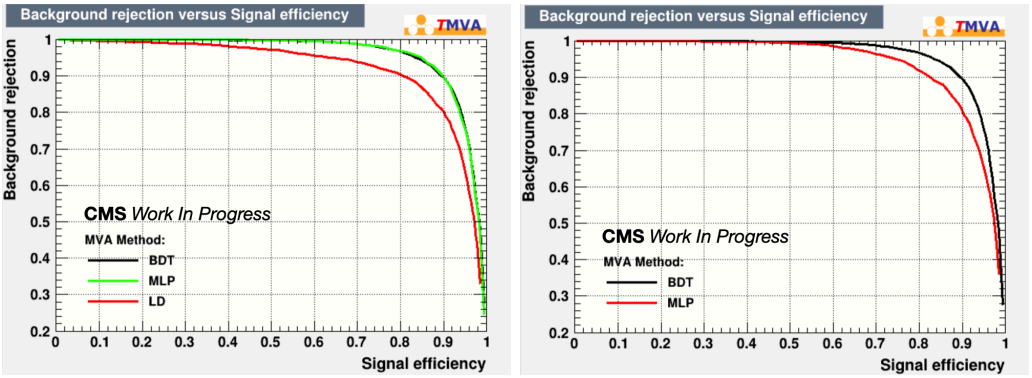


Figure 5.23: Comparison of the performance of different MVA algorithms for the Pion ID. The comparison was performed with the same set of input variables and optimized parameters for each algorithm. The final performance of the ID changed due to differences in the reconstruction HGCal clusters. Left: LD vs BDT vs MLP (normalized input variables). Right: BDT vs MLP (non-normalized input variables).

The training and evaluation of the BDT performance is carried out on three different samples, based on the object that was selected in each case:

- Photons: Matched to MC generator objects in a `DoublePhoton_FlatPt-1To100` sample at PU 200
- Electrons: Matched to MC generator objects in a `DoubleElectron_FlatPt-1To100` sample at PU 200
- Charged pions: Matched to MC generator objects in a `SinglePion_PT0to200` sample at PU 200
- PU: Not matched to MC generator objects in a `DoublePhoton_FlatPt-1To100` at PU 200

The selection of the objects is made based on a matching procedure of the MC generator objects (gen) to the HGCal clusters. To account for the curvature of the trajectory of charged objects, the matching is performed on the surface of the HGCal subdetector, where the MC objects are propagated. Each gen particle is matched to the closest HGCal cluster (reco) as long as the latter is inside a cone of radius $\Delta R = 0.1$ and it fulfils relation $p_{\mathrm{T}}^{\mathrm{reco}}/p_{\mathrm{T}}^{\mathrm{gen}} > 0.3$. The reco clusters used in the training are preselected to have $5 < p_{\mathrm{T}} < 60\,\mathrm{GeV}$ and $1.55 < |\eta| < 2.85$, to avoid bias effects caused by the different response of the different objects, which are more important at boundary values of the sample p_{T} spectrum, and by inefficiencies at the edge of the HGCal subdetector ($1.48 < |\eta| < 3.0$), respectively.

Out of each of the above samples, 10% of their events is reserved for the validation of the BDT performance and the rest is used for the training and testing. Out of the latter 90% of events, 10% is kept for testing and the rest is used for training. The training procedure is performed as follows:

1. Train with the full list of variables (24 in total). The definition of each variable is included in Appx. D.
2. Find the lowest performing variable for the BDT and remove it.
3. Train with N-1 variables and check the performance:
   - If it degrades significantly, keep the variable.
   - Otherwise, drop the variable.
4. Repeat steps 2 and 3 until the minimum set of variables is left.
5. Check the importance/correlation among the remaining variables.
6. Decide the WP based on MET trigger rate curves: First for PU ID, then for Pion ID.

In step 6, the training of the PU ID is prioritized, since this ID is responsible for accepting or rejecting input clusters, therefore it is expected to have more significant impact on the figure of merit, which is the performance of the $p_{\mathrm{T}}^{\mathrm{miss}}$ L1 seed. For step 2, there are two variables that can be defined to evaluate the best performing variables for the BDT training. The first variable is called *separation* and is a generic measure of the dissimilarity of two distribution. It is defined as:

$$\left\langle S^2 \right\rangle = \frac{1}{2} \int \frac{\left(y_{\mathrm{S}}(x) - y_{\mathrm{B}}(x)\right)^2}{y_{\mathrm{S}}(x) + y_{\mathrm{B}}(x)} dx \tag{5.7}$$

where y is the distribution of the variable x and "S" and "B" indicate the signal and background distributions respectively.

Higher separation for a variable implies that this variable is more attractive for usage in a BDT. Nevertheless, separation does not take into account the correlations between the different variables that a BDT, as a non-linear method, can take advantage of. As a result, a second variable, specific for BDTs, called *importance*, is defined as the percentage that a signal variable has contributed to the discrimination between signal and background during the BDT training. The separation of the input variables will be quoted at the beginning of each training as a general estimate of their usefulness. The final decision in step 2 is taken based on the importance, which, however, is not quoted explicitly, since it changes after each training iteration. The importance of the final variables will be given at the end of the training of each ID.

Starting off with PU ID BDT training, the signal is defined as the combination of photons and charged pions, weighted so that their contribution in the signal sample is approximately equal. The background is defined as all the cluster originating from PU. Given the huge amount of such clusters, only a subset of those is used. After the preselection, ∼75000 signal and ∼135000 background clusters are inputted in the PU ID BDT. The entire set of input variables, both for the signal (blue) and the background (red), is given in Figs. 5.24 and 5.25. The linear correlation matrix of these variables is given in Fig. 5.26 for the signal (upper) and the background (lower), while their separation is quoted in the left part of Table 5.7.

Given that the PU clusters are generally soft, while the signal clusters have, by construction, p_{T} up to 60 GeV, it is expected that the p_{T} variable has great separation. Including it in the training would lead to p_{T}-driven results, effectively discarding low-p_{T} clusters. However, a central goal of this ID is to maintain adequate efficiency even for low-p_{T} clusters, therefore the p_{T} is not included in the training set. PU clusters with low p_{T} do not usually extend to multiple layers, in contrast with photons and charged pions from the PV. As a result, coreShowerLength also ranks in separation. In general, the energy of PU clusters is more dispersed, both in the longitudinal and transverse directions when compared to clusters from hard scattering, which explains

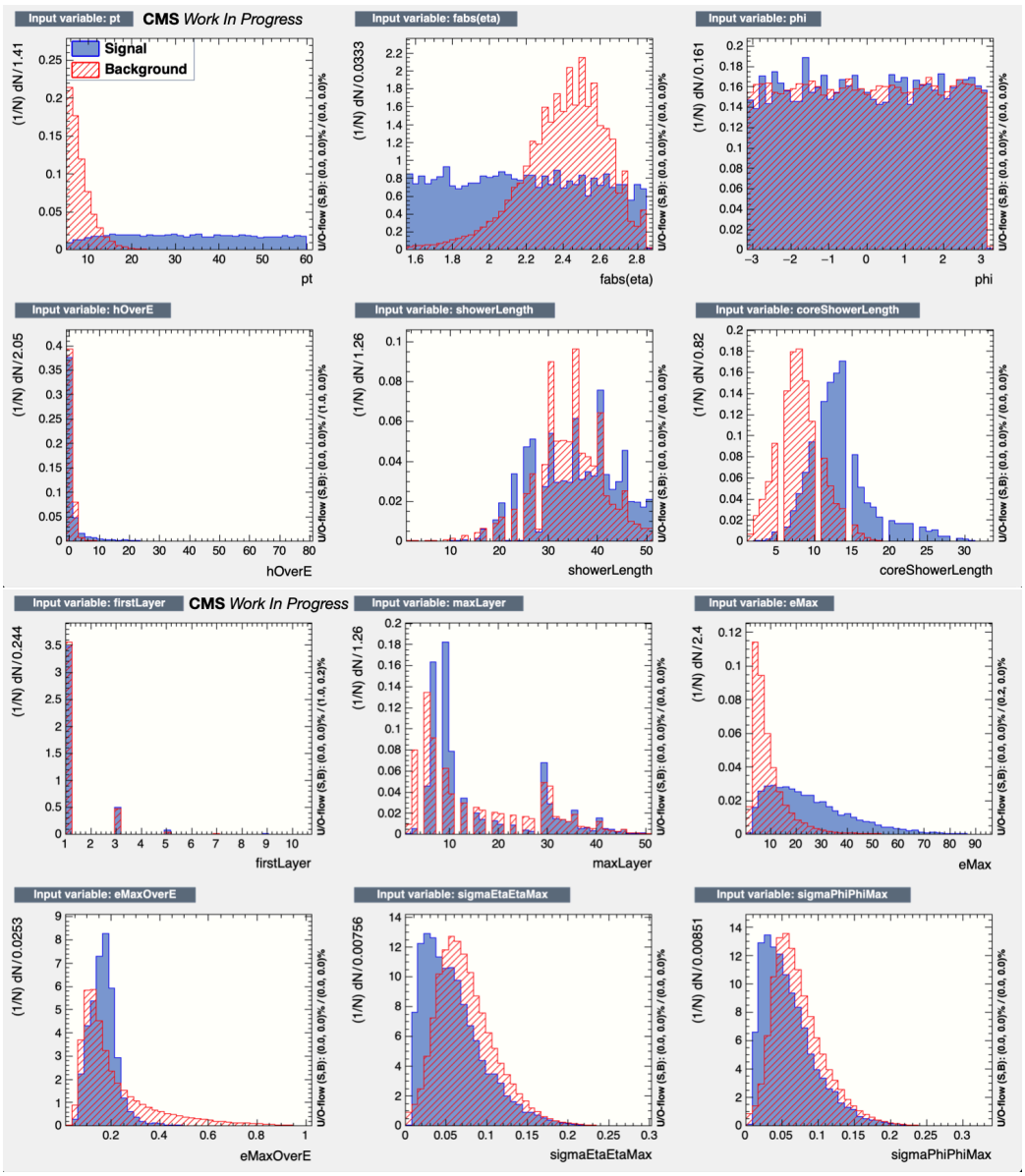


Figure 5.24: PU ID input variables in blue for the signal and in red for the background (part 1).

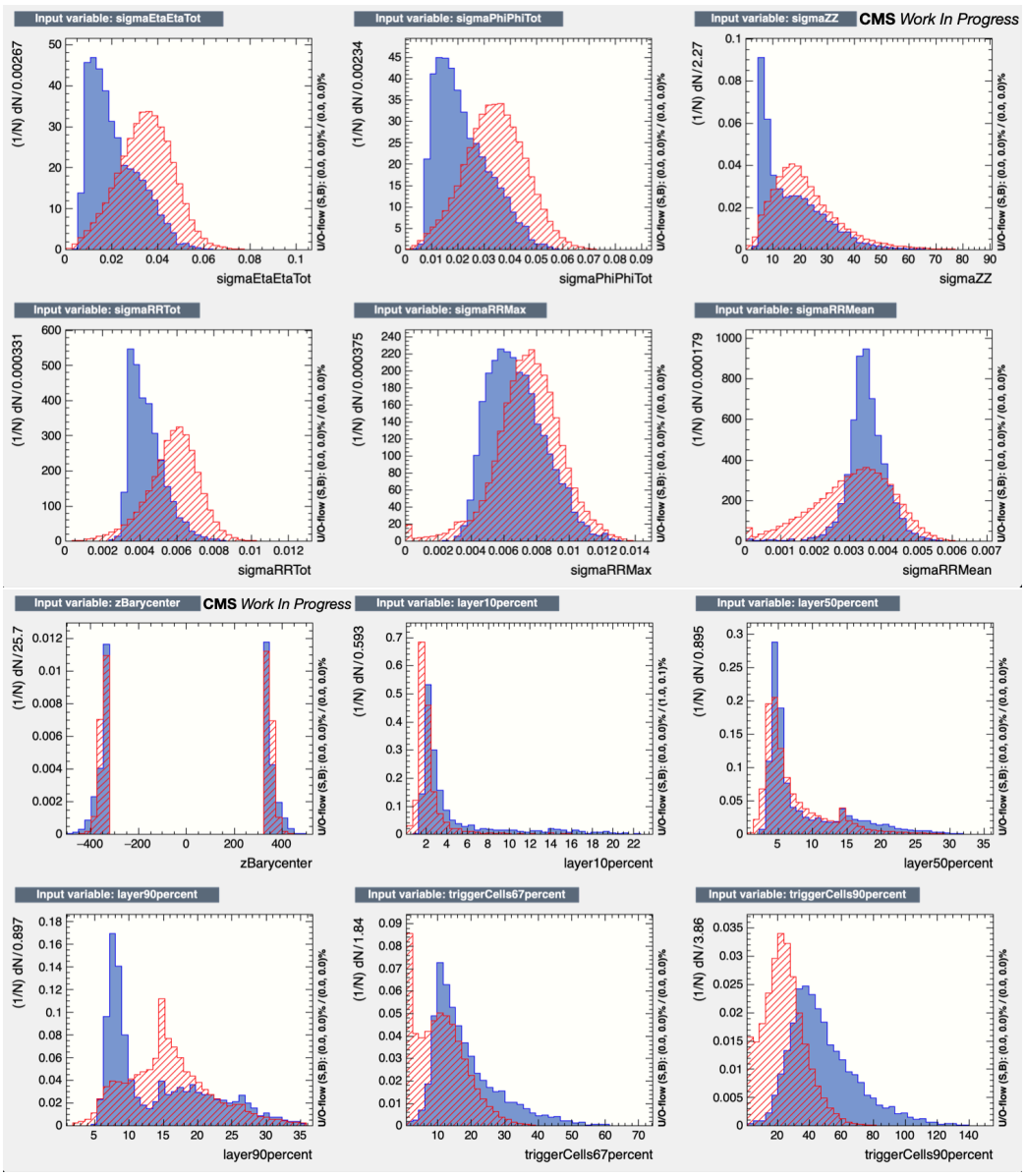


Figure 5.25: PU ID input variables in blue for the signal and in red for the background (part 2).

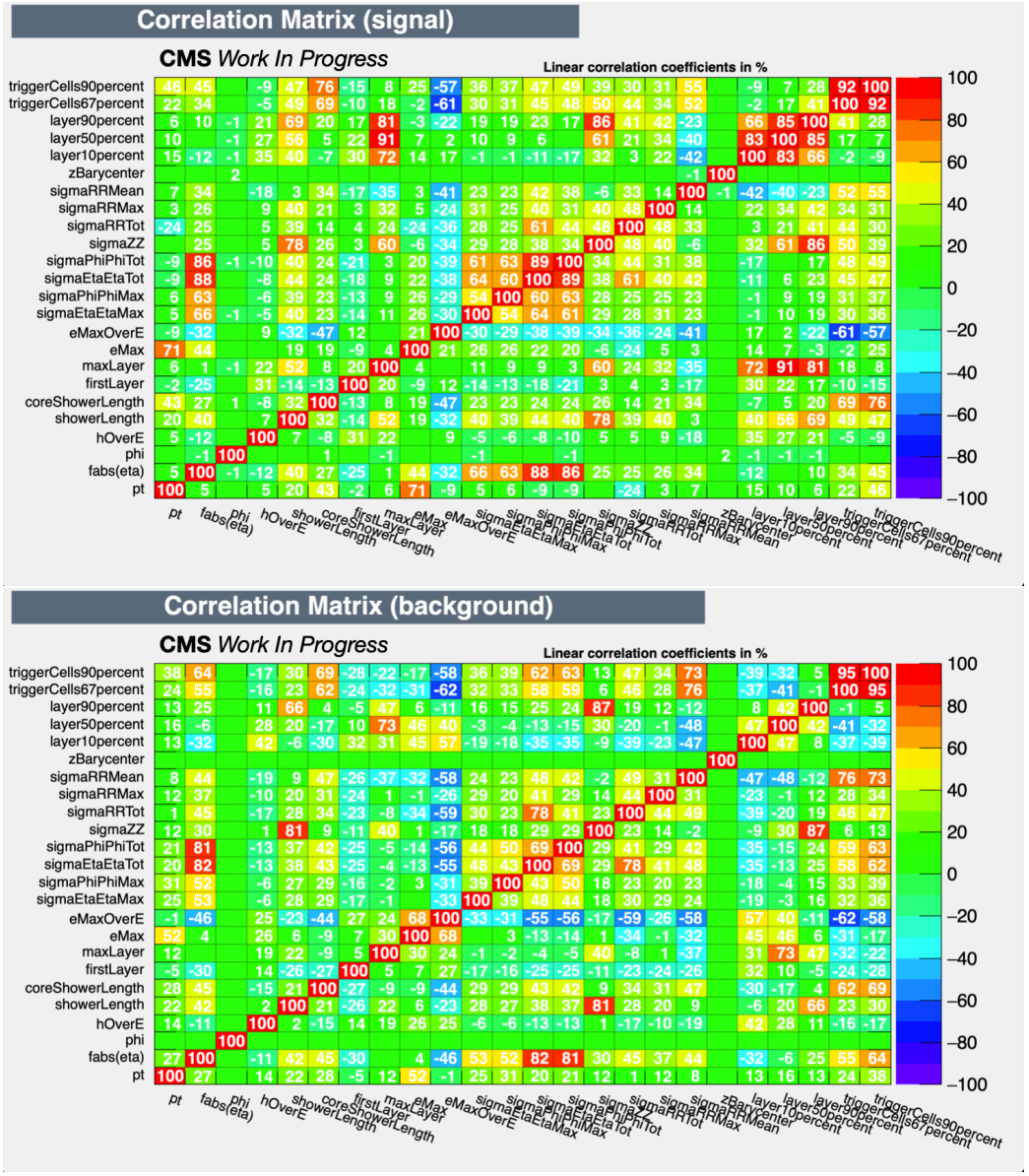


Figure 5.26: Linear correlation matrices of the PU ID input variables for signal (upper) and background (lower).

Rank	Variable	Separation	Rank	Variable	Separation
1	p_{T}	0.715	1	layer90percent	0.699
2	coreShowerLength	0.422	2	layer50percent	0.586
3	triggerCells90percent	0.385	3	sigmaZZ	0.510
4	sigmaRRTot	0.362	4	maxLayer	0.486
5	eMax	0.354	5	zBarycenter	0.478
6	sigmaEtaEtaTot	0.284	6	layer10percent	0.361
7	sigmaPhiPhiTot	0.262	7	showerLength	0.335
8	layer10percent	0.239	8	triggerCells67percent	0.328
9	triggerCells67percent	0.207	9	sigmaRRTot	0.309
10	$\lvert\eta\rvert$	0.196	10	coreShowerLength	0.266
11	sigmaRRMean	0.185	11	eMaxOverE	0.265
12	layer90percent	0.158	12	sigmaRRMax	0.198
13	maxLayer	0.155	13	triggerCells90percent	0.179
14	eMaxOverE	0.149	14	hOverE	0.138
15	sigmaZZ	0.113	15	sigmaRRMean	0.099
16	sigmaEtaEtaMax	0.086	16	sigmaEtaEtaTot	0.083
17	sigmaRRMax	0.086	17	sigmaEtaEtaMax	0.080
18	layer50percent	0.081	18	sigmaPhiPhiMax	0.069
19	sigmaPhiPhiMax	0.072	19	sigmaPhiPhiTot	0.044
20	showerLength	0.061	20	eMax	0.044
21	zBarycenter	0.049	21	firstLayer	0.029
22	hOverE	0.045	22	$\lvert\eta\rvert$	0.008
23	firstLayer	0.003	23	p_{T}	0.008

Table 5.7: Separation of input variables of the PU ID (left) and the Pion ID BDT (right). The ϕ variable is not quoted, since its usefulness is limited.

why the triggerCells$X$percent, the layer$X$percent and the σ variables rank high. This also means that low eMax is expected for PU clusters. Finally, PU clusters tend to populate the more forward regions of the detector, hence the separation in the $|\eta|$ variable.

Following the training procedure outlined above, the best performing variables for the PU ID BDT training are given in Table 5.8:

Rank	Variable	Importance
1	sigmaPhiPhiTot	21.1%
2	eMax	21.1%
3	triggerCells90percent	21.0%
4	sigmaRRTot	19.7%
5	eMaxOverE	17.2%

Table 5.8: Importance of chosen input variables for the final PU ID BDT training.

For the Pion ID BDT training, the signal and background sample definitions need to change: The signal sample consists of clusters from prompt photons, while the background sample is made up of clusters from prompt charged pions. In terms of clusters given as input to the BDT, there remain $\sim$40000 signal and $\sim$35000 background clusters after the preselection requirements are applied. As in the case of the PU ID, all of the Pion ID input variables are given in Figs. 5.27 and 5.28, following the same color code as before. Fig. 5.29 shows the linear correlation matrices of the input variables for the different samples and Table 5.7 (right part) quotes their separation.

In contrast with the PU ID case, the p_{T} and $|\eta|$ are very similar for γ and $\pi^{\pm}$ cluster. However, it is expected that other, more discriminating variables can be a function of $|\eta|$ and, hence, it can be beneficial to include it in the training, so that the BDT identifies correlations with it. From the distributions of input variables, it is evident that the main difference between γ and $\pi^{\pm}$ clusters is their longitudinal profile: layer$X$percent, sigmaZZ, maxLayer, zBaycenter, showerLength and triggerCells$X$percent all point to the observation that clusters from charged pions are longer that those of photons in the z direction. Transverse shape variables also provide a handle for separation. Surprisingly, the hOverE variable is not that discriminating, explaining the only moderate performance of the EM and hadronic tagging of clusters by a simple selection on this variable. Finally, even though the eMax variable is

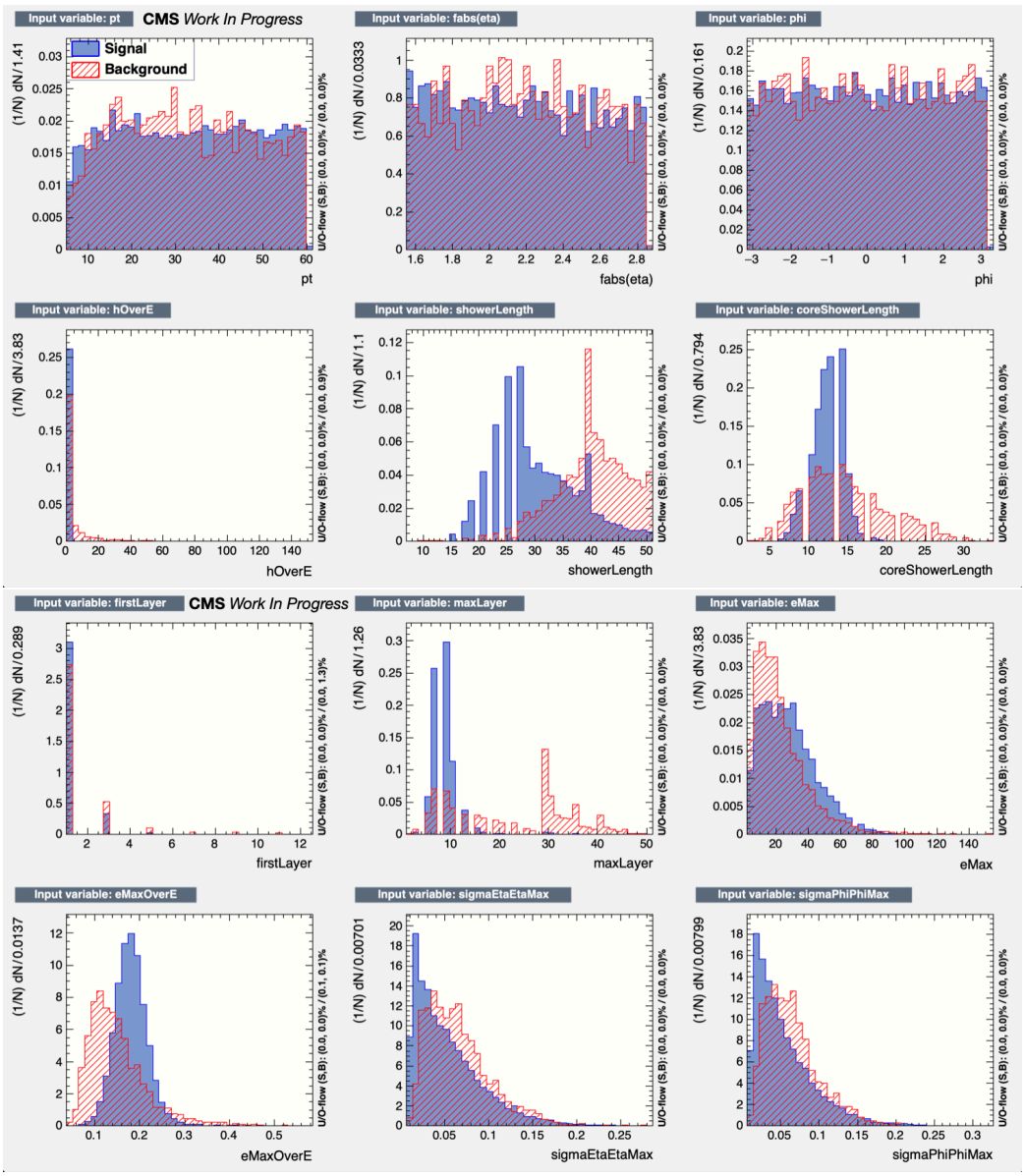


Figure 5.27: Pion ID input variables in blue for the signal and in red for the background (part 1).

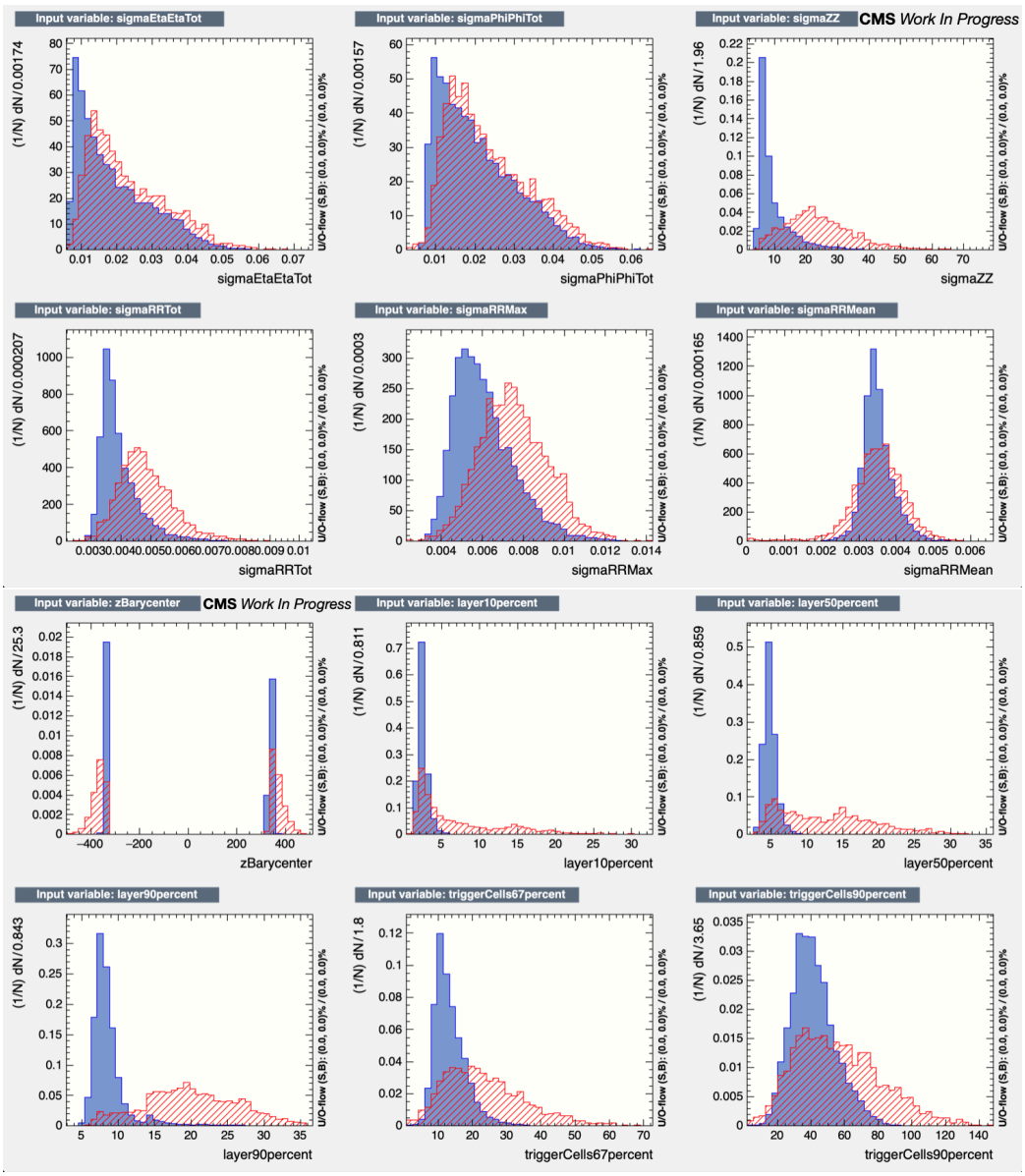


Figure 5.28: Pion ID input variables in blue for the signal and in red for the background (part 2).

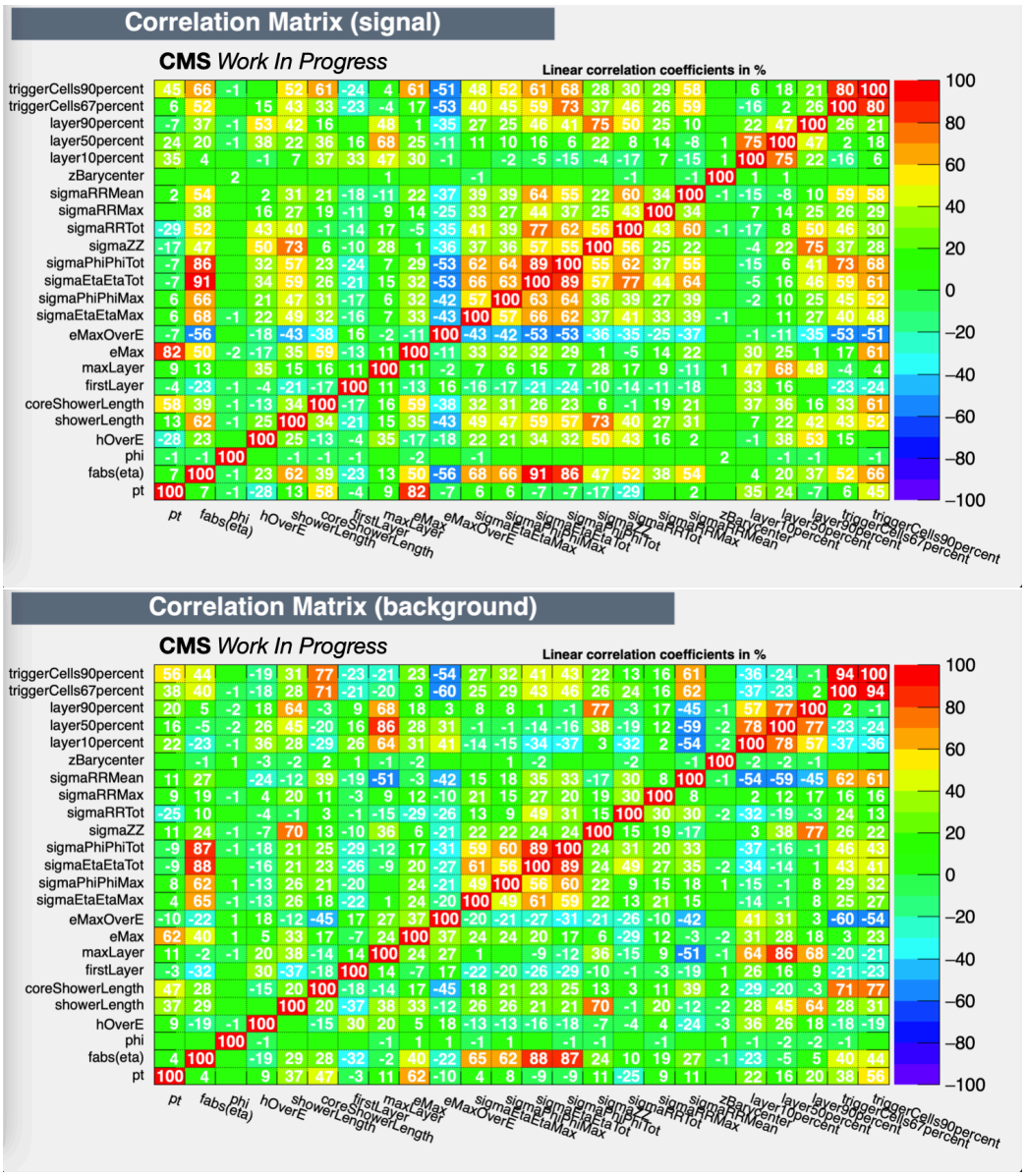


Figure 5.29: Linear correlation matrices of the Pion ID input variables for signal (upper) and background (lower).

ranked very low in terms of separation, the BDT distinguishes correlations with other variables, leading to an increased ranking in terms of importance.

The final set of input variables for the Pion ID BDT training is chosen as quoted in Table 5.9:

Rank	Variable	Importance
1	layer50percent	21.7%
2	triggerCells67percent	17.7%
3	$\|\eta\|$	15.6%
4	sigmaZZ	15.2%
5	sigmaPhiPhiTot	15.1%
6	eMax	14.8%

Table 5.9: Importance of chosen input variables for the final Pion ID BDT training.

The performance of the PU and Pion ID BDTs is evaluated with the use of ROC curves, i.e. the plot of signal efficiency versus background rejection. These ROC curves are shown in Fig. 5.30 for the PU ID on the left and the Pion ID on the right. The plots include two curves: The final BDT training with the chosen variables as specified above for each ID (labeled as "New Training") and an older training with a limited set of variables for the final choice and with preliminary version of the clustering algorithm for the HGCal (labeled as "Old Training") for comparison. The PU ID BDT shows excellent results in terms of signal to background separation, achieving 94% signal efficiency with 99% background rejection, at the optimal ROC curve point in terms of $\mathrm{S}/\sqrt{\mathrm{B}}$. This performance is driven by correlations of the input variable with the p_{T}, identified by the BDT. This leads to the rejection of the majority of low-p_{T} PU clusters, while maintaining sufficient efficiency for the same p_{T} photon and charged pion clusters. The Pion ID BDT also displays exceptionally good performance, with 95% signal efficiency and 95% background rejection at the optimal ROC curve point. For its performance, the driving factor is the longitudinal information, which accounts for more than 50% of the discrimination between signal and background.

One most important targets of the L1CorT is the control of the multi-object L1 seeds. Hence, the optimization of the PU and Pion ID working points is performed by scanning the BDT WP and checking the H_T, quad-jet and $p_{\mathrm{T}}^{\mathrm{miss}}$ L1 seed rate at the 95% efficiency threshold. The optimization

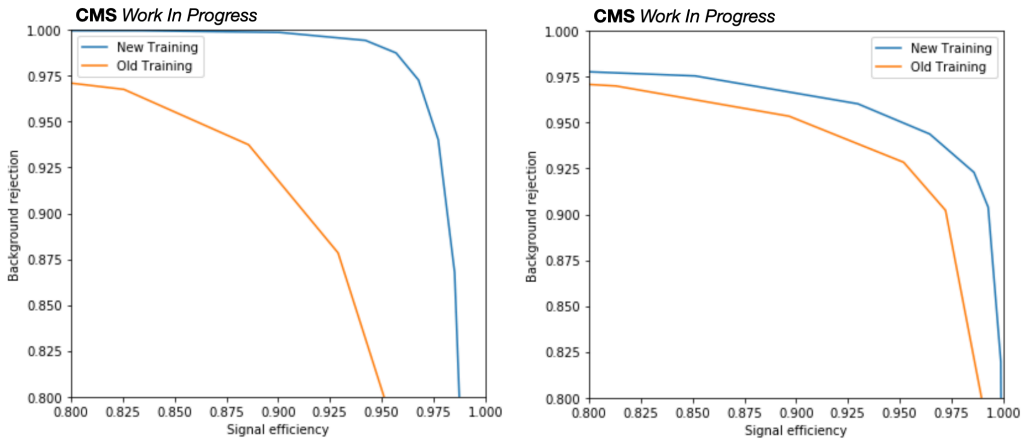


Figure 5.30: ROC curves of the PU ID (left) and the Pion ID BDT (right). The latest training (labeled as "New Training") is marked in blue, while a preliminary training with only subset of the variables and with a previous, non-optimized version of the HGCal clustering is shown in orange for comparison.

is done in a $t\bar{t}$ sample, which includes a lot of jets, moderate $p_{\mathrm{T}}^{\mathrm{miss}}$ and generally central objects, and in a VBF H $\rightarrow$ Inv sample, which includes hadronic activity in the forward region of the detector and a lot of $p_{\mathrm{T}}^{\mathrm{miss}}$. Both of the samples are produced with 200 PU.

Figs. 5.31 and 5.32 show the rate of the H_T, quad-jet and $p_{\mathrm{T}}^{\mathrm{miss}}$ L1 seeds in the $t\bar{t}$ and the VBF H $\rightarrow$ Inv samples respectively. The different lines correspond to the cases without the application of the PU ID, with the application of the Old Training PU ID and with different WP of the New Training PU ID, ranging from -0.3 to 0.3. For the H_T and quad-jet rate, little difference is observed in all of the cases. This is because the PUPPI jet creation algorithm is robust against the presence of PU. On the other hand, $p_{\mathrm{T}}^{\mathrm{miss}}$ is very sensitive to low-p_{T} jets, which can be the result of PU clusters. Without the application of a PU ID, the $p_{\mathrm{T}}^{\mathrm{miss}}$ rate is intolerable for the L1 trigger rate budget. Progressively increasing the WP leads to the reduction of the $p_{\mathrm{T}}^{\mathrm{miss}}$ L1 seed rate and reduction starts to saturate for WP = 0.1. Further increasing the WP does not result in any gain in the $p_{\mathrm{T}}^{\mathrm{miss}}$ performance, while it increases the danger over discarding too many clusters, even genuine ones.

With the choice of 0.1 for the WP of PU ID, the efficiency of photons (Fig. 5.33), electrons (Fig. 5.34), charged pions (Fig. 5.35) and PU (Fig. 5.36)

is calculated as a function of p_{T} (left) and η (right). The efficiency in η is generally uniform, with the efficiency of PU being slightly lower in the low-η HGCal edges. For photons and electrons, the efficiency is very similar, starting at 20% even at 5 GeV and quickly rising to the plateau of 100% at ∼15 GeV. The efficiency of charged pions is a slightly lower over the whole p_{T} range and has a longer turn-on but still reaches 100% at the plateau. For the PU clusters, the efficiency of identifying them and discarding them is ∼100% for $p_{\mathrm{T}} < 10$ GeV, where the bulk of the PU cluster distribution is, and progressively goes down for higher values of the p_{T}.

The strategy for defining the WP for the Pion ID is similar, however its determination is performed after applying the PU ID. All of the curves in Figs. 5.37 ($\mathrm{t\bar{t}}$) and 5.38 (VBF H → Inv) include both the PU ID with the 0.1 WP and different configuration of the Pion ID. The "NoID" curve is an exception, including neither the PU nor the Pion ID. In this case, only small differences are observed in all the L1 seed rates, even in the $p_{\mathrm{T}}^{\mathrm{miss}}$ one, with varying Pion ID WP. As a result, the WP with the best $\mathrm{S}/\sqrt{\mathrm{B}}$ is used, 0.05.

Figs. 5.39–5.41 display the efficiency of photons, electrons and charged pion after the application of the Pion ID with the WP determined above. As in the case of the PU ID, the efficiency is fairly stable all over the η range. The efficiency for all objects is high (> 70%) even at low p_{T} (< 5 GeV). The turn-on is not very steep but it reaches > 95% efficiency at ∼35 GeV.

Following the decision on the WP of both IDs, it is important that the PUPPI parameters in Eq. 5.5 are retuned, so that the PUPPI algorithm does not downweight clusters too much. The prior parameters (c_0), the p_{T} slope parameters ($c_{p_{\mathrm{T}}}$) and the p_{T} priors ($p_{\mathrm{T}0}$), for photons and all the other types of particles, are configured to improve the response and resolution of PUPPI jets, while keeping the efficiency and rate of the H_T and $p_{\mathrm{T}}^{\mathrm{miss}}$ L1 seeds at the same level. After testing the behaviour of each PUPPI parameter separately, three different sets of their combinations were tested, summarized in Table 5.10: The "Boldest" tune relaxes the parameters by a lot, the "Shiest" tune is the most conservative one with respect to the initial values ("No Tuning"), while the "Tuned" one is the middle solution.

Fig. 5.42 shows the efficiency (right) and rate (left) of the H_T (upper) and the $p_{\mathrm{T}}^{\mathrm{miss}}$ L1 seed (lower) in a $\mathrm{t\bar{t}}$ and a VBF H → Inv sample respectively. The threshold for these seeds is much improved compared to the case where no IDs were applied (labeled as "Before" in the legend) but the tuning slightly increases threshold with respect to the aggressive initial tuning. The benefit from the tuning can be seen in Figs. 5.43–5.45, where the re-

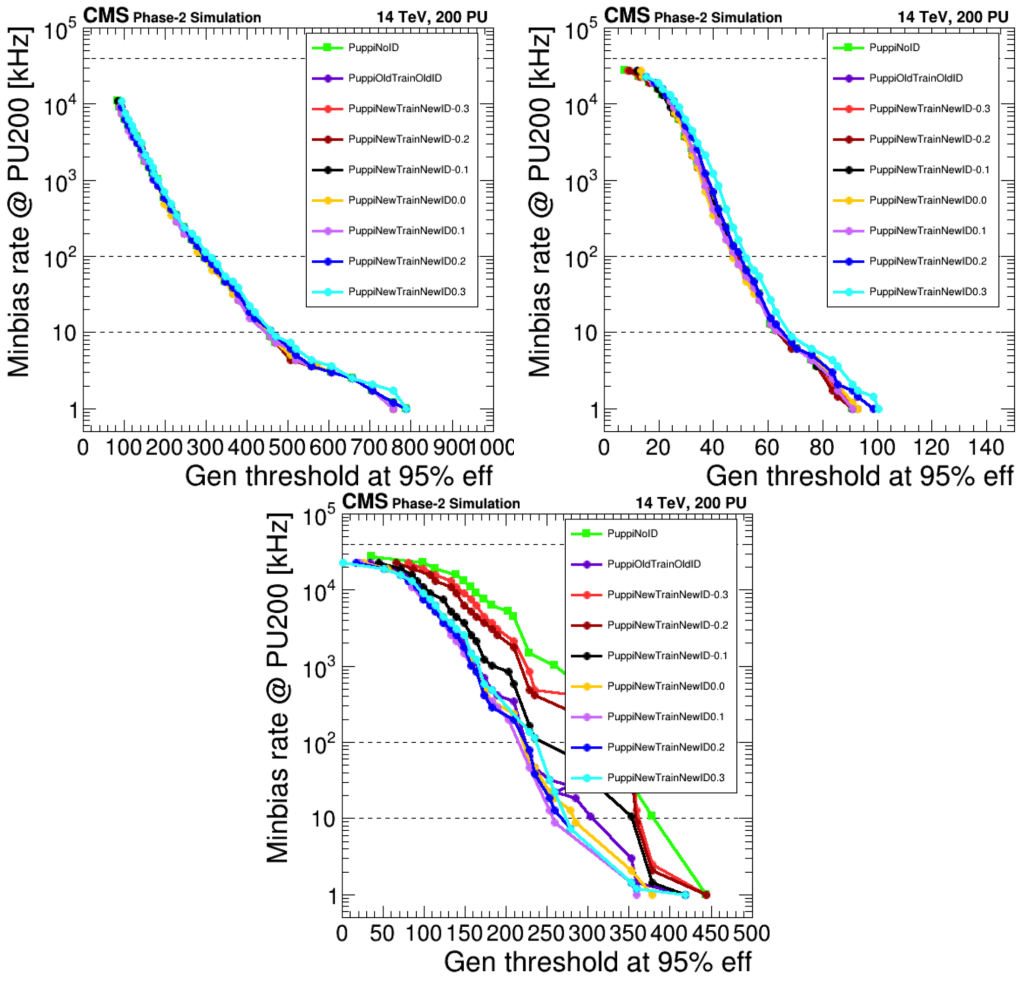


Figure 5.31: Rate of an H_T (upper left), quad-jet (upper right) and $p_{\mathrm{T}}^{\mathrm{miss}}$ (lower) L1 seed versus the corresponding generator variable threshold to achieve 95% efficiency for different WP of the PU ID BDT response in a $\mathrm{t\bar{t}}$ sample at PU 200. The cases where no PU ID and where the preliminary PU ID training are applied are also shown for comparison.

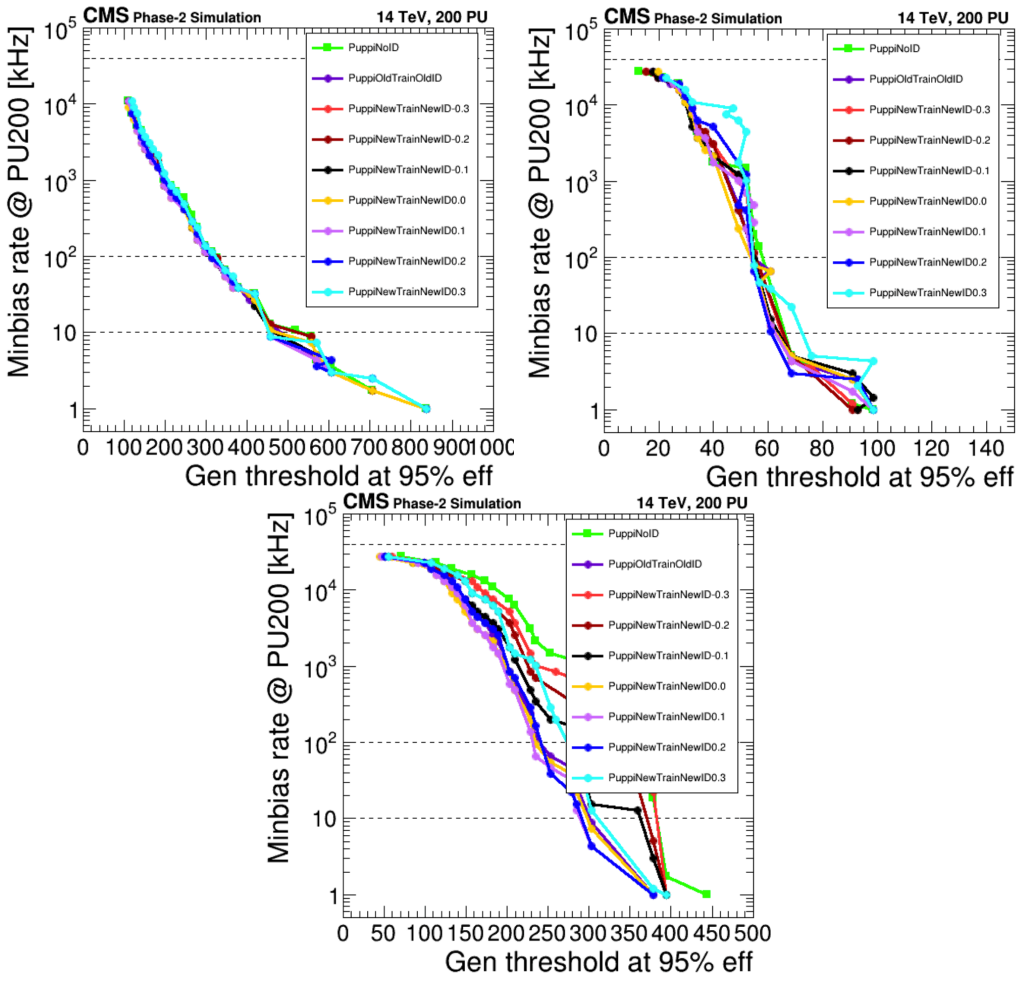


Figure 5.32: Rate of an H_T (upper left), quad-jet (upper right) and $p_{\mathrm{T}}^{\mathrm{miss}}$ (lower) L1 seed versus the corresponding generator variable threshold to achieve 95% efficiency for different WP of the PU ID BDT response in a VBF H $\rightarrow$ Inv sample at PU 200. The cases where no PU ID and where the preliminary PU ID training are applied are also shown for comparison.

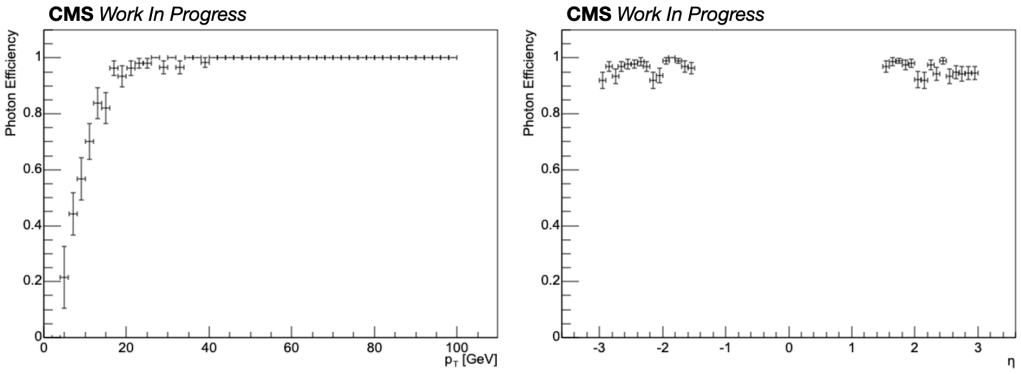


Figure 5.33: Efficiency as a function of p_{T} (left) and η (right) of clusters from genuine photons, having passed the PU ID.

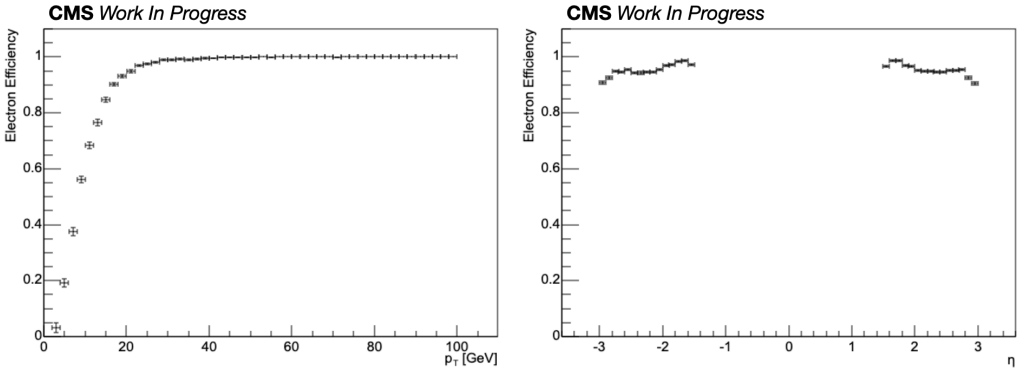


Figure 5.34: Efficiency as a function of p_{T} (left) and η (right) of clusters from genuine electrons, having passed the PU ID.

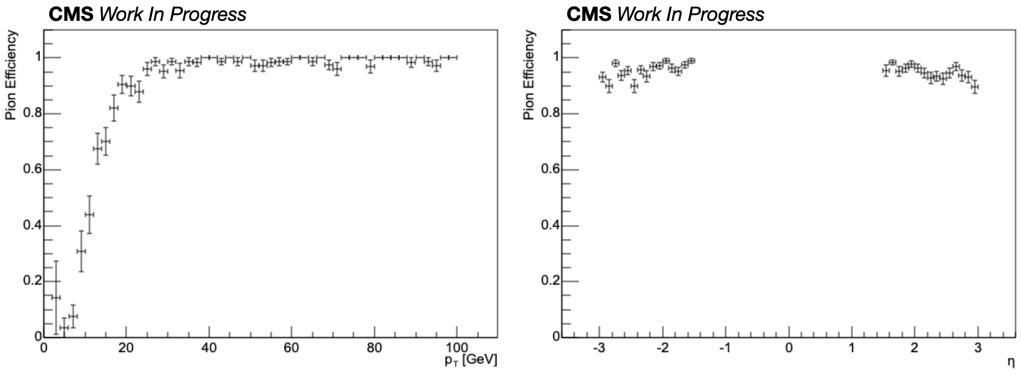


Figure 5.35: Efficiency as a function of p_{T} (left) and η (right) of clusters from genuine pions, having passed the PU ID.

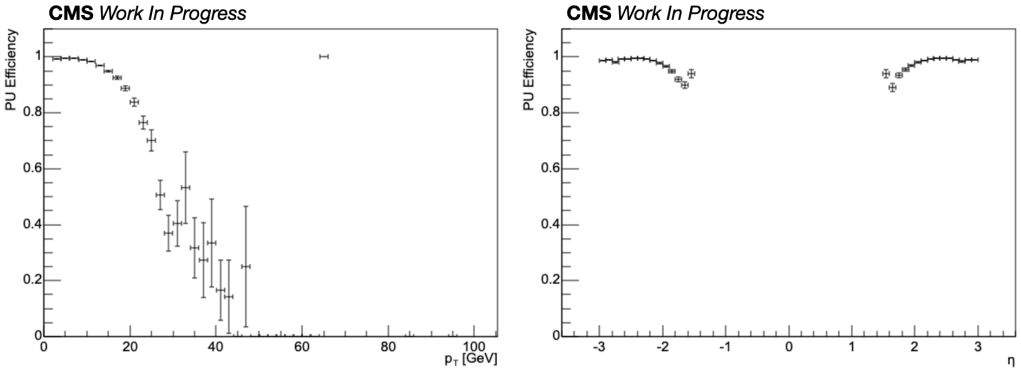


Figure 5.36: Efficiency as a function of p_{T} (left) and η (right) of clusters from PU, having failed the PU ID.

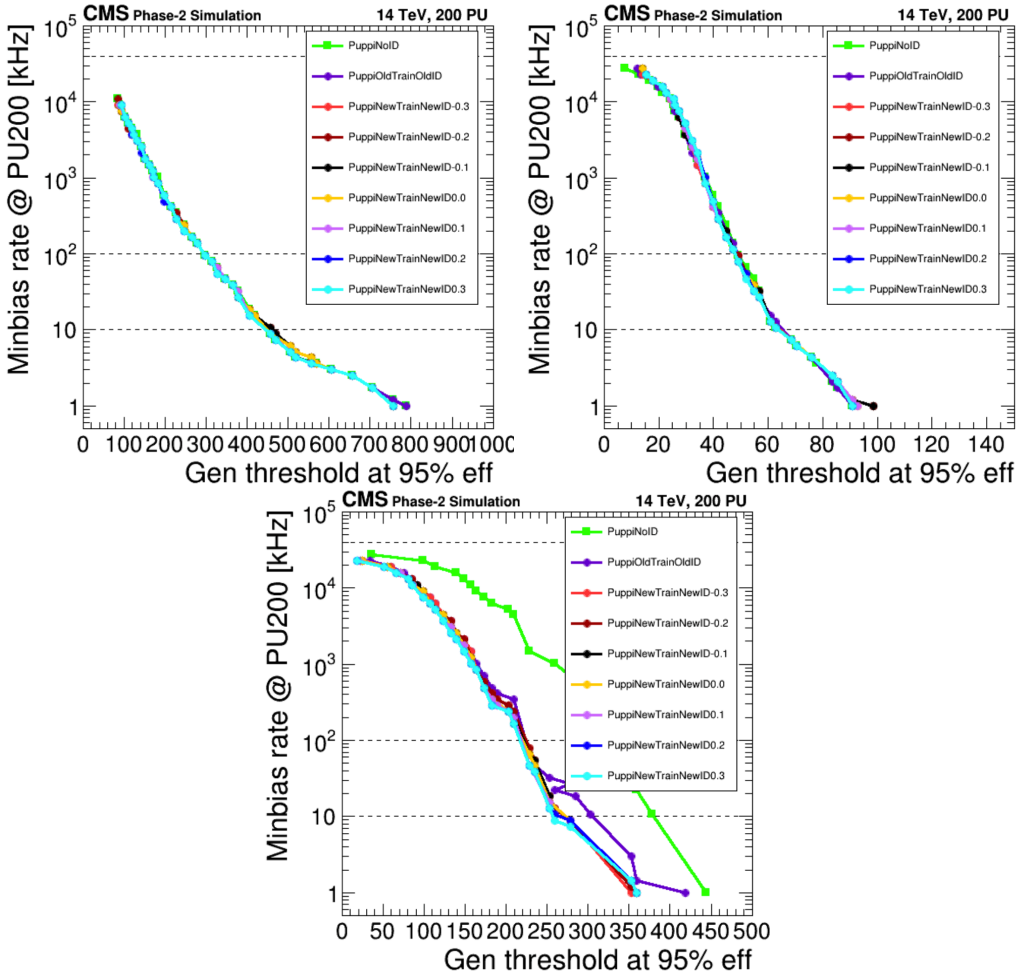


Figure 5.37: Rate of an H_T (upper left), quad-jet (upper right) and $p_{\mathrm{T}}^{\mathrm{miss}}$ (lower) L1 seed versus the corresponding generator variable threshold to achieve 95% efficiency for different WP of the Pion ID BDT response in a $\mathrm{t\bar{t}}$ sample at PU 200. The cases where no Pion ID and where the preliminary Pion ID training are applied are also shown for comparison. The PU ID has already been applied with its WP set at 0.1.

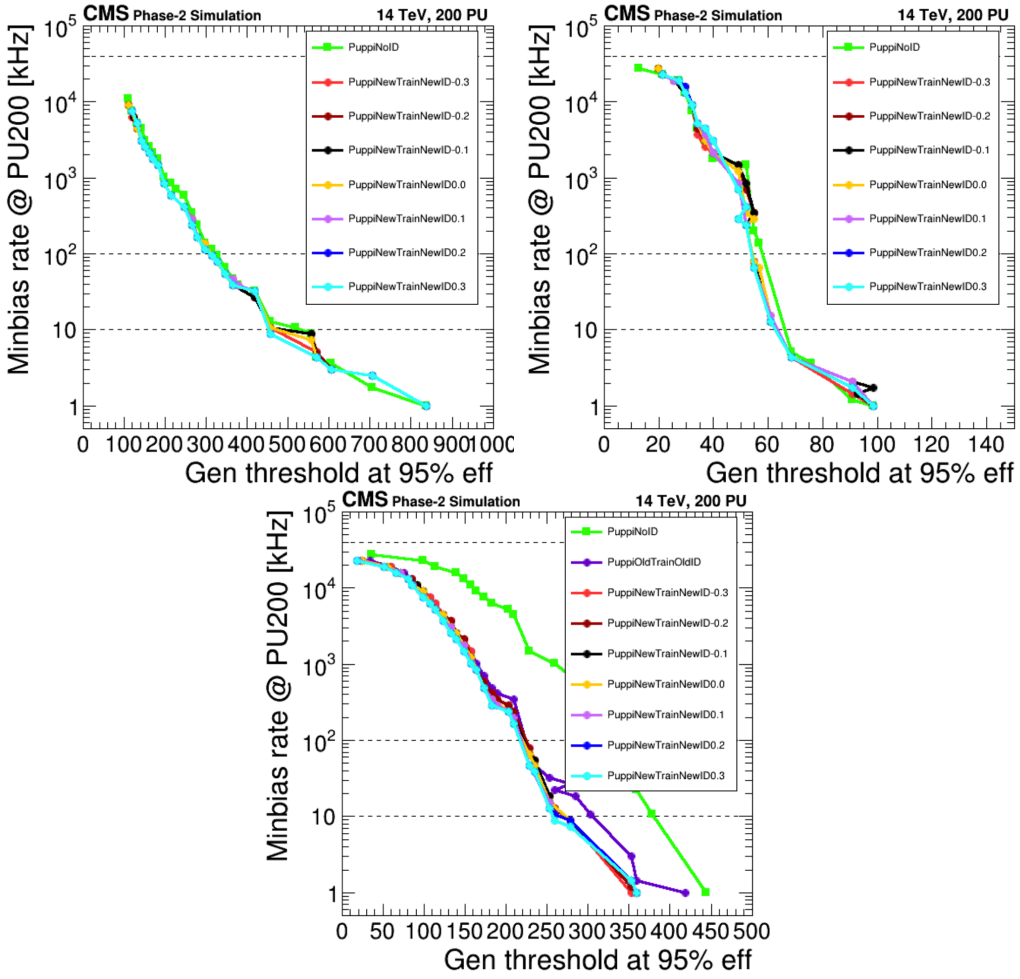


Figure 5.38: Rate of an H_T (upper left), quad-jet (upper right) and $p_{\mathrm{T}}^{\mathrm{miss}}$ (lower) L1 seed versus the corresponding generator variable threshold to achieve 95% efficiency for different WP of the Pion ID BDT response in a VBF H $\rightarrow$ Inv sample at Pion 200. The cases where no PU or Pion ID and where the preliminary Pion ID training are applied are also shown for comparison. The PU ID has already been applied with its WP set at 0.1.

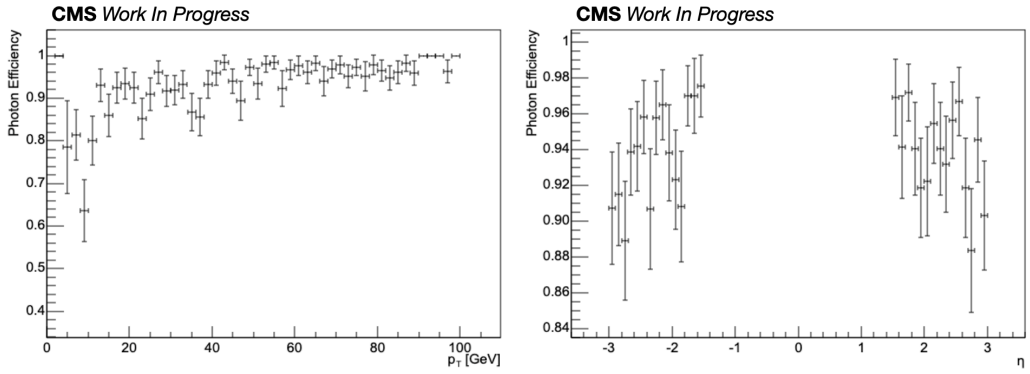


Figure 5.39: Efficiency as a function of p_{T} (left) and η (right) of clusters from genuine photons, having passed the Pion ID.

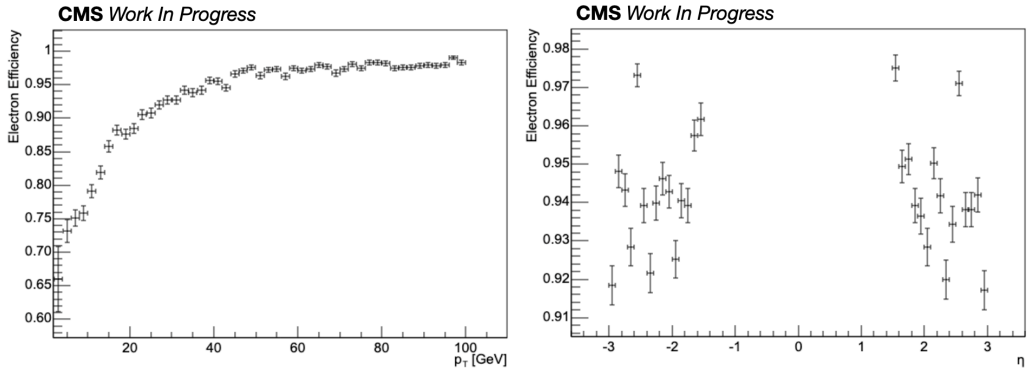


Figure 5.40: Efficiency as a function of p_{T} (left) and η (right) of clusters from genuine electrons, having passed the Pion ID.

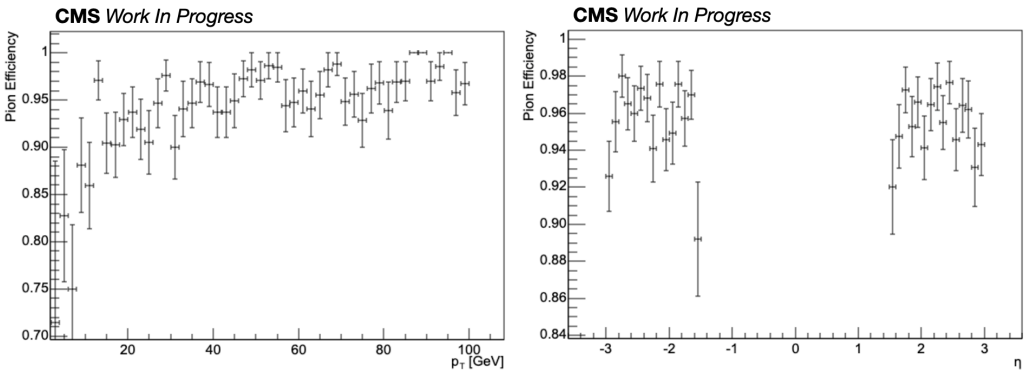


Figure 5.41: Efficiency as a function of p_{T} (left) and η (right) of clusters from genuine pions, having failed the Pion ID.

Parameter	No Tuning	Boldest	Shiest	Tuned
c_0 (photons)	$(3.5, 3.5, 7.0)$	$(-1.5, -1.5, 0.0)$	$(1.5, 1.5, 5.0)$	$(1.5, 1.5, 5.0)$
c_0 (others)	$(5.0, 5.0, 7.0)$	$(0.0, 0.0, 5.0)$	$(2.5, 2.5, 6.0)$	$(4.0, 4.0, 6.0)$
$c_{p_{\mathrm{T}}}$ (photons)	$(0.3, 0.3, 0.3)$	$(0.9, 0.9, 0.9)$	$(0.4, 0.4, 0.4)$	$(0.4, 0.4, 0.4)$
$c_{p_{\mathrm{T}}}$ (others)	$(0.3, 0.3, 0.3)$	$(0.6, 0.6, 0.4)$	$(0.4, 0.4, 0.35)$	$(0.4, 0.4, 0.35)$
p_{T0} (photons)	$(3.0, 4.0, 5.0)$	$(0.0, 0.0, 0.0)$	$(1.5, 2.0, 2.5)$	$(1.5, 2.0, 2.5)$
p_{T0} (others)	$(5.0, 7.0, 9.0)$	$(0.0, 0.0, 3.0)$	$(2.5, 3.5, 4.5)$	$(2.5, 3.5, 4.5)$

Table 5.10: Summary of the different sets of PUPPI parameters used for the optimization of the PUPPI algorithm tuning, after the application of the PU and Pion IDs. The first number corresponds to the parameter value in $1.5 < |\eta| < 2.0$, the second number to the parameter value in $2.0 < |\eta| < 2.4$ and the third number to the parameter value in $2.4 < |\eta| < 3.0$

sponse is observed to be improved by the tuning, especially at low p_{T}, and the resolution is also at the same level or slightly better. The "Tuned" set of PUPPI parameters seems to maintain the best balance between L1 seed efficiency/rate and performance in terms of response/resolution of jets and, hence, is applied in the PUPPI algorithm.

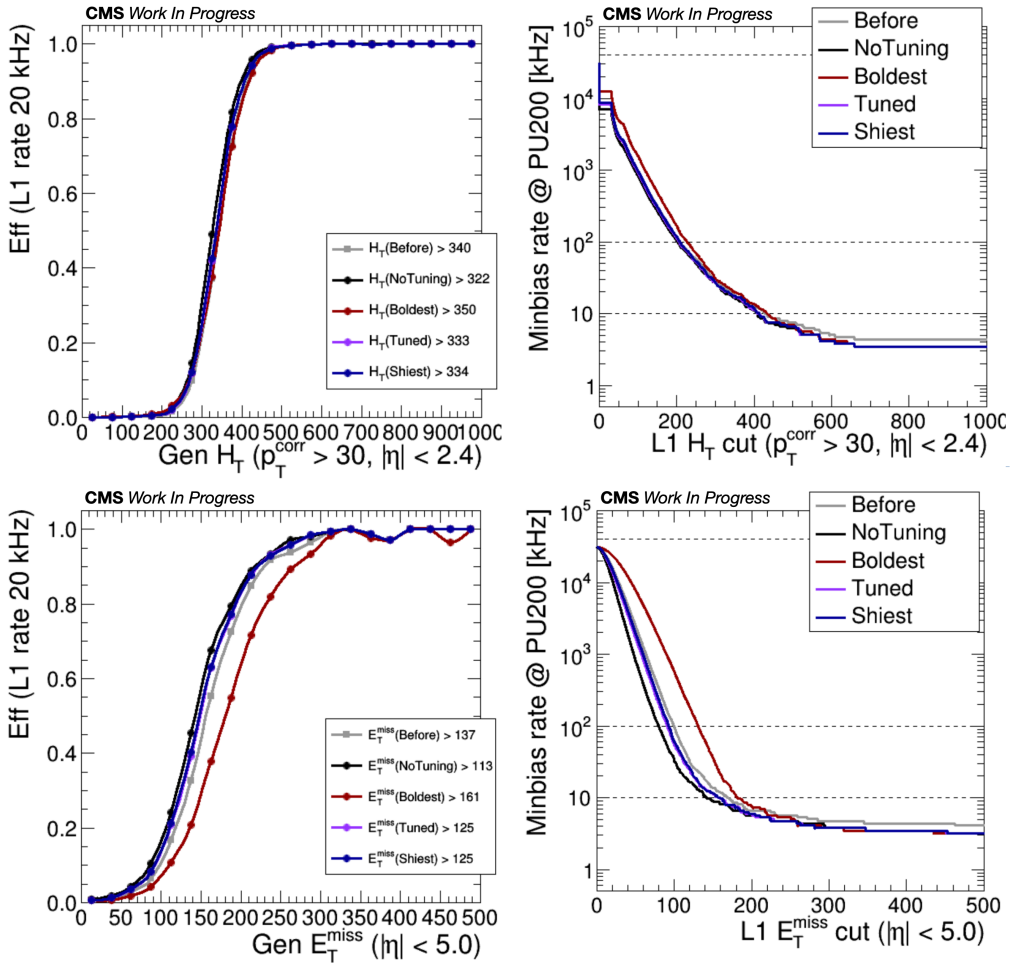


Figure 5.42: The efficiency, at L1 rate of 20 kHz, (left) and rate, at 95% efficiency, (right) of the H_T (upper) and the p_T^{miss} L1 seed (lower) in a $t\bar{t}$ and a VBF H $\rightarrow$ Inv sample, respectively.

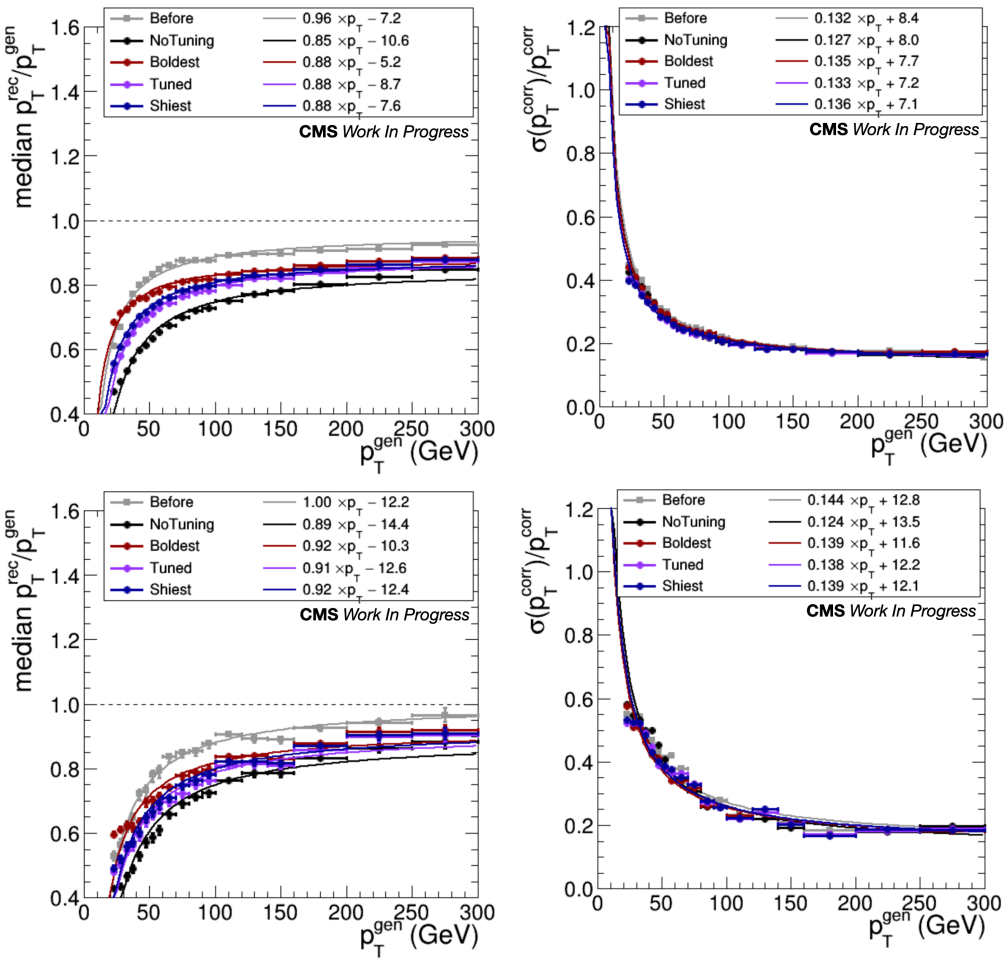


Figure 5.43: The response (left) and resolution (right) of jets as a function of their generator p_{T} in a $\mathrm{t\bar{t}}$ (upper) and a VBF H $\rightarrow$ Inv sample (lower). The plots are for the range $1.5 < |\eta| < 2.0$.

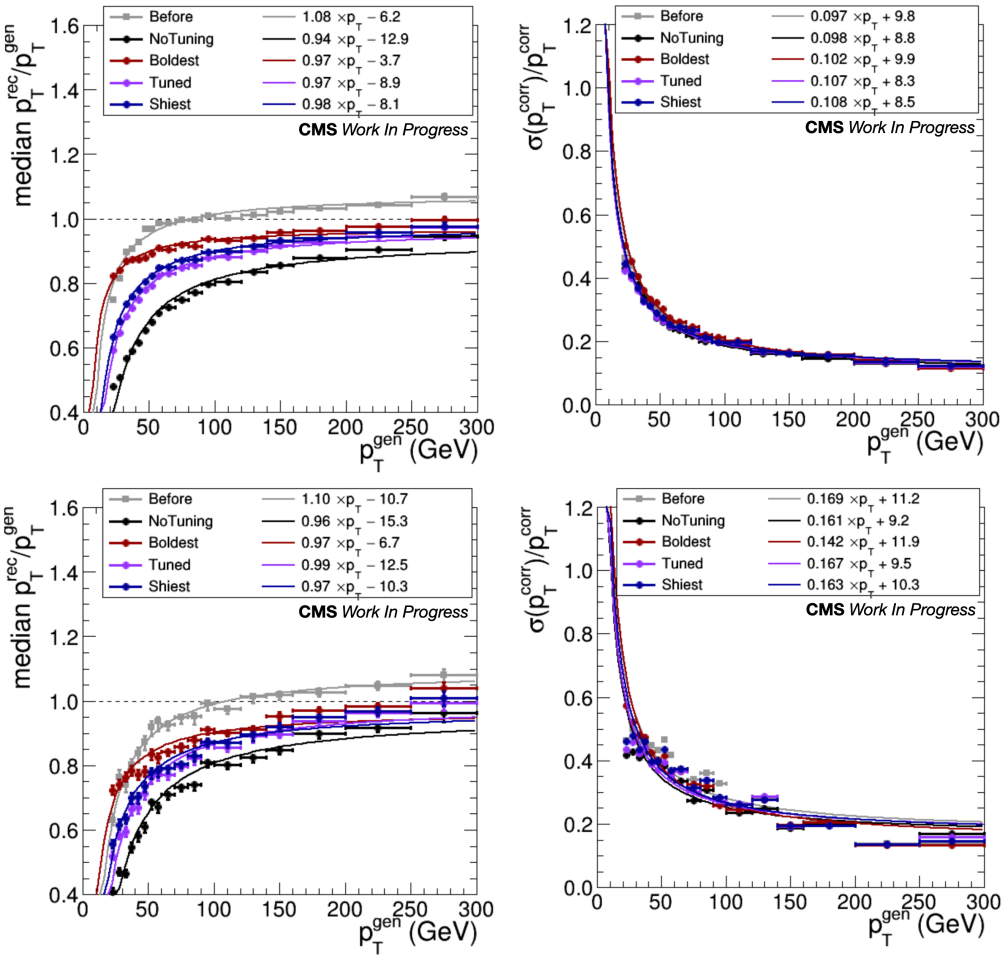


Figure 5.44: The response (left) and resolution (right) of jets as a function of their generator p_{T} in a $\mathrm{t\bar{t}}$ (upper) and a VBF H $\rightarrow$ Inv sample (lower). The plots are for the range $2.0 < |\eta| < 2.4$.

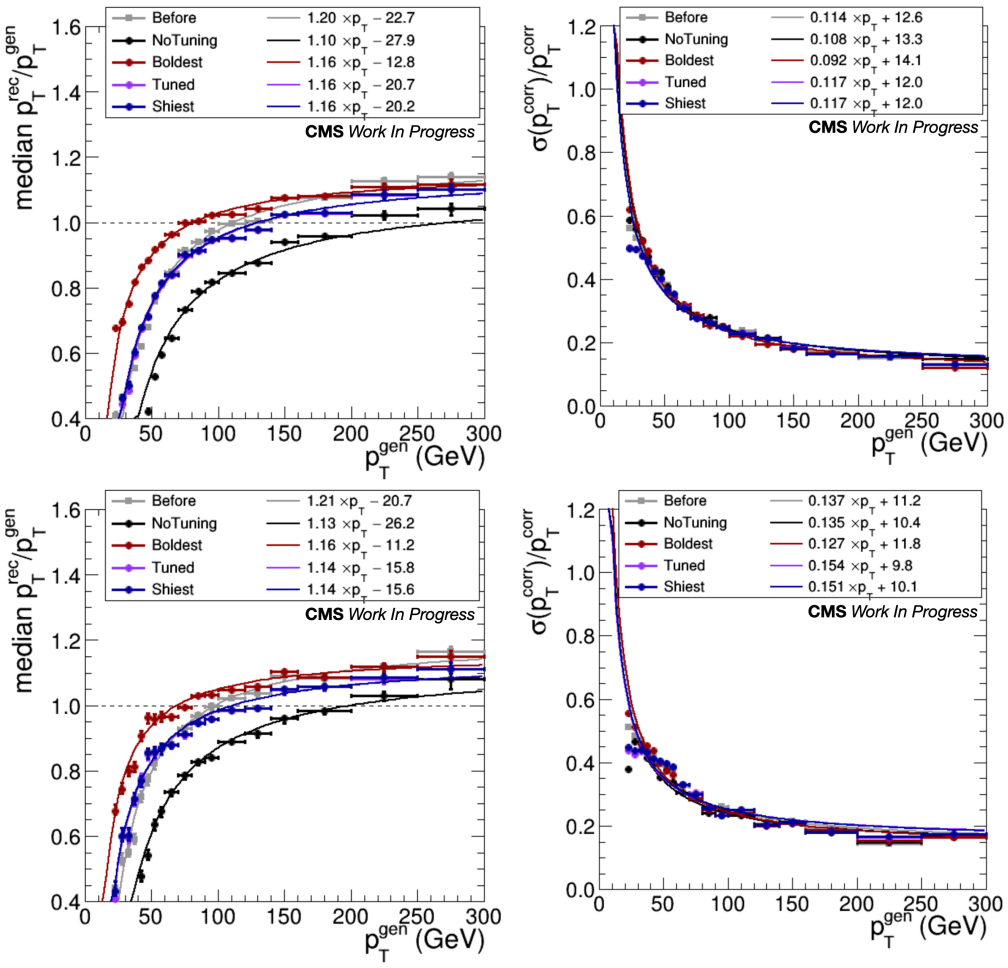


Figure 5.45: The response (left) and resolution (right) of jets as a function of their generator p_{T} in a $\mathrm{t\bar{t}}$ (upper) and a VBF H $\rightarrow$ Inv sample (lower). The plots are for the range $2.4 < |\eta| < 3.0$.

5.4.2 Development of the Emulator for the Deregionizer Firmware Module of the L1CT Layer-2

As briefly mentioned already, the L1CorT Layer-2, which is not regionalized, receives its inputs from the L1CorT Layer-1, which is split into 6 different η-regions, and from the L1CT, which sends the HF η-region. The L1CorT Layer-2 connects with 3 links to each of its upstream boards. As a result, with this architecture, the different links into Layer-2 map into different region of η. Subregions corresponding to different ϕ coordinates are streamed sequentially to Layer-2, hence, in terms of hardware, the ϕ variable is mapped in different slices, i.e. clocks. In summary, the grid of (η, ϕ) blocks of Fig. 5.46 is the input to the Layer-2 for a full event. The PF η-regions are reduced from 7 to 6 by merging the endcap regions with no tracker information with the HF, as described below.

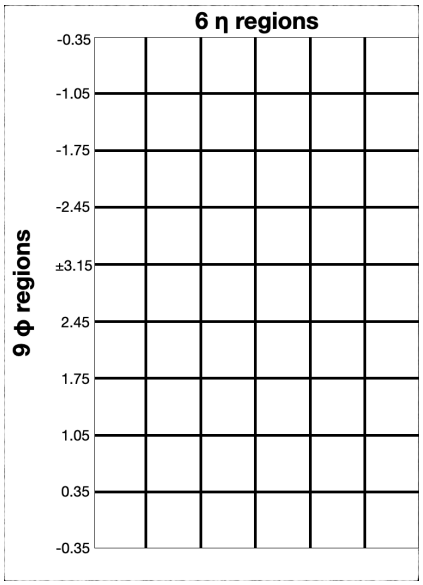


Figure 5.46: The regionalized input of L1CorT Layer-2 over a full event.

Fig. 5.47 (left sketch) shows a strip of PF subregions for L1CorT Layer-1 with the same ϕ. Multiple PF η-regions are displayed, separated by the thick, vertical lines. Internally, the three barrel PF η-regions are further split into two η-subregions, as indicated by the thin lines. As a result, the barrel includes two times the PF subregions per PF η-region compared to the other PF η-regions. In order to stream the PF subregions from a specific

ϕ strip over a common time period for all PF η-regions, a transformation of the input is required, as shown in the right sketch of Fig. 5.47. For the barrel, which includes the additional PF subregions, half of them are streamed during 3 clocks (based on the total bandwidth of the links and the size of the transmitted data) and the other half are streamed in the next 3 clocks. Hence, each Layer-2 input ϕ-block requires 6 clocks to be sent. Given that the detector is split in 9 ϕ blocks and that there is a continuous flow of data from the barrel blocks, 54 clocks are needed for the transmission of the full event. For all the other Layer-2 input η-blocks, only three clocks are needed for each ϕ slice. The endcap regions with tracker are sent separately for positive and negative η, transmitting 3 clocks of data, followed by 3 empty clocks. On the other hand, given the low occupancy of the different endcap regions without tracker and each side of the HF, their PF subregions are merged together in a common η-block for better management of the hardware resources. Therefore, they are streamed over six clocks, similarly to the barrel blocks, as shown in the last column of Fig. 5.47.

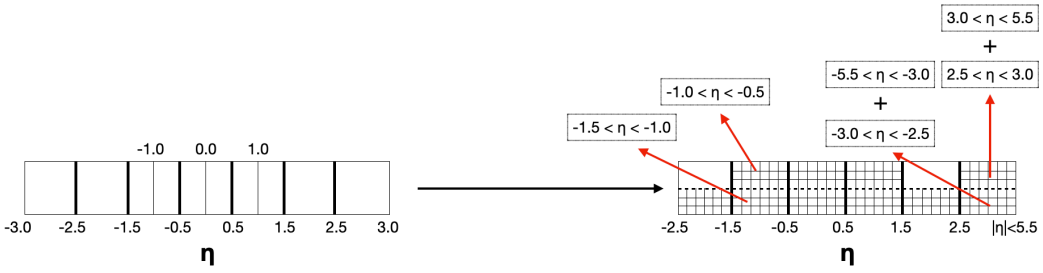


Figure 5.47: The mapping of PF subregions from different η and same ϕ coordinates to different input blocks into the Layer-2 deregionizer.

Fig. 5.48 shows the whole detector grid of Layer-2 input blocks, after it has been transformed to account for the different number of PF subregions, as explained above. Each horizontal line corresponds to a different clock and every different small square to a different PUPPI candidate. In each rectangle, separated by either dashed or solid lines, the PUPPI candidates are p_{T} sorted, starting from the bottom left. These rectangles indicate the different PF subregions and contain 18 PUPPI candidates at most.

The goal of the deregionizer module is to "flatten" this detector grid, i.e. put all of the PUPPI candidates contained in it into a single array (buffer). This is imperative so that the global-objects algorithms can be run on the

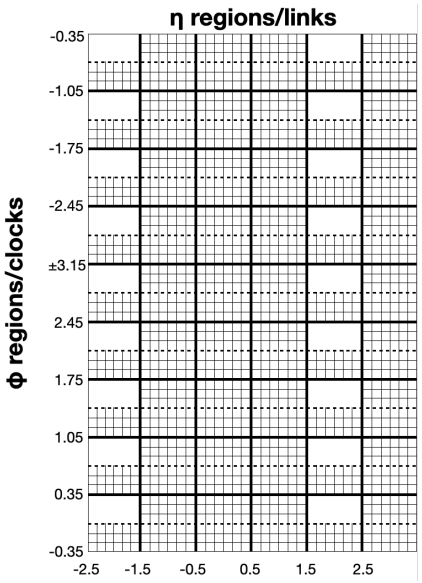


Figure 5.48: The transformed regionalized input of L1CorT Layer-2 over a full event (54 clocks).

final buffer, which holds the candidates from the whole detector. Fig. 5.49 displays the detector grid of a dummy event, which contains PUPPI candidates (colored differently for the different η-blocks) in multiple PF subregions. The deregionizer module consists of the series of buffers on the bottom of the grid: The first set of buffers take the 6 PUPPI candidates from each of the η-blocks. In the next set of buffers, the buffers from the first line are merged. The third set of buffers takes as input the merged or non-merged (depending on their provenance) buffers from the second line, in accordance with the arrows in the sketch. One more merge of buffers follows, then the buffer of the fourth step is accumulated in the final buffer, which ends up including the data from the entire detector after 54 clocks (time needed to stream the event) plus 4 clocks (time needed to stream the data through the deregionizer module). Fig. 5.50 shows an example of how the "merge" and "accumulate" operations are performed within the deregionizer module. It worth noting that the "merge" operation prioritizes the PUPPI candidates on the left buffers and the "accumulate" operation prioritizes the PUPPI candidates already in its target buffer.

On the left side of Fig. 5.49, the (maximum) number of regions that go

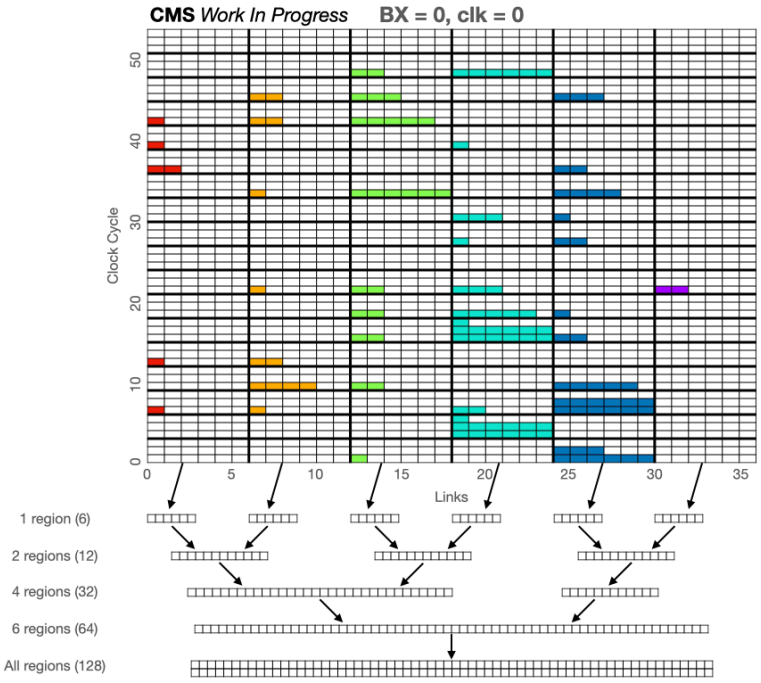


Figure 5.49: The detector grid, as inputted to the L1CorT Layer-2, for a dummy event. PUPPI candidates from different η-blocks are signified by different colors. The structure of the deregionizer module is shown on the bottom. The text on the left indicates how many regions and how many PUPPI candidates each array can contain.

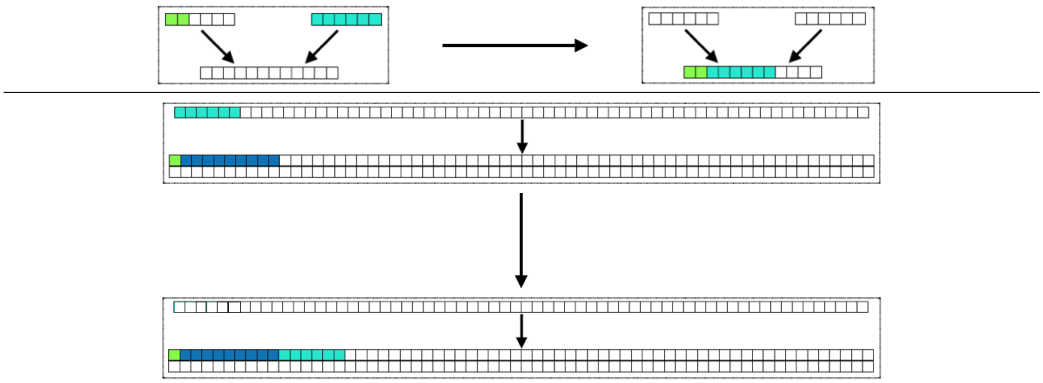


Figure 5.50: The "merge" (upper) and "accumulate" (lower) operations of the deregionizer module.

into each buffer is written, along with the (maximum) size of each buffer in PUPPI candidates. As it is clear by those numbers, the buffers in the first four rows of the deregionizer are sufficiently large to include all the PUPPI candidates streamed to them, without any truncation. On the other hand, the final buffer has a size limit of 128 PUPPI candidates to control the hardware resources needed. The number of PUPPI candidates in a busy topology could potentially exceed 128, so this limit is a subject of study in the following.

In the specific event of Fig. 5.49, more than 128 PUPPI candidates are included. Appx. E shows the clock-by-clock procedure of streaming the event through the deregionizer module. The final result is displayed in Fig. 5.51. As it is shown in the top of the figure, the event is streamed over 6 bunch crossings (BX) and the entire procedure takes 58 clocks. The truncated PUPPI candidates are shown on the right and are discarded in the hardware implementation of the deregionizer module. It is worth mentioning that, even though the input PUPPI candidates are p_{T}-sorted in their PF subregions, the PUPPI candidates in the final buffer are not globally p_{T}-sorted.

The deregionizer module described above has been implemented in firmware. To verify that its firmware implementation works as expected and to study its performance in MC events of different topologies, a software emulator for it was also developed. The deregionizer emulator takes as input the PUPPI candidates, as they are outputted by the L1CorT Layer-1, sorts them in the transformed detector grid, as in Fig. 5.48, and streams them through the series of buffer that constitute the deregionizer module. With respect to the deregionizer firmware, the emulator has two extra features: It makes the size of every buffer configurable and it tracks truncated PUPPI candidates, putting them in a separate array, instead of completely discarding them.

The verification of the firmware implementation of the deregionizer module using the emulator lead to the discovery and solution of minor de-sync issues in the firmware. More specifically, due to an error in setting the latency of running the final deregionizer steps, after the whole event has been streamed to it, to 4 instead of 5 clocks, the final frame of one event was being moved to the next event. After the fix, 100% agreement was observed between the emulator and firmware simulation.

The extra capabilities of the deregionizer emulator allow for the study of the truncations, as a function of the buffer size, and understanding their provenance. The studies are performed in a $\mathrm{t\bar{t}}$ sample at PU 200, in a VBF $\mathrm{H} \rightarrow$ Inv sample at PU 200 and in two four-top quark (TTTT) SUSY sam-

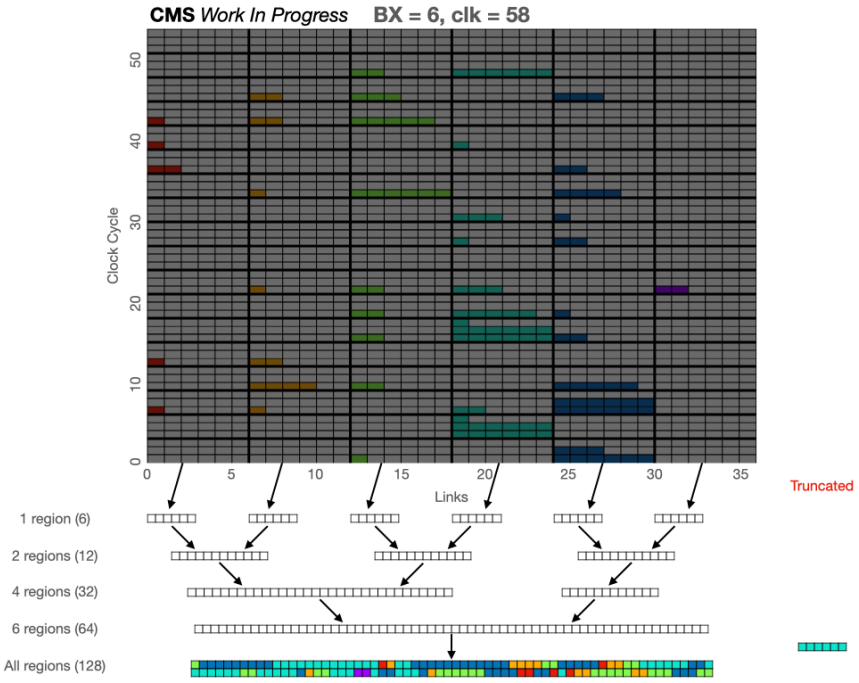


Figure 5.51: The contents of the final buffer, after the whole event has been streamed through the deregionizer module. Since the number of PUPPI candidates is larger than 128, the six last PUPPI candidates are truncated (shown on the right side of the sketch).

ples, one at PU 200 and one at PU 300. Only the results on the TTTT sample at PU 200 are shown in the following, as they exhibit all the important conclusions that can be drawn. To quantify the effect of the inclusion of the deregionizer module, instead of using the entirety of the PUPPI candidates in the idealistic case where they are not subject to the order and size limitation of the deregionizer, the distributions of L1 jets were compared: The *L1PuppiJets* correspond to the jet creation case without the deregionizer and with an offline-equivalent jet finding algorithm (unfeasible for firmware implementation in the L1 trigger). The *scPuppiJets* correspond to the case without the deregionizer but with a L1 trigger jet finding algorithm and the *scDeregPuppiJets* to the case with both the deregionizer and the L1 jet finding algorithm. The *scTruncPuppiJets* correspond to jets that would be reconstructed by the PUPPI candidates truncated due to the presence of the size limit on the final deregionizer buffer.

Focusing on a specific event of the TTTT PU 200 sample with 142 PUPPI candidates, Fig. 5.52 shows a series of tables with explanations on the logic and the effect of truncations. The upper table includes the p_{T}-sorted list of the different jet collections used for these tests. The "L1PuppiJets" column exists as a reference to the ideal case, while the "scPuppiJets" and the "scDeregPuppiJets" collections serve the comparison between the cases where the deregionizer is applied or not. This comparison leads to some interesting results, annotated on the tables. As indicated in the upper table of Fig. 5.52, the highest p_{T} jet, common in both the "L1PuppiJets" and "scPuppiJets" collections, is not present in the "scDeregPuppiJets". Noticing its coordinates, the reason behind this is the value of its ϕ variable: The ϕ slice that this "lost jet" falls in is one of the last to be processed (c.f. Fig 5.48), hence at least some of the PUPPI candidates that constitute it are discarded.

As a first conclusion to be drawn, PUPPI candidates with $-\pi < \phi < 0$ are more likely to be truncated, due to the sequence that the PF subregions are inputted in the L1CorT Layer-2. Depending on the number of PUPPI candidates truncated in a region of the detector, lost jets may be produced with underestimated p_{T} or may not be produced at all. In the specific case discussed above, it seems that almost the entire set of PUPPI particles of the lost jet are rejected. This follows from the fact that, checking the collection of jets produced by the PUPPI candidates that have been truncated by the deregionizer ("scTruncPuppiJets"), included in the lower table of Fig. 5.52, the lost jet is reconstructed to a very good approximation in it.

It is crucial to note that the L1CorT Layer-2 does not have the resources to reconstruct all the jets in a event, in contrast to the ideal case. Because of this, the "scPuppiJets" and "scDeregPuppiJets" collections are cropped to at most 10 objects. This leads to a second-order effect of the truncations: When a jet is truncated, another jet, in most cases but not always with lower p_{T}, takes its place in the jet collection. This jet can come from any region of the detector with $\phi \lesssim \phi_{\text{lost jet}}$. As a result, the truncations from the deregionizer do not really reduce the number of jets in an event but rather "transfer" jets from the negative-ϕ region of the detector to the rest of the detector.

Having understood the origin of the truncations in individual events and their repercussions, the total impact of truncations can be quantified by studying the p_{T}, η and ϕ distributions of scPuppiJets, scDeregPuppiJets and scTruncPuppiJets in the TTTT PU 200 sample. The upper left plot of Fig. 5.53 indicates that most of the truncated jets are of low p_{T}, less than $\sim$20 GeV. However, there is also a nonnegligible amount of truncated jets

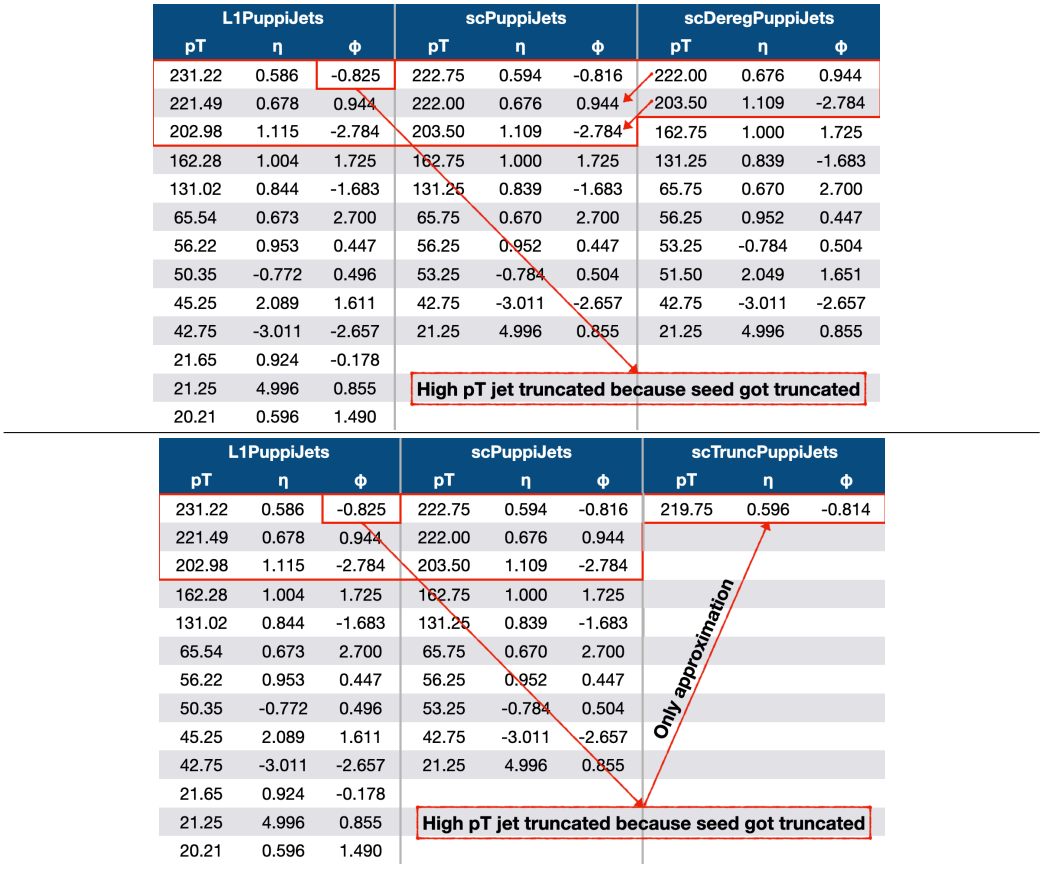


Figure 5.52: List of the first, p_{T} sorted L1PuppiJets, scPuppiJets, scDeregPuppiJets and scTruncPuppiJets for the event under discussion in the main text. The annotations on the tables are discussed in the main text as well.

with higher p_{T}. This can be problematic in the case where such a truncated jet would have otherwise caused a L1 seed to fire. The η distribution (upper right plot of Fig. 5.53) does not point to a specific bias. Finally, the ϕ distribution (lower plot of Fig. 5.53) confirms the conclusions above that more and more jets are lost as in the range $-\pi \to 0$. The "transferring" of jets is confirmed as well: Even though the "Puppi" distribution tends to be higher in the $-\pi < \phi < 0$ region, in the rest of the detector, the "DeregPuppi" distribution tends to be higher, due to the inclusion of jets there that replace the lost ones in the collection.

Both the rejection of PUPPI candidates that form high p_{T} jets as well as

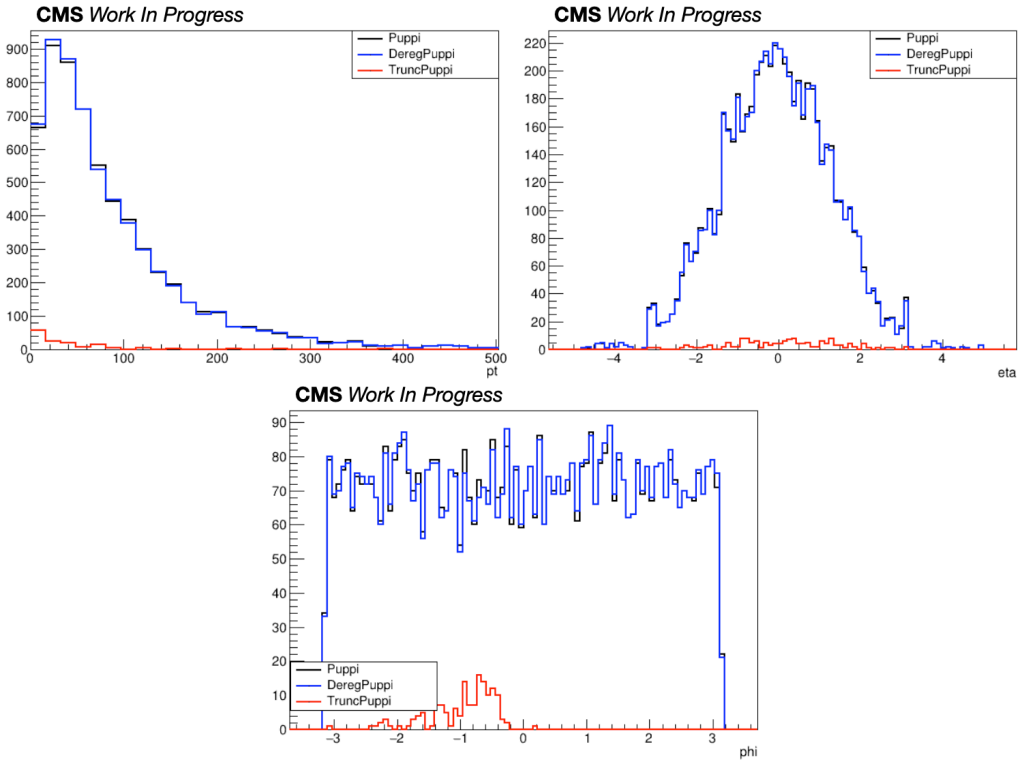


Figure 5.53: Distributions of the p_{T}, η and ϕ variables of scPuppiJets (labeled as "Puppi"), scDeregPuppiJets (labeled as "DeregPuppi") and scTruncPuppiJets (labeled as "TruncPuppi") in a TTTT PU 200 sample.

the bias in the ϕ distribution call for improvement in the deregionizer design. A straightforward solution is the increase of the final buffer size from 128 to 256 PUPPI candidates (a multiple of 2 to facilitate the resource allocation in hardware). Repeating the previous plots with a deregionizer with a final buffer containing up to 256 PUPPI candidates removes all truncations and leads to 100% agreement between the case with and without the application of the deregionizer (black and blue lines overlap in Fig. 5.53). The same result is confirmed in the most busy topology of TTTT at PU 300.

Although it is conceptually simple, the expansion of the deregionizer final buffer requires attention in its hardware implementation. The increase in hardware resources due to this change is summarized in Table 5.11. Multiple firmware modules are affected and their resource usage approximately doubles. However, given the small requirements of those modules to begin

with, the increase in resources is still acceptable. The change is accompanied by a small increase in latency (∼30 ns), which can also be accommodated. Finally, some tweaks were needed to make the new deregionizer design meet timing. In view of the benefits that the final buffer expansion brings and the feasibility of its application in firmware, the deregionizer module is now included in the L1CorT Layer-2 design with 256-long final buffer.

N = 128				
Module	LUTs (%)	FFs (%)	DSPs (%)	BRAMs
Total	6.8	6.4	7.5	∼0
Deregionizer	3.4	2.3	—	—
JetLoop	1.9	2.1	3.7	—
JetCompute	0.9	1.0	3.8	∼0
N = 256				
Module	LUTs (%)	FFs (%)	DSPs (%)	BRAMs
Total	9.8	10.8	15.0	∼0
Deregionizer	3.4	2.3	—	—
JetLoop	3.9	4.7	7.5	—
JetCompute	1.7	2.6	7.5	∼0

Table 5.11: The hardware resource usage of the deregionizer and jet finding algorithm modules in a VU9P FPGA for the cases of 128 (upper) and 256 (lower) PUPPI candidates in the final deregionizer buffer.

The different deregionizer configurations were tested in terms of L1 seed performance as well. Fig. 5.54 shows the efficiency and rate of the $p_{\mathrm{T}}^{\mathrm{miss}}$ L1 seeds in the case of ideal jets ("L1Jets") and of L1 jets with the application of the deregionizer with 128 ("DeregPuppi") and 256 ("Dereg256Puppi") PUPPI candidates in the final buffer. The threshold values for the rate calculation have been derived within a conservative 10% uncertainty margin to emulate potential differences between the data and simulation. Only small differences are present: The differences between the "L1Jets" and the "Dereg256Puppi" mainly originate from the better resolution of the offline-equivalent jets of the former collection. The differences between "DeregPuppi" and "Dereg256Puppi" come from the fact that the former collection is afflicted by truncation, hence creating $p_{\mathrm{T}}^{\mathrm{miss}}$ imbalances in some events, mostly evident at lower $p_{\mathrm{T}}^{\mathrm{miss}}$ values. The differences are even more subtle in the H_T and the quad-jet L1 seed performance.

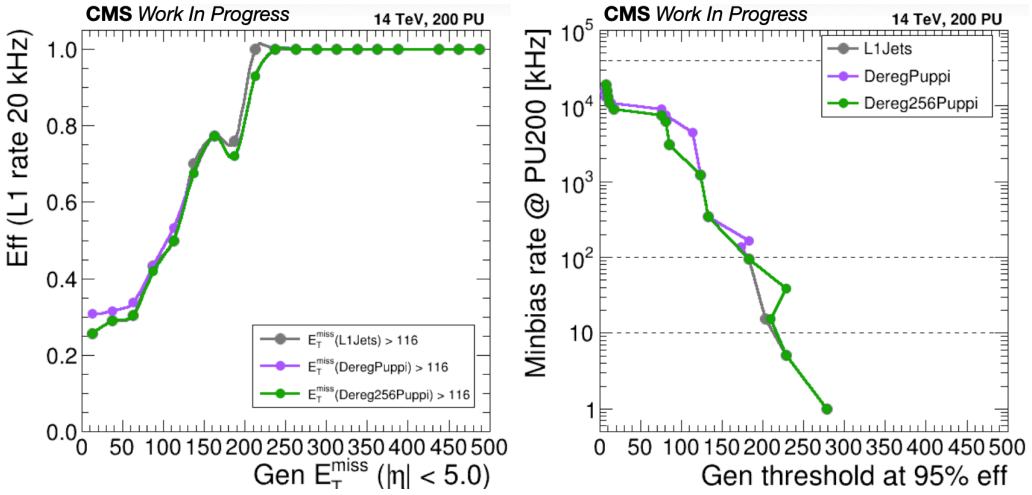


Figure 5.54: The efficiency (left), at 20 kHz L1 rate, and the rate (right), at 95% efficiency of the $p_{\mathrm{T}}^{\mathrm{miss}}$ L1 seed, using L1PuppiJets (labeled as "L1 Jets"), scDeregPuppiJets (labeled as "DeregPuppi") and scDeregPuppiJets with the final buffer size increased to 256 PUPPI candidates (labeled as "Dereg256Puppi") in a TTTT PU 200 sample. Higher rates for higher thresholds seen at some points are fluctuations within the 10% uncertainty margin discussed in the text.

Fig. 5.55 aims to investigate the jets that take the place of the truncated ones. The upper table in the figure highlights the jet that was added in the "scDeregPuppiJets" collection to replace the one that was lost due to truncations. It is worth noticing that this jet is not added to the bottom of the collection but in a higher spot. This implies that it is not necessarily the lower p_{T} jets that are not produced due to the 10-jet limit. The explanation of this is included in the lower table of the same figure, where the p_{T}-sorted list of the jet seeds is quoted. This shows that the creation of jets is performed by order of the seed p_{T} and not of the jet p_{T}, a limitation caused by the firmware implementation of the jet finding algorithm. In a significant portion of the cases, higher-p_{T} seeds do not lead to higher-p_{T} jets. This is the case for the event discussed above, where a 51.5 GeV jet is created by a 18.25 GeV seed, and, therefore, is lower in the collection with respect to a 21.25 GeV jet starting from a seed with the same p_{T}.

To avoid discarding high-p_{T} jets as much as possible, some changes in the jet finding algorithm are implemented. First of all, the number of output jets

L1PuppiJets			scPuppiJets			scDeregPuppiJets		
pT	η	ϕ	pT	η	ϕ	pT	η	ϕ
231.22	0.586	-0.825	222.75	0.594	-0.816	222.00	0.676	0.944
221.49	0.678	0.944	222.00	0.676	0.944	203.50	1.109	-2.784
202.98	1.115	-2.784	203.50	1.109	-2.784	162.75	1.000	1.725
162.28	1.004	1.725	162.75	1.000	1.725	131.25	0.839	-1.683
131.02	0.844	-1.683	131.25	0.839	-1.683	65.75	0.670	2.700
65.54	0.673	2.700	65.75	0.670	2.700	56.25	0.952	0.447
56.22	0.953	0.447	56.25	0.952	0.447	53.25	-0.784	0.504
50.35	-0.772	0.496	53.25	-0.784	0.504	51.50	2.049	1.651
45.25	2.089	1.611	42.75	-3.011	-2.657	42.75	-3.011	-2.657
42.75	-3.011	-2.657	21.25	4.996	0.855	21.25	4.996	0.855
21.65	0.924	-0.178						
21.25	4.996	0.855	Extra jet that took the place of the truncated one					
20.21	0.596	1.490						

# Jet	Jet pT	Seed pT
1	203.5000	67
2	65.7500	56
3	42.75	42.75
4	131.2500	37
5	222	34
6	56.2500	28.75
7	162.75	21.75
8	21.25	21.25
9	53.2500	19.25
10	51.5	18.25

scPuppiJets			scDeregPuppiJets		
pT	η	ϕ	pT	η	ϕ
222.75	0.594	-0.816	222.00	0.676	0.944
222.00	0.676	0.944	203.50	1.109	-2.784
203.50	1.109	-2.784	162.75	1.000	1.725
162.75	1.000	1.725	131.25	0.839	-1.683
131.25	0.839	-1.683	65.75	0.670	2.700
65.75	0.670	2.700	56.25	0.952	0.447
56.25	0.952	0.447	53.25	-0.784	0.504
53.25	-0.784	0.504	51.50	2.049	1.651
42.75	-3.011	-2.657	42.75	-3.011	-2.657
21.25	4.996	0.855	21.25	4.996	0.855
Extra jet that took the place of the truncated one					

Figure 5.55: List of the first, p_{T} sorted L1PuppiJets, scPuppiJets and scDeregPuppiJets for the event under discussion in the main text. The list of the p_{T}-ordered jet seeds is included, along with the p_{T} of the jets that they create. The annotations on the tables are discussed in the main text as well.

is increased from 10 to 12 to limit the number of truncated jets. Additionally, improvements in the jet loop firmware module, which reduce its latency from 14 to 8 clocks, allow the processing of 16 seeds for jet creation (760 ns) in the same time that previously only 10 seeds were processed (792 ns). Combining these updates, the jet finding algorithm now uses a list of 16 seeds to create 16 jets, p_{T}-sorts them and then send the first 12 of them to downstream systems. This way, the probability of truncating high-p_{T} jets reduces significantly.

5.4.3 Development of the Firmware Module and the Emulator for the Sorting of L1 e/γ Objects

In contrast with the PF and PUPPI candidates, which are inputted in the deregionizer module to get properly formatted for the purposes of L1CorT Layer-2, as described in the previous section, e/γ objects are handled differently. Even though e/γ objects are p_{T}-sorted in the PF subregions, just like PF and PUPPI candidates, the algorithms of Layer-2 require them to also come p_{T}-sorted per Layer-1 board. At the same time, a large amount of such objects can potentially be present in a single board, creating the need to reduce their number for downstream calculations to keep the hardware resource usage under control. Consequently, a firmware module is needed in order to sort the e/γ objects for all PF subregions in a single Layer-1 board and then possibly truncate the resulting list of objects to discard low-p_{T} ones. This firmware module along with its corresponding emulator were developed as part of this thesis.

The firmware module, called *e/γ final sorter*, takes as input the N highest p_{T} e/γ objects per PF subregion for all subregions in a single board. Its output is the list of M highest p_{T} e/γ objects per board. In accordance with the rest of the firmware modules of Layer-1, it should be run at TMUX 6. As a result, if the e/γ final sorter is run at 360 MHz, it then has 360/6 clocks per event. Given that 9 PF subregions are processed per board, this means that the e/γ final sorter has $\lfloor 360/6/9 \rfloor = 6$ clocks per region.

The e/γ final sorter is optimized for the endcap region, where the number of e/γ objects per PF subregion is at most 6, hence $N = 6$ is set. The fact that number of input objects are less than the number of clocks available for the module to run eliminates the need for multiple lists to keep track of the input objects as they are waiting to be sorted in the final output list. Because of this, the e/γ final sorter module has been implemented as an incremental sorter: It takes one new object per clock and adds it to the final sorted list on the fly. This sorting method is resource-light, since, apart from the input and output list of e/γ objects, the only other computational object it requires is an additional boolean input to mark the beginning of the events. Due to the different multiplicity of e/γ objects in the barrel regions, the implementation of the e/γ final sorter is more complicated there. The firmware module version optimized for these regions is beyond the scope of this thesis.

The firmware implementation has been tested with the VU9P FPGA

as the target device, using the Xilinx Vivado High-Level Synthesis libraries (version 2019.2) [139]. As expected, the algorithm mainly uses LUTs, to store the input and output list, while FF are also reserved for the comparison of the p_{T} among the different e/γ objects. The utilization of these resources by the e/γ final sorter amounts to less than 1% of the total VU9P resources, conveniently fitting to the existing L1CorT Layer-1 design. The timing is also met, estimated to 2.34±0.34 ns, well within the target of 2.70 ns. The latency was measured to be 19 clock cycles or, equivalently, 51.30 ns, completely overshadowed by the 54 clock cycles needed to stream an event from Layer-1 to Layer-2. Fig. 5.56 shows the floorplan of the Layer-1 algorithms on a VU9P FPGA, in which the final e/γ sorter only takes a small amount of space on the upper right part.

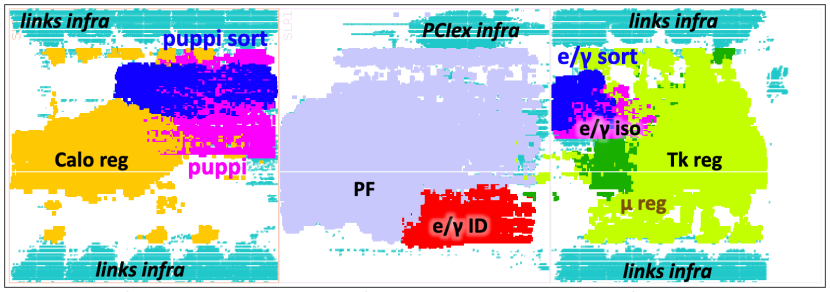


Figure 5.56: Floorplan of the L1CorT Layer-1 firmware algorithms on a VU9P FPGA: The "links infra" and "PCIex infra" labels refer to the needed logic for the I/O of the FPGA, the "μ reg", "Tk reg" and "Calo reg" labels refer to the regionizer logic for muons, tracks and calorimetric inputs respectively, while the rest of the labels are self-explanatory.

The e/γ final sorter emulator was developed in parallel with the relevant firmware and agrees 100% with it, tested on 1000 events of a $\mathrm{t\bar{t}}$ sample at PU 200. The only notable feature of the incremental sorter algorithm that required attention during the development of the emulator is the reverse sorting of objects with equal values. Apart from that, the emulator was implemented with configurable number of incoming and outgoing e/γ objects, so that the impact of truncations in the input and output of the algorithm can be studied in simulation. This kind of studies is part of a larger collection studies requiring the integration of multiple e/γ-related algorithms in L1CorT Layer-1 and, as a result, is beyond the scope of this thesis.

Summary and Outlook

This thesis presented a search for New Physics, more specifically for Supersymmetry (SUSY), using the full Run 2 dataset (2016–2018) of proton-proton collisions recorded by the Compact Muon Solenoid (CMS) experiment operating at the Large Hadron Collider (LHC). The analyzed data correspond to a total integrated luminosity of up to $137\,\mathrm{fb}^{-1}$, at a center-of-mass energy, $\sqrt{s} = 13\,\mathrm{TeV}$.

The search is designed to probe signatures with two or three low transverse momentum (p_T) leptons (electrons or muons) and moderate to high missing transverse momentum (MET). At least one of the lepton pairs must have opposite sign (OS) charge and same flavor (SF), while the MET is induced by an initial state radiation jet. Such final states occur in SUSY models involving the pair production of the lightest chargino ($\widetilde{\chi}_1^\pm$) and neutralinos ($\widetilde{\chi}_2^0$ and $\widetilde{\chi}_1^0$), collectively called "electroweakinos" (EWKinos), in the case where these are almost mass degenerate. Due to the small mass difference (Δm) between the produced SUSY particles and the lightest SUSY particle, these models are said to have a *compressed mass spectrum*. Multiple such simplified models (SMS) were presented in the thesis, depending on the composition of the degenerate SUSY particles. In the TChiWZ SMS, the lightest EWKinos are mostly either bino or wino, while in the Higgsino SMS, they are mostly higgsino. The latter case is also explored in the framework of a more general model, the phenomenological minimal supersymmetric standard model (Higgsino pMSSM). Additionally, the search is sensitive to SUSY models with top squark pair production and a small Δm between the top squark and the neutralino LSP. The case where the top squark undergoes a four-body decay is described in the T2bff$\widetilde{\chi}_1^0$ SMS, while its decay via a chargino is described in the T2bW SMS. All the models probed were motivated in the thesis by theoretical considerations.

The analysis strategy is optimized for compressed mass scenarios, by requiring the invariant mass of the OSSF lepton pair ($M(\ell\ell)$) to be in the range between 1 and 50 GeV and the MET to be greater than 125 GeV. The $M(\ell\ell)$ exhibits an endpoint at the value of the signal Δm, and this is exploited to further enhance the analysis sensitivity. The event selection is carefully designed to reject the dominant standard model (SM) backgrounds to the analysis: nonprompt processes, i.e. leptons that do not come from the primary vertex, the $\mathrm{t\bar{t}}$, DY and $\mathrm{WZ} \to 3\ell$ processes. Tight identification criteria on the impact parameters and the isolation of leptons reduce the contribution from nonprompt leptons, while a veto for jets originating from b quarks reduces the contribution of the $\mathrm{t\bar{t}}$ process. The DY process mainly contributes to the analysis background with the decay to τ leptons, which includes MET due to the presence of neutrinos in the final state. It is suppressed by rejecting the events having the approximately reconstructed ditau mass in a range around the Z boson mass. Finally, the $\mathrm{WZ} \to 3\ell$ process is effectively minimized by vetoing events with $M(\ell\ell)$ near the Z boson mass. Robust, data-driven methods combining information from multiple regions are used to estimate the contribution of the aforementioned SM processes. The smaller background contributions of other, rare SM processes are taken directly from simulation.

The statistical and systematic uncertainties, including both experimental effects as well as imperfections of the modeling of the simulation, are carefully studied. Within these uncertainties, the data are found to be in agreement with the SM predictions. A binned maximum likelihood fit with the uncertainties included as nuisance parameters is used to extract exclusion limits on SUSY particle production cross sections. At 95% confidence level, the $\widetilde{\chi}_1^{\pm}$ and $\widetilde{\chi}_2^0$ are excluded for masses up to 275 GeV for $\Delta m = 10$ GeV in the TChiWZ SMS, and up to 150 GeV for $\Delta m = 3$ GeV in the Higgsino SMS. In terms of the Higgsino pMSSM model, the higgsino mass parameter, μ, is excluded up to 180 GeV for bino mass parameter $M_1 = 800$ GeV. For the models involving top squark production, for $\Delta m = 30$ GeV, top squark masses are excluded up to 540 GeV in the T2bff$\widetilde{\chi}_1^0$ SMS and up to 480 GeV in the T2bW SMS.

In addition to the search for New Physics, this thesis also presented a series of software and firmware algorithms developed for the Phase 2 upgrade of the Level-1 Trigger (L1T) system of the CMS experiment, planned to be deployed during the High Luminosity LHC (HL-LHC) operation.

The thesis discussed the implementation of a dedicated L1T scouting sys-

tem which is planned for the Phase 2 L1T upgrade. The L1T scouting system will be able to record data at real time, without any preselection, albeit with lower resolution. The vast amount of data envisioned to be captured by it has the potential to probe rare processes, such as the $H \rightarrow \phi\gamma$ decay, and facilitate experimental procedures, such as detector monitoring. The feasibility and utility of such a system has been tested in a muon scouting demonstrator implemented at the end of the data-taking period of 2018. The thesis described the extraction of data-driven corrections for improving the resolution of muons captured by the scouting system and presented studies for their performance.

A significant portion of the work presented in this thesis is dedicated to the L1T Correlator system, which is also planned for Phase 2 L1T upgrade and will be responsible for executing the higher level object reconstruction for the L1T system. Within the L1T Correlator, a multivariate algorithm, more specifically a Boosted Decision Tree, was developed in order to classify the clusters incoming from the future CMS calorimeter, the High Granularity Calorimeter, as electromagnetic, hadronic or pile-up, i.e. not originating from the hard interaction. The performance of this algorithm in terms of rate reduction as well as of trigger efficiency was presented. Furthermore, the development of the software emulator for one of the L1T Correlator data organization algorithms, called "deregionizer", was discussed. Bias studies based on this emulator led to improvements of the data organization logic and were presented in the thesis. Last but not least, this thesis described the development of the firmware module, and its corresponding emulator, responsible for the sorting of the electromagnetic of the L1T Correlator according to their p_{T}.

Appendix A

Extra Material on the Trigger Efficiencies

This appendix is dedicated to adding more detailed information on the subject of trigger efficiency, summarized in Sec. 4.2.1.

A.1 Pure $p_{\mathrm{T}}^{\mathrm{miss,corr}}$ trigger paths

For the pure $p_{\mathrm{T}}^{\mathrm{miss,corr}}$ trigger paths, their efficiency on data has been estimated in a data set including events captured by single muon triggers, so that the measurement is unbiased. More concretely, the events were required to fire the trigger paths required an isolated muon with p_{T} greater than 24 GeV in 2016 and 2018 and 27 GeV in 2017. A requirement for H_{T} greater than 100 GeV, compatible with the offline ISR jet selection of the analysis and with the purpose to also cover for the backup paths, was set.

The efficiency for each year is shown in Fig. A.1. Due to the limited resolution, especially at the L1 trigger, an approximately stable efficiency plateau is reached at the offline value of $p_{\mathrm{T}}^{\mathrm{miss,corr}}$ of ~ 200 GeV compared with the 125 GeV online threshold. This is what drives the offline $p_{\mathrm{T}}^{\mathrm{miss,corr}}$ requirements of the analysis (more details in Sec. 4.3.4). To better estimate the efficiency (ϵ) at the turn-on and the plateau, the following parametrization is used for the efficiency curve:

$$\epsilon = 0.5 \times \epsilon_{\infty} \times \left(Erf\left(\frac{p_{\mathrm{T}}^{\mathrm{miss,corr}} - \mu}{\sigma} \right) + 1 \right) \tag{A.1}$$

where ϵ_∞, μ and σ are parameters of the fit, which are given separately for each year in the plots of Fig. A.1.

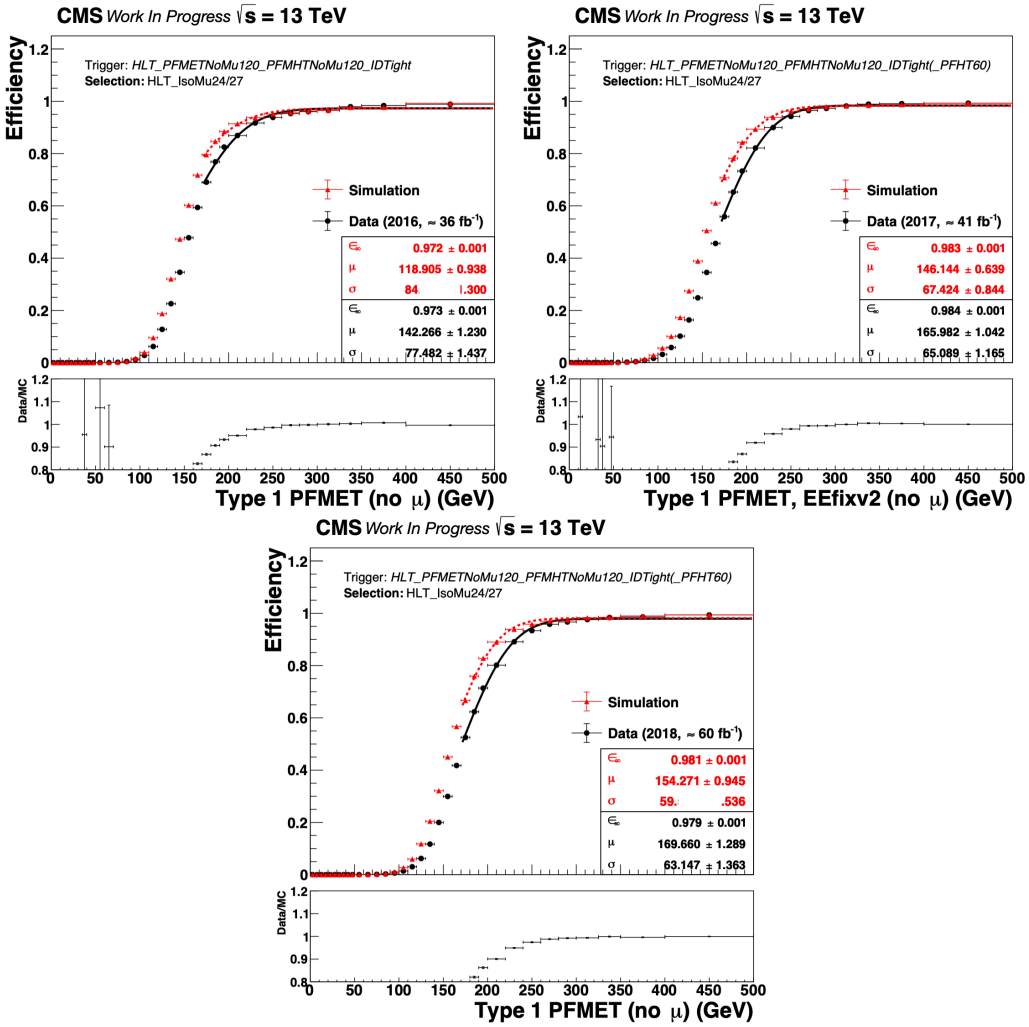


Figure A.1: Trigger efficiency as a function of the offline $p_{\mathrm{T}}^{\mathrm{miss,corr}}$ measured in both data and simulation for the pure $p_{\mathrm{T}}^{\mathrm{miss,corr}}$ trigger paths of 2016 (upper left), 2017 (upper right) and 2018 (bottom).

A.2 Double-μ+MET trigger path

Given the presence of many more requirements in the double-μ+MET trigger path, needed to lower the MET thresholds, the efficiency measurement is more complicated. In fact, it is broken down to several separate components:

$$\epsilon = \epsilon_{\mu_1}\epsilon_{\mu_2} \times \epsilon_{\mu,\text{distance}} \times \epsilon_{\text{MET}} \tag{A.2}$$

where the $\epsilon_{\mu_1}\epsilon_{\mu_2}$ (leptonic) part includes the efficiency of the muon and dimuon requirements, except for the invariant dimuon mass requirement, whose inefficiency was found to be negligible (Fig. A.2) and, as a result, its efficiency is assumed to be 1. The $\epsilon_{\mu,\text{distance}}$ term accounts for the efficiency of the DCA or the Δz requirement in 2016 and 2017/2018 respectively. Finally, the ϵ_{MET} component measures the efficiency of the purely MET (hadronic) part (both $p_{\text{T}}^{\text{miss}}$ and $p_{\text{T}}^{\text{miss,corr}}$) of the trigger path.

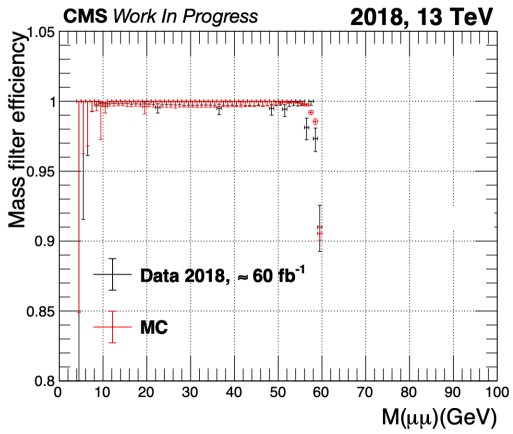


Figure A.2: Trigger efficiency as a function of the offline invariant dimuon mass measured in both data and simulation for the double-μ+MET trigger path of 2018.

The leptonic part of the efficiency is computed with the *tag-and-probe* method [114]. To this end, events passing the trigger path requiring an isolated muon with p_{T} greater than 24 GeV in 2016 and 2018 and 27 GeV in 2017 are used. The events must also meet the MET requirements of the double-μ+MET trigger path. The isolated triggered muon is identified as the tag muon and the efficiency is measured as the fraction of the probe muons that pass the full muon selection of the double-μ+MET trigger path.

Specifically for the 2016 trigger path, each of the ϵ_{μ_i} terms also includes the DCA efficiency. Since the DCA efficiency is a property of the muon pair and not of each muon separately, it should be included in the calculation only once. As a result, to avoid double counting it, only for 2016, Eq. A.2 gets modified to:

$$\epsilon = \frac{\epsilon_{\mu_1}\epsilon_{\mu_2}}{\epsilon_{\mu,\mathrm{DCA}}} \times \epsilon_{\mathrm{MET}} \tag{A.3}$$

The ϵ_{μ_i} efficiency as a function of the muon p_{T} and η is shown for data and simulation in Fig. A.3 (2016), Fig. A.4 (2017) and Fig. A.5 (2018).

The DCA (2016) and Δz (2017 and 2018) efficiencies are computed in the following way: The events used in the denominator of the efficiency ratio must pass a double-μ trigger path with p_{T} thresholds at 17 and 8 GeV for the leading and subleading muon respectively. They are also required to pass the double-μ+MET L1 seed and the rest of the muon requirements of the double-μ+MET HLT path, such as the opposite charge and the invariant dimuon mass requirement. The events in the numerator must fulfil the same requirements as the denominator and additionally pass the DCA/Δz requirement. The result of the DCA/Δz efficiency measurement is shown in Figs. A.6 (2016), A.7 (2017) and A.8 (2018). The fact that only the lower right half of the plots is populated originates from the choice of ordering the muons in descending η order for this plot, that is $\eta(\mu_1) > \eta(\mu_2)$ always. It is also worth noting that some bins, those in which the muons have a very large separation in η, i.e. $\eta(\mu_1) - \eta(\mu_2) \gtrsim 2.5$, are empty. This is a physical restriction coming from the kinematical characteristics of the muons probed: Muons with moderate p_{T} (at least 17 or 8 GeV, as per trigger requirements), fulfilling the invariant dimuon mass requirement of $4 < M(\ell\ell) < 56$ GeV do not give such large η separations. However, the softer muons used in the analysis can in fact end up in these corners of the $\eta(\mu_1)$-$\eta(\mu_2)$ phase space. To account for this, the bins for which the relation $\eta(\mu_1) - \eta(\mu_2) > 2.5$ is true are filled with the average value of efficiency calculated over the bins which satisfy $\eta(\mu_1) - \eta(\mu_2) \leq 2.5$. This is a valid extrapolation, given the small variation of efficiency in the bins where the measurement can be performed with good statistical power. In fact, the variation is so small in the Δz efficiency of 2017 and 2018 that the average value of the efficiency is used even in bins with $\eta(\mu_1) - \eta(\mu_2) \leq 2.5$. The average values of DCA/Δz efficiency extracted from the efficiency maps of Figs. A.6-A.8 are given in Table A.1.

The hadronic part of the efficiency is measured in a similar way. The

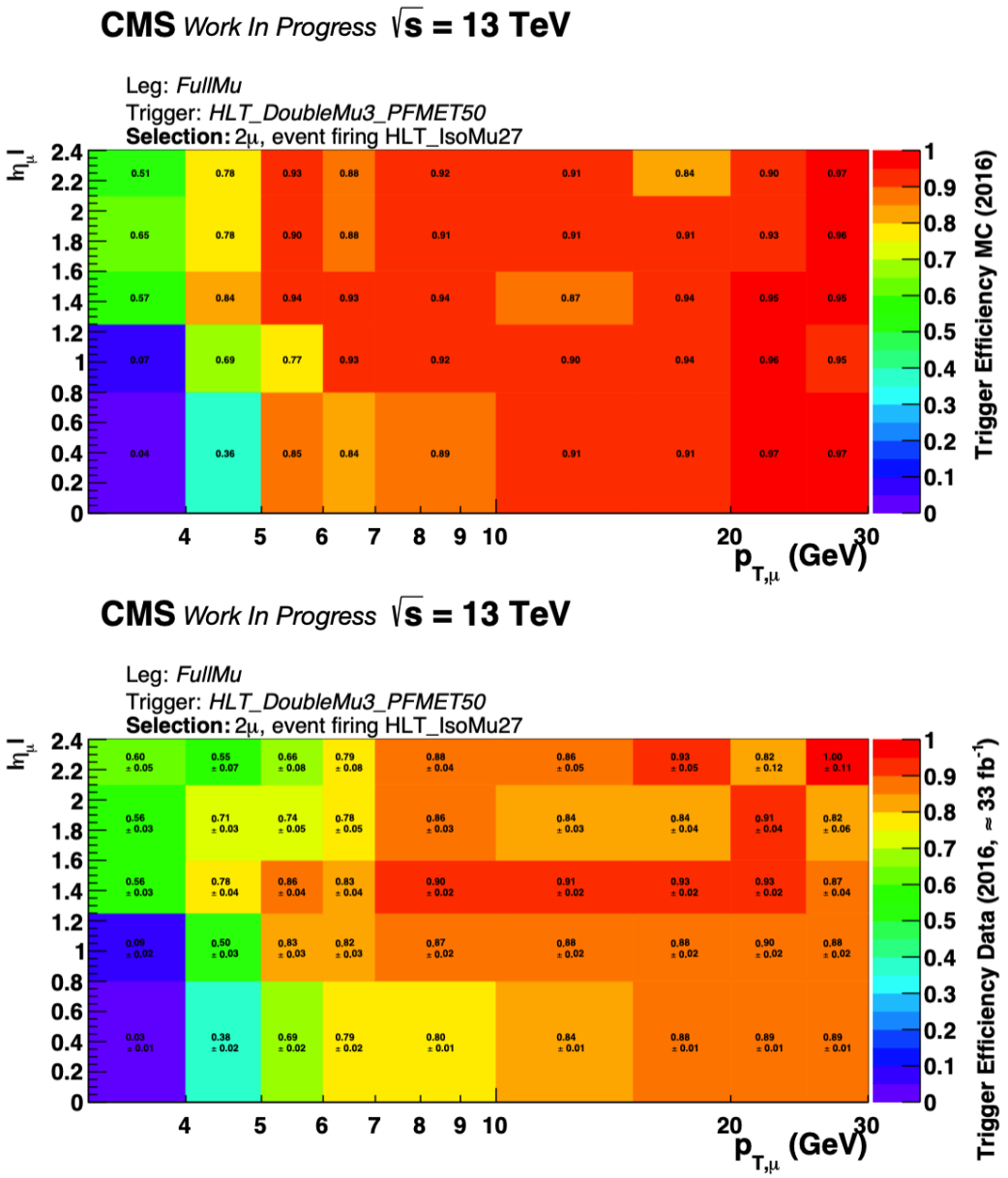


Figure A.3: Trigger efficiency of the leptonic requirements for the double-μ+MET trigger path of 2016 as a function of the offline muon p_{T} and η measured in both data (top) and simulation (bottom).

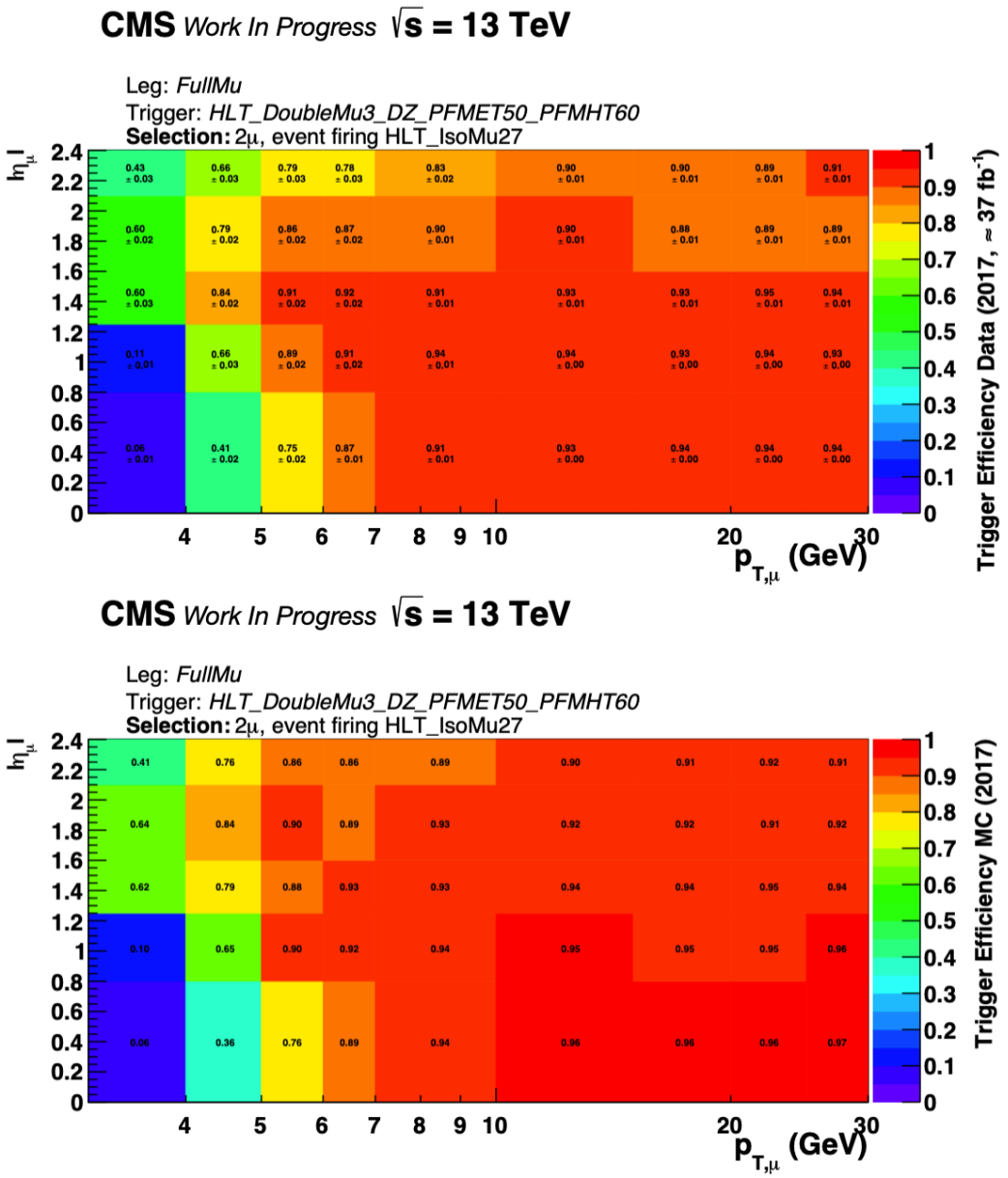


Figure A.4: Trigger efficiency of the leptonic requirements for the double-μ+MET trigger path of 2017 as a function of the offline muon p_{T} and η measured in both data (top) and simulation (bottom).

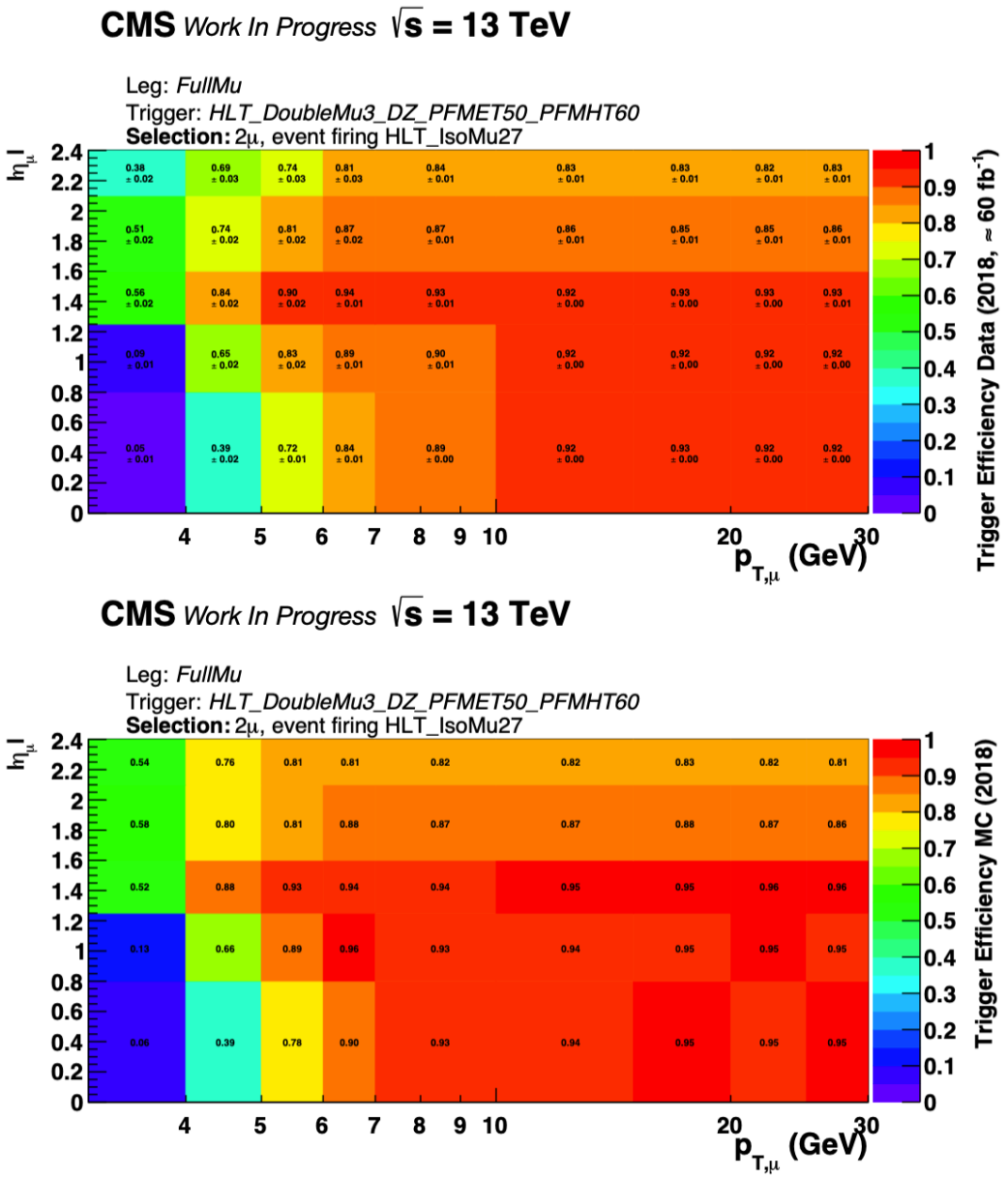


Figure A.5: Trigger efficiency of the leptonic requirements for the double-μ+MET trigger path of 2018 as a function of the offline muon p_{T} and η measured in both data (top) and simulation (bottom).

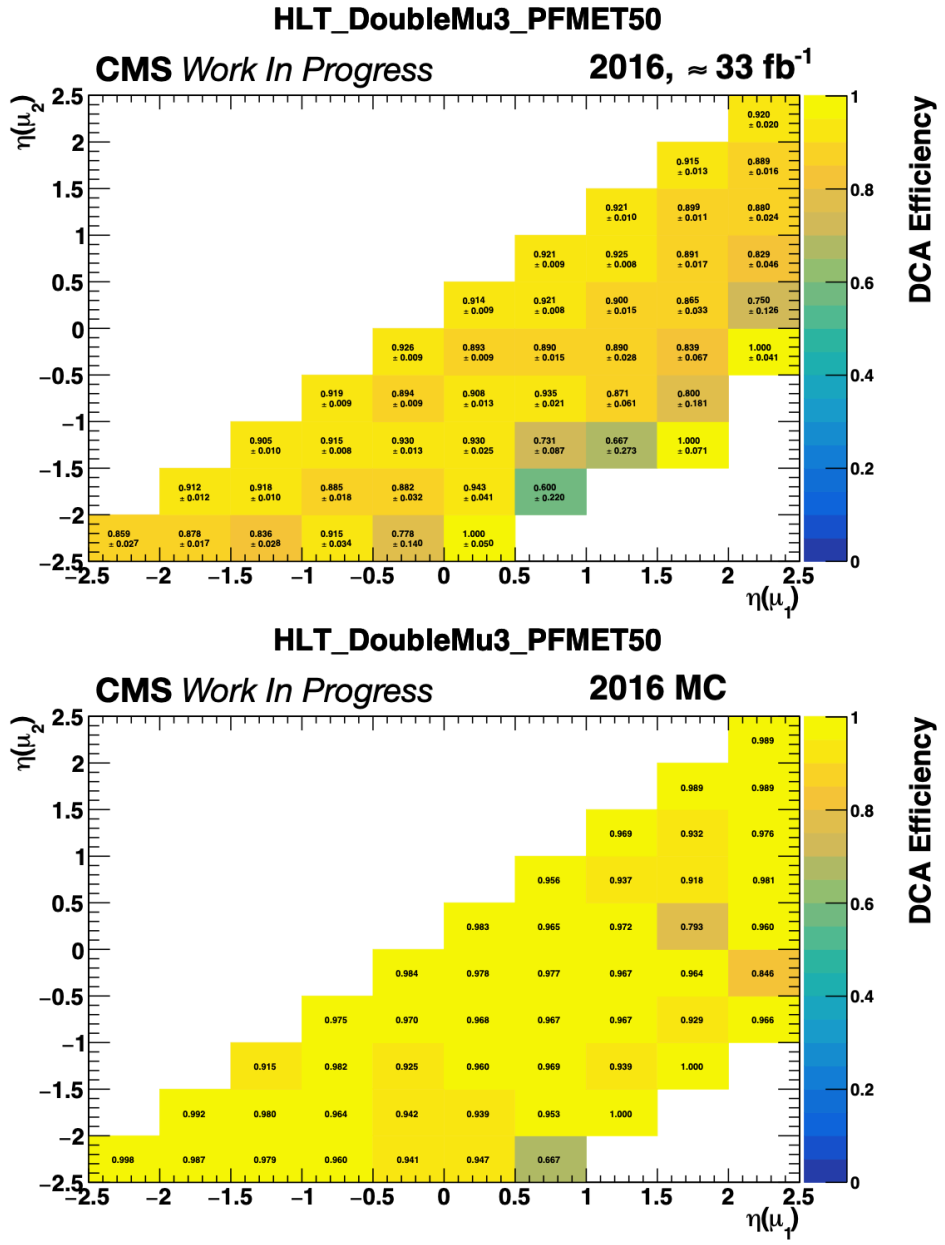


Figure A.6: Trigger efficiency of the DCA requirement as a function of the offline η of the muon with the highest η (μ_1) and the lowest η (μ_2) measured in both data (top) and simulation (bottom) for the double-μ+MET trigger path of 2016.

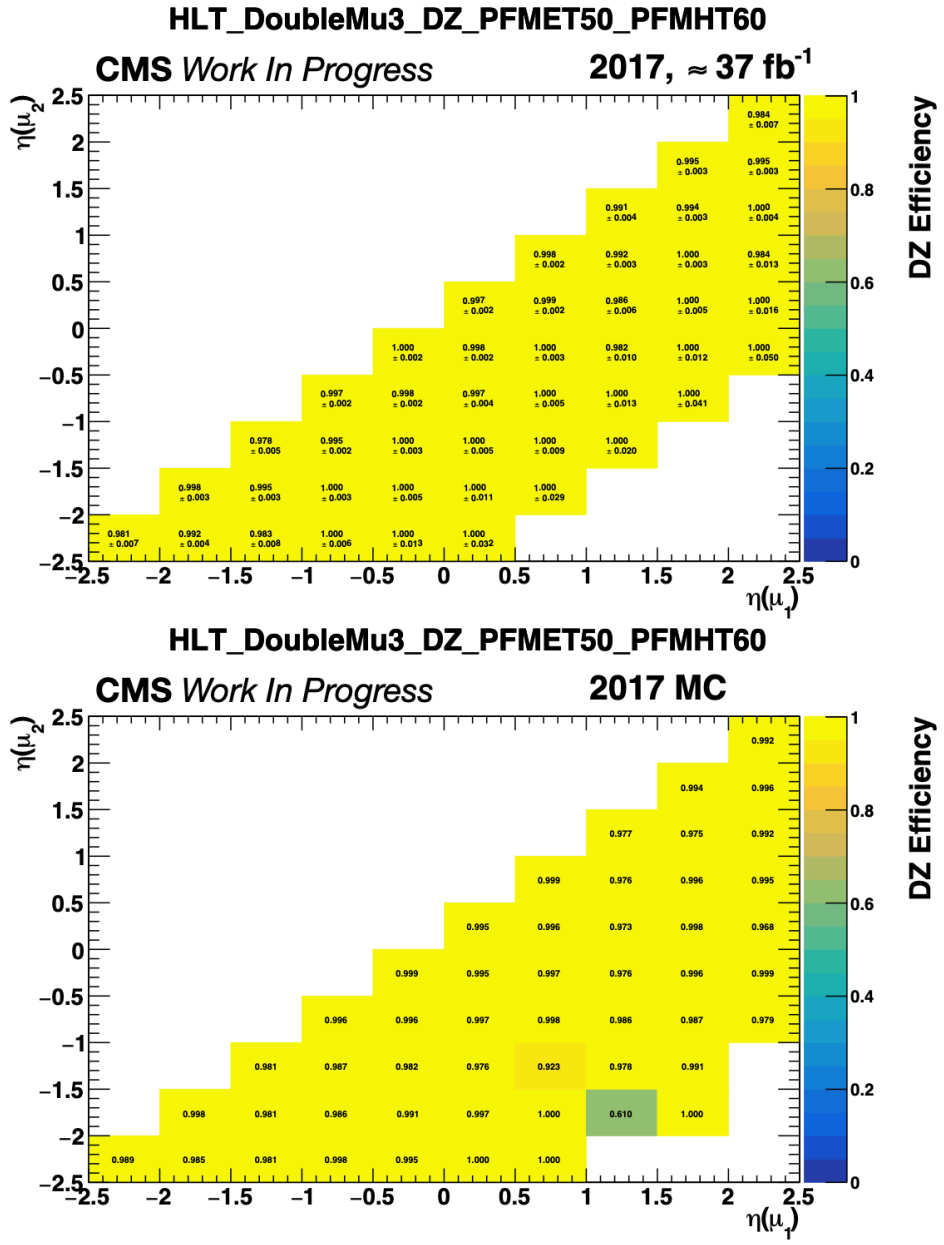


Figure A.7: Trigger efficiency of the Δz requirement as a function of the offline η of the muon with the highest η (μ_1) and the lowest η (μ_2) measured in both data (top) and simulation (bottom) for the double-μ+MET trigger path of 2017.

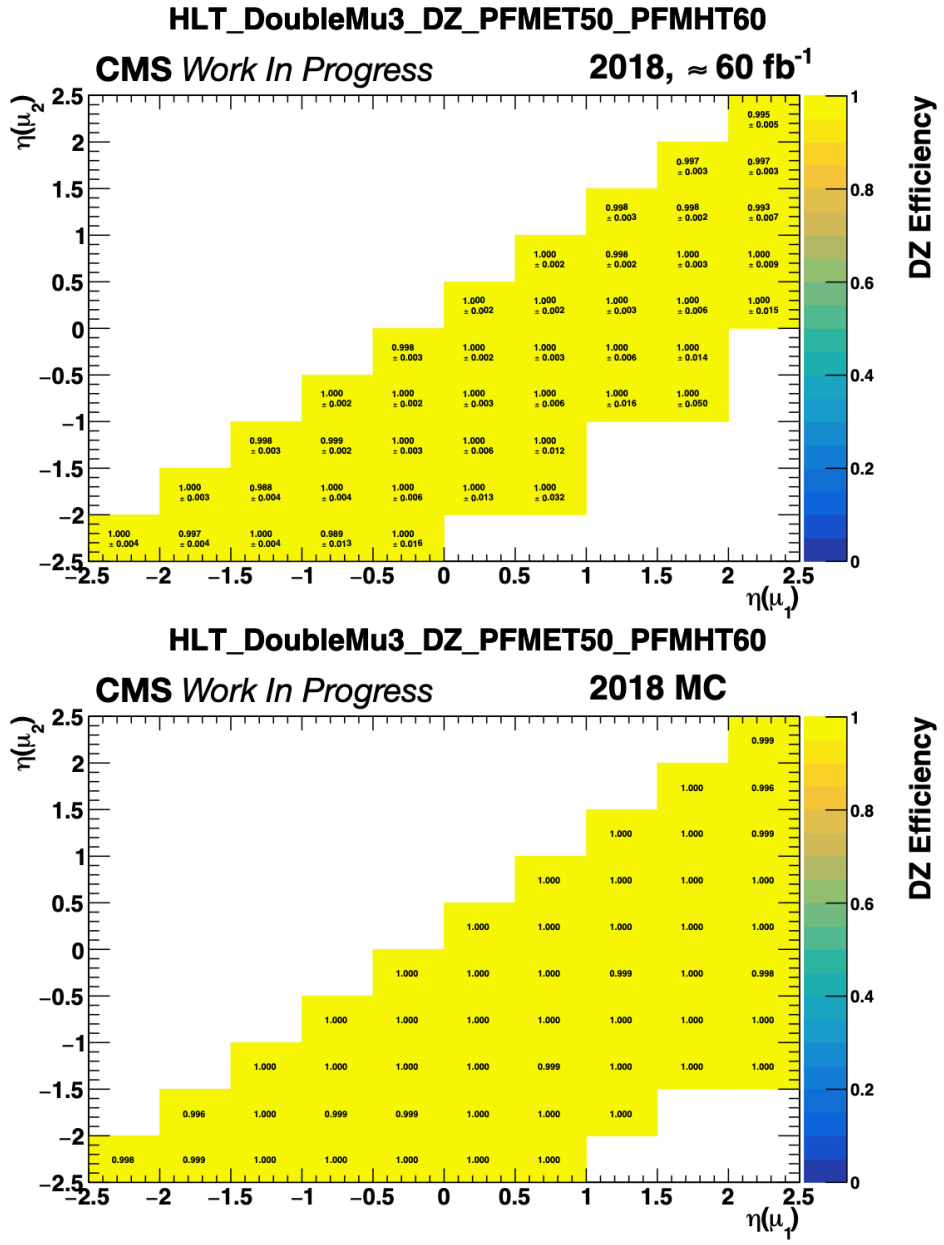


Figure A.8: Trigger efficiency of the Δz requirement as a function of the offline η of the muon with the highest η (μ_1) and the lowest η (μ_2) measured in both data (top) and simulation (bottom) for the double-μ+MET trigger path of 2018.

Year	$< \epsilon_{\mu,\mathrm{DCA}} >$	$< \epsilon_{\mu,\Delta z} >$
2016 (Data)	90.6%	-
2016 (MC)	96.2%	-
2017 (Data)	-	99.5%
2017 (MC)	-	99.8%
2018 (Data)	-	99.0%
2018 (MC)	-	99.9%

Table A.1: Average values of the DCA and Δz efficiency for the double-μ+MET trigger path of each year, in data and simulation.

events used in the measurement are required to have fired a double-μ trigger path with $p_{\mathrm{T}}(\mu_1) > 17\,\mathrm{GeV}$ and $p_{\mathrm{T}}(\mu_2) > 8\,\mathrm{GeV}$. To perform the measurement with as soft muons as possible while also retaining adequate statistical power, muons with $10 < p_{\mathrm{T}}(\mu) < 50\,\mathrm{GeV}$ are used. This, in combination with charge and invariant dimuon mass requirements, ensures that the leptonic part of the double-μ+MET trigger path is satisfied and that the DCA/Δz component is $\sim$100% efficient. This selection defines the efficiency ratio denominator, while the extra requirement of passing the double-μ+MET trigger path is set for the numerator. In this way, both the $p_{\mathrm{T}}^{\mathrm{miss}}$ and the $p_{\mathrm{T}}^{\mathrm{miss,corr}}$ parts are measured simultaneously, as shown in Figs. A.9 (2016), A.10 (2017) and A.11 (2018). It can be noticed that the efficiency plateau is achieved for $p_{\mathrm{T}}^{\mathrm{miss}} > 125\,\mathrm{GeV}$ and $p_{\mathrm{T}}^{\mathrm{miss,corr}} > 125\,\mathrm{GeV}$. In addition, empty bins are again present at the edges of the plots. This originates from the fact that $p_{\mathrm{T}}^{\mathrm{miss}}$ and $p_{\mathrm{T}}^{\mathrm{miss,corr}}$ are strongly correlated when the muons are relatively soft and, as a result, the phase space where these two variables have large difference ($\gtrsim 100\,\mathrm{GeV}$ in this case) is sparsely populated. These bins are filled with values extrapolated from neighbouring bins.

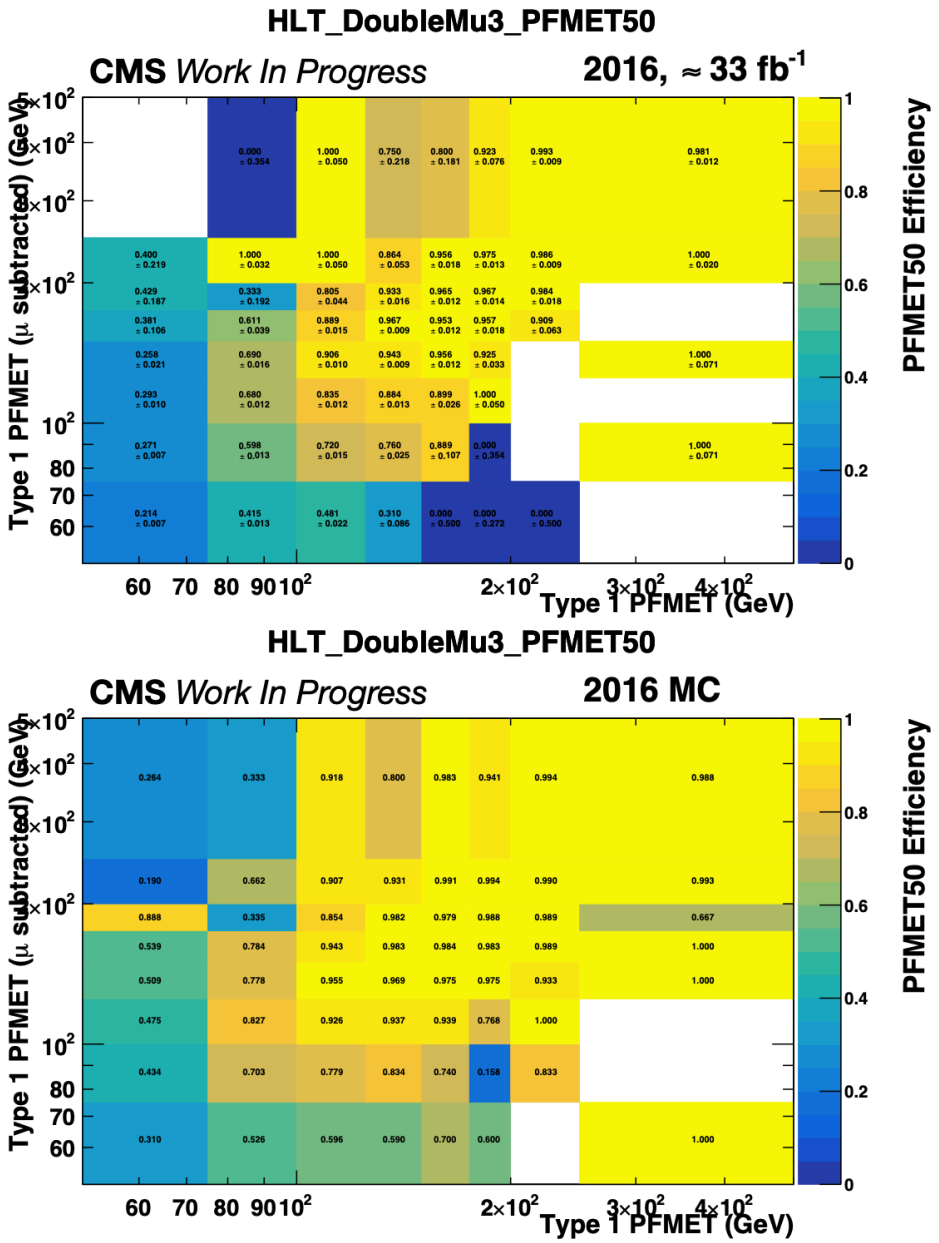


Figure A.9: Trigger efficiency of the MET requirements for the double-μ+MET trigger path of 2016 as a function of the offline $p_{\mathrm{T}}^{\mathrm{miss}}$ and $p_{\mathrm{T}}^{\mathrm{miss,corr}}$ measured in both data (top) and simulation (bottom).

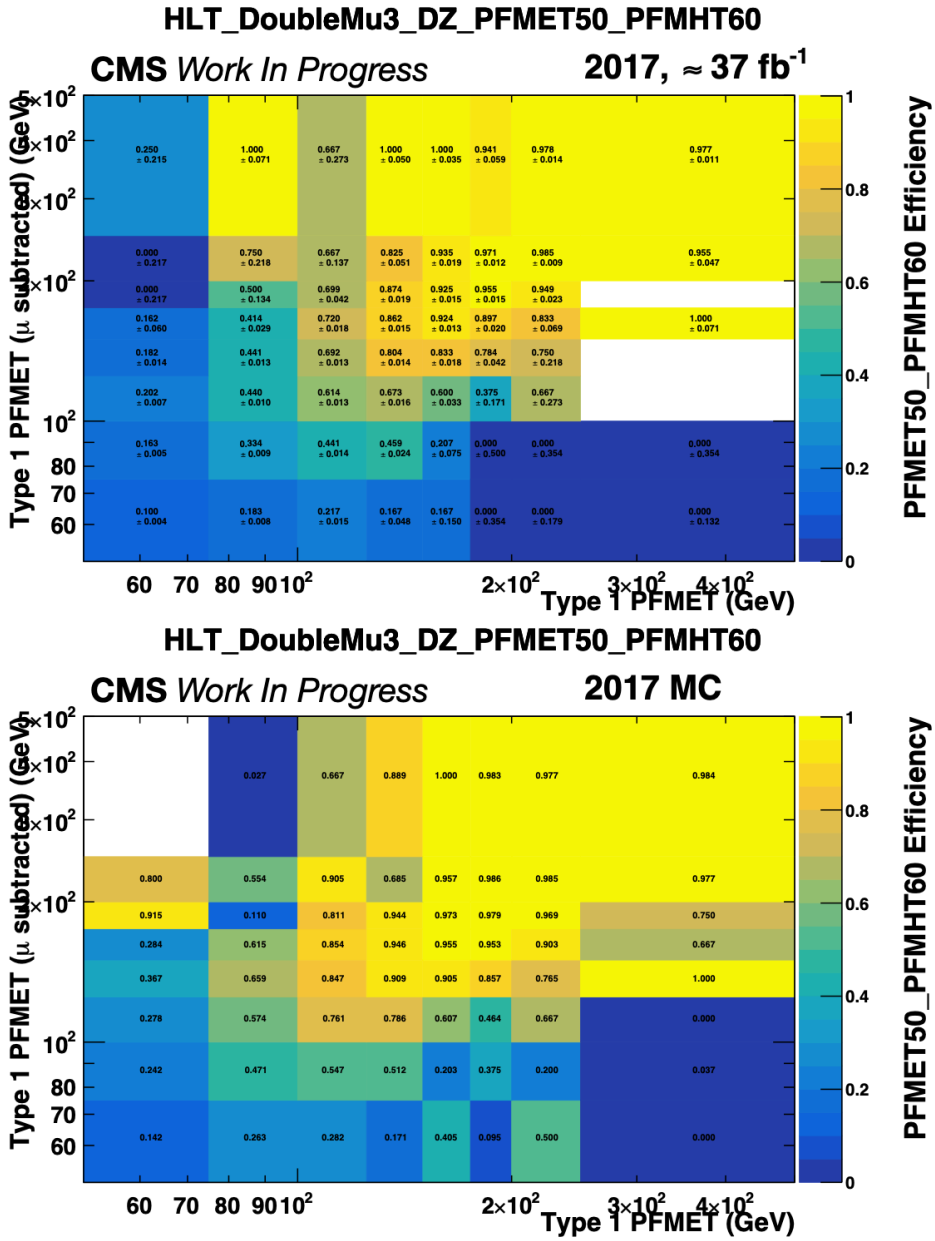


Figure A.10: Trigger efficiency of the MET requirements for the double-μ+MET trigger path of 2017 as a function of the offline $p_{\mathrm{T}}^{\mathrm{miss}}$ and $p_{\mathrm{T}}^{\mathrm{miss,corr}}$ measured in both data (top) and simulation (bottom).

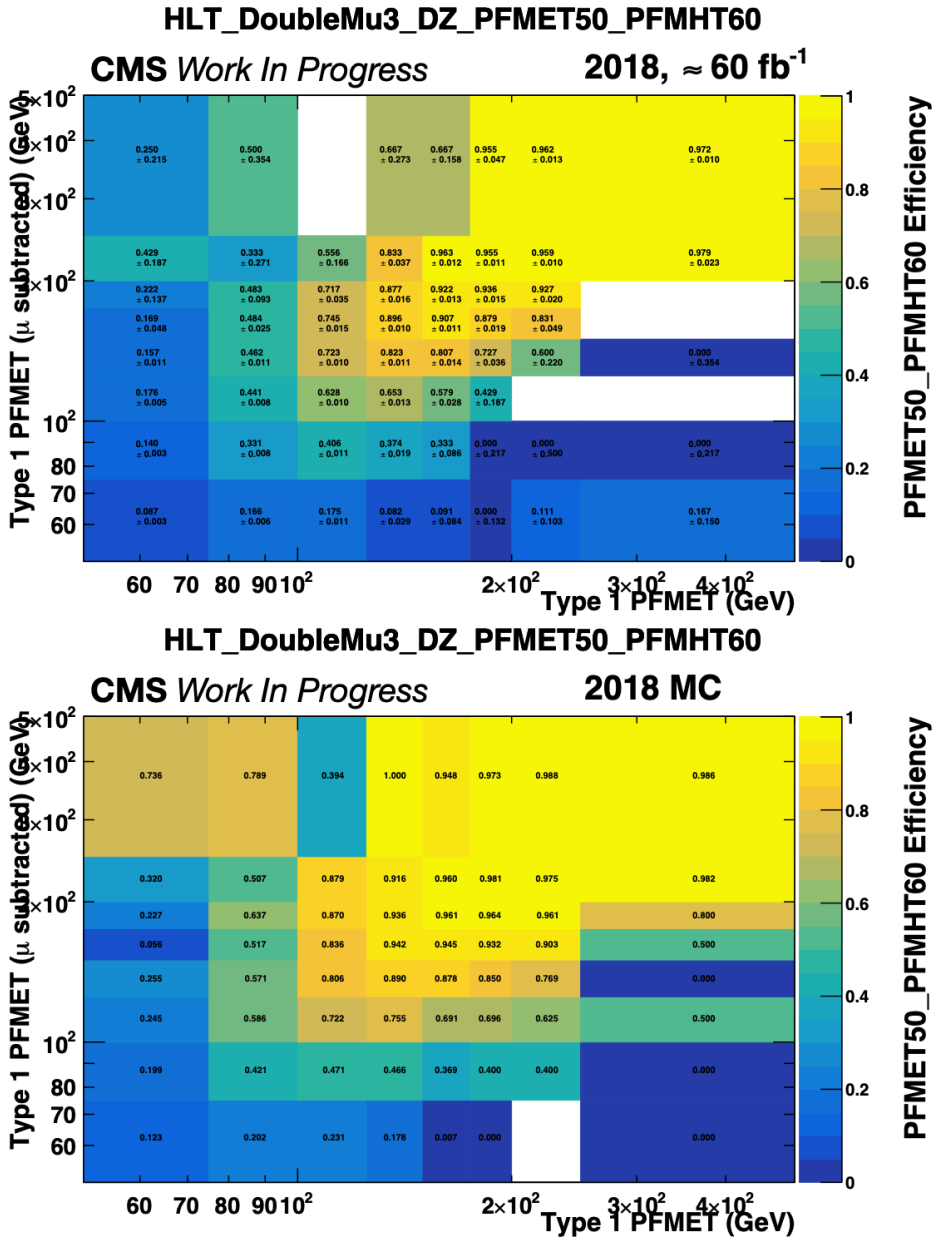


Figure A.11: Trigger efficiency of the MET requirements for the double-μ+MET trigger path of 2018 as a function of the offline $p_{\mathrm{T}}^{\mathrm{miss}}$ and $p_{\mathrm{T}}^{\mathrm{miss,corr}}$ measured in both data (top) and simulation (bottom).

Appendix B

Extra Material on the Nonprompt Background Estimation

This appendix is dedicated to providing extra material for the estimation of the nonprompt background. Figs. B.1-B.3 show the FR measurement of muons (for the two different muon trigger paths used separately) and electrons in η bins, as a function of lepton p_{T}. The different lines on the plots correspond to the different methods used for the FR measurement: "QCD MC" uses QCD simulation, "QCD MC, cut" uses QCD simulation with the requirement $M_{\mathrm{T}}^{\mathrm{fix}}(\ell, p_{\mathrm{T}}^{\mathrm{miss}}) < 20\,\mathrm{GeV}$, while the rest of the lines use the three different methods, as these were defined in Sec. 4.4.1, on data. The simulation FR measurements are used as a reference and it is only the data-driven FR measurements that are used in the analysis.

Figs. B.4-B.6 show the FR, measured in simulation, for muons (for the two different muon trigger paths used separately) and electrons originating from bottom, charm or lighter quarks, in the different η bins, as a function of the lepton p_{T}. The FR for leptons coming from the decay of different flavor quarks has the same general behavior after tuning the analysis loose-not-tight ID, as mentioned in Sec. 4.4.1.

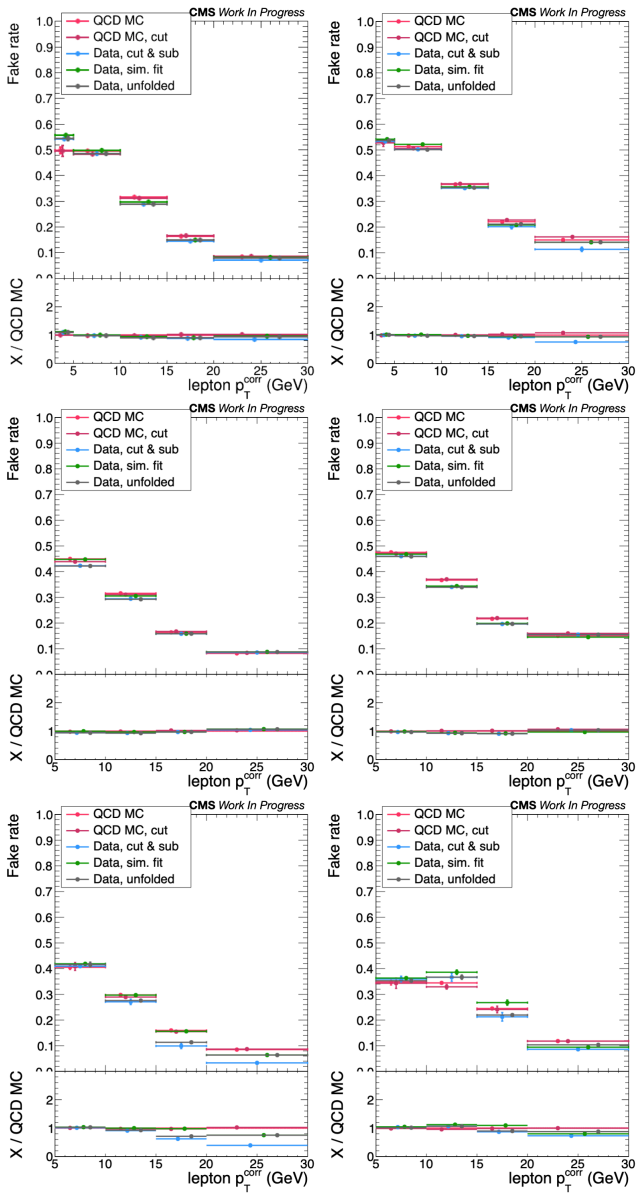


Figure B.1: The FR of muons of the `HLT_Mu3_PFJet40` (upper row) and the `HLT_Mu8` trigger path (middle row) and of electrons (lower row) in the barrel (left) and endcap (right) in 2016, as a function of the lepton p_{T}, using the three different methods mentioned in the text.

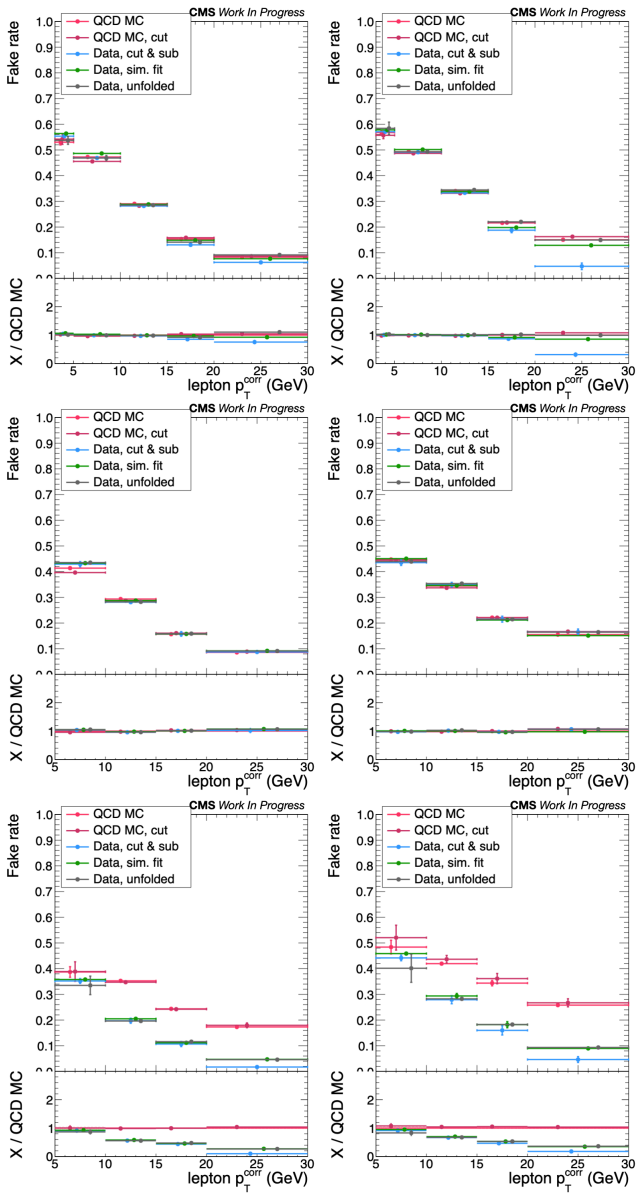


Figure B.2: The FR of muons of the `HLT_Mu3_PFJet40` (upper row) and the `HLT_Mu8` trigger path (middle row) and of electrons (lower row) in the barrel (left) and endcap (right) in 2017, as a function of the lepton p_{T}, using the three different methods mentioned in the text.

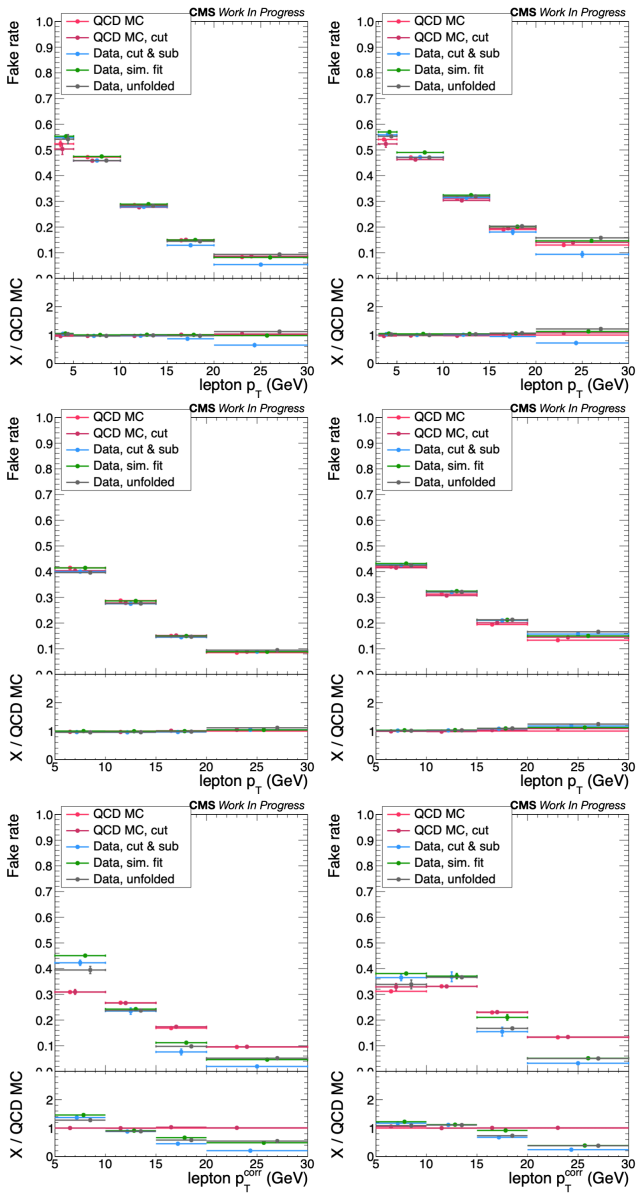


Figure B.3: The FR of muons of the `HLT_Mu3_PFJet40` (upper row) and the `HLT_Mu8` trigger path (middle row) and of electrons (lower row) in the barrel (left) and endcap (right) in 2018, as a function of the lepton p_{T}, using the three different methods mentioned in the text.

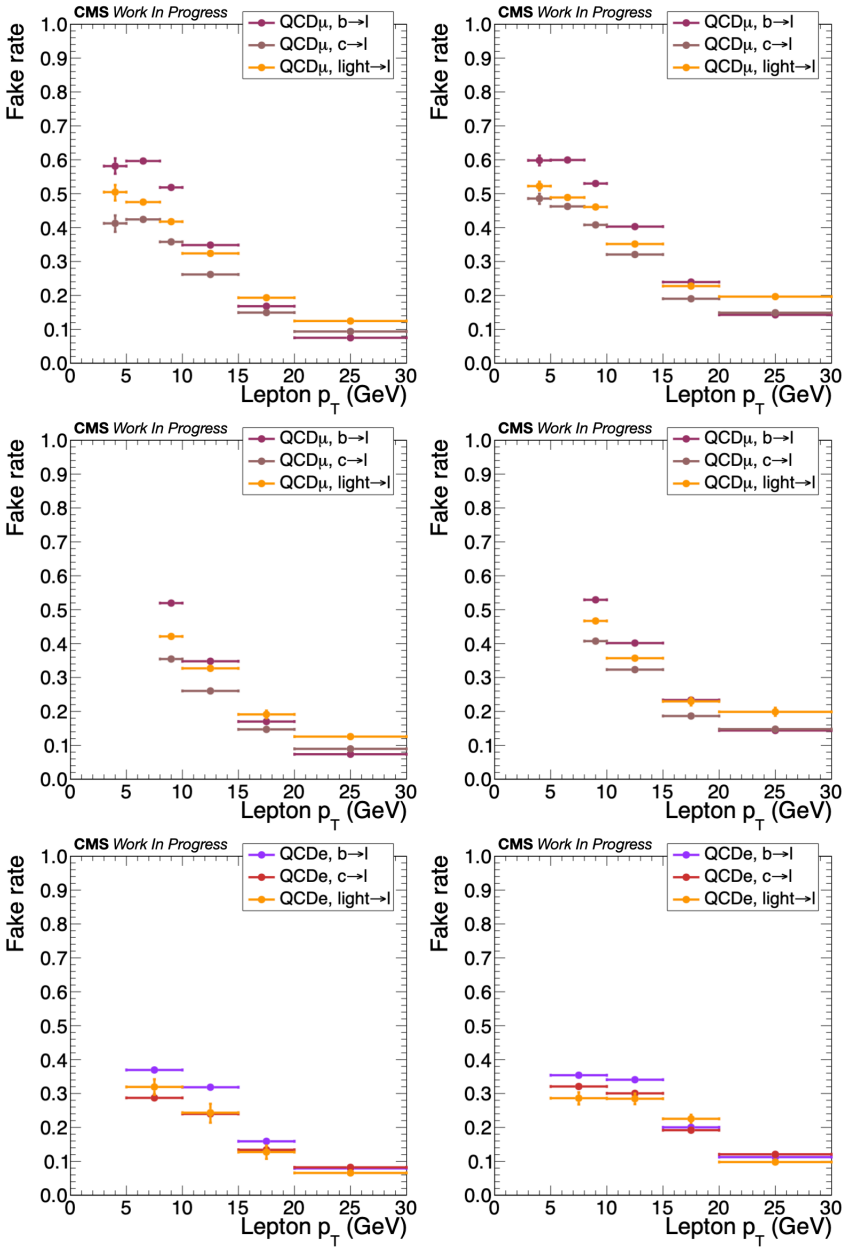


Figure B.4: The FR of muons of the `HLT_Mu3_PFJet40` (upper row) and the `HLT_Mu8` trigger path (middle row) and of electrons (lower row) in the barrel (left) and endcap (right) in 2016, as a function of the lepton p_{T}, for leptons originating from different quark flavors in simulation.

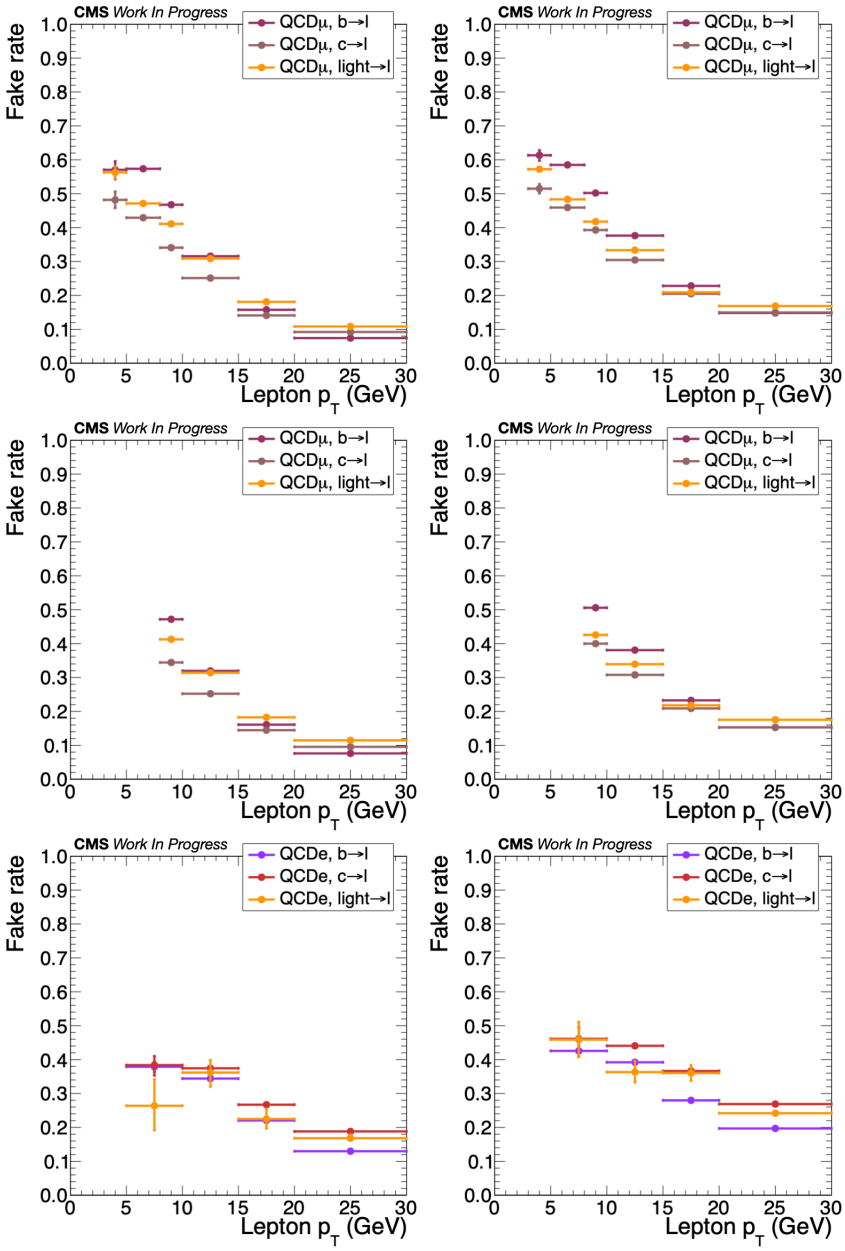


Figure B.5: The FR of muons of the `HLT_Mu3_PFJet40` (upper row) and the `HLT_Mu8` trigger path (middle row) and of electrons (lower row) in the barrel (left) and endcap (right) in 2017, as a function of the lepton p_{T}, for leptons originating from different quark flavors in simulation.

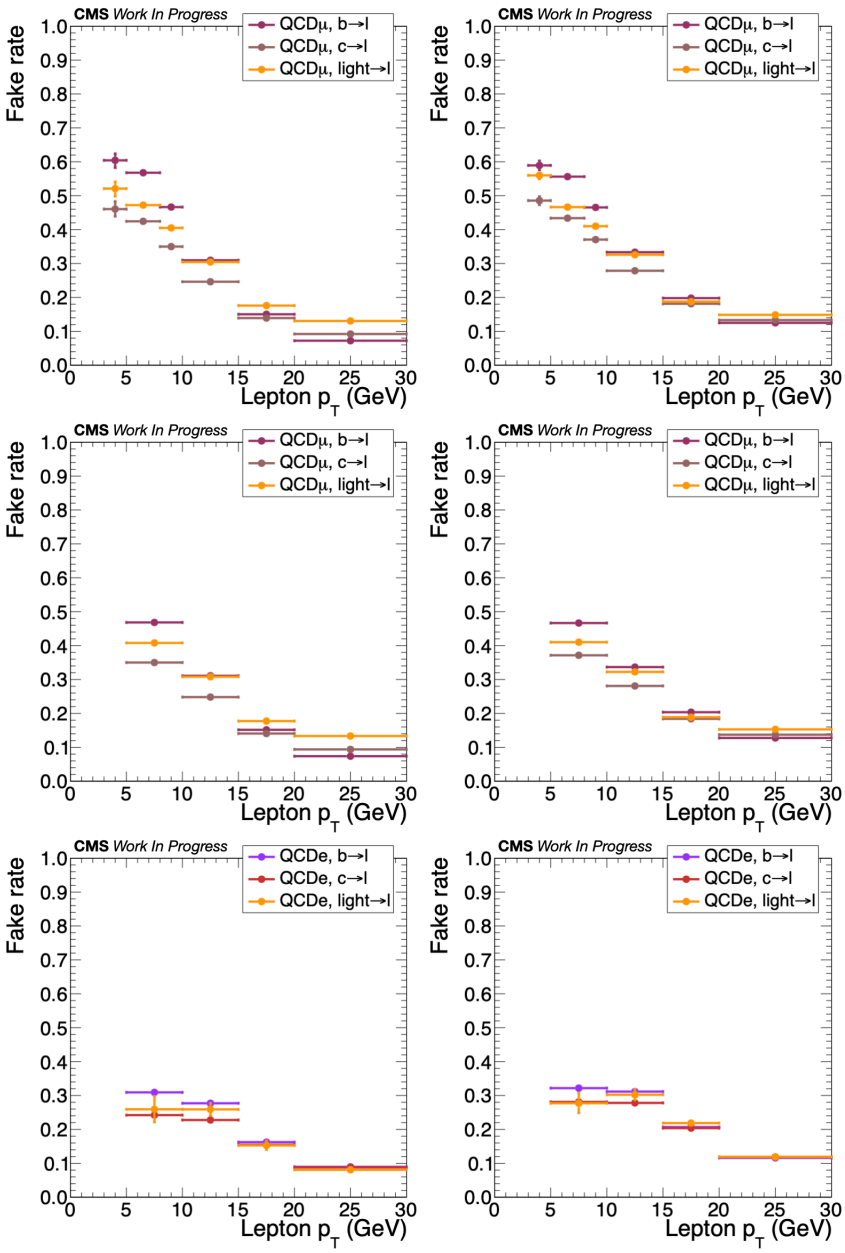


Figure B.6: The FR of muons of the `HLT_Mu3_PFJet40` (upper row) and the `HLT_Mu8` trigger path (middle row) and of electrons (lower row) in the barrel (left) and endcap (right) in 2018, as a function of the lepton p_{T}, for leptons originating from different quark flavors in simulation.

Appendix C

Some Technical Details of the MLP Implementation

This appendix is dedicated to providing some extra details on the technical implementation of the MLP for the ϕ correction of muons in 40 MHz muon scouting system. The MLP has been developed, trained and tested with the use of the TMVA toolkit [138], included in the ROOT analysis framework [140].

The MLP consists 1 input layer with 4 input variables, 1 hidden layer with 24 neurons and 1 output layer. The neuron response function of a layer is chosen to be the weighted sum of the neuron response of the previous layer, activated with a tanh function:

$$y_i^l = w_{i0}^{l-1} + \sum_{j=1}^{n} \tanh(y_j^{l-1}) w_{ij}, \text{ for } l > 1 \tag{C.1}$$

where l corresponds to the layer ($l = 1$ for the input layer) and n is the number of neurons in the lth layer (i.e. $n = 24$ for the hidden layer and $n = 1$ for the output layer of this MLP). The back-propagation of the weights is performed with the Broyden-Fletcher-Goldfarb-Shannon (BFGS) method [141–144].

The number of training cycles is set to 20000. However, a convergence test is performed at the end of each cycle and, if the improvement of the MLP response estimator, a mean square error estimator in this case, is not greater than 10^{-6} for 15 consecutive cycle, the training is stopped early. Tests for overtraining are performed every 6 cycles.

A sampling technique is employed for the training. For the 80% of the total cycles, i.e. 16000 cycles, only the 30% of the total events, randomly selected, are used for the training. In the last 20% of the cycles, all the events are included in the training. This way, the training is significantly sped-up. Sampling has been also shown to improve the robustness of the minimizing algorithm used in the MLP [138].

Appendix D

Definitions of the HGCal Variables

This appendix is dedicated in defining the variables provided by the HGCal subsystem and were used as input to the PU and Pion ID BDTs. These variables can be broadly split in four categories:

- Generic variables:
  - p_{T}
  - $|\eta|$
  - ϕ
- Energy-related variables:
  - hOverE: $E_{\mathrm{had}}/E_{\mathrm{EM}}$.
  - eMax: Energy of layer with maximum energy.
  - eMaxOverE: eMax divided by the total cluster energy.
  - zBarycenter: $\langle z \rangle_{\mathrm{tot}}$
- Longitudinal shape variables:
  - firstLayer: The first layer of the cluster with nonzero energy.
  - maxLayer: The layer with the maximum energy.
  - showerLength: The length of the shower, in layers.

  - coreShowerLength: Maximum number of consecutive layers with nonzero energy.
  - sigmaZZ: σ_{zz}
  - layer$X$percent: Layer up to which $X\%$ of the total cluster energy is contained.
  - triggerCells$X$percent: Number of trigger cells that contain $X\%$ of the total cluster energy.

- Transverse shape variables:
  - sigmaEtaEtaMax: $\sigma_{\eta\eta,\text{max}}$
  - sigmaEtaEtaTot: $\sigma_{\eta\eta,\text{tot}}$
  - sigmaPhiPhiMax: $\sigma_{\phi\phi,\text{max}}$
  - sigmaPhiPhiTot: $\sigma_{\phi\phi,\text{tot}}$
  - sigmaRRMax: $\sigma_{RR,\text{max}}$
  - sigmaRRTot: $\sigma_{RR,\text{tot}}$
  - sigmaRRMean: $\sigma_{RR,\text{mean}}$

In the above list, the following definition apply:

$$\langle X \rangle_Y = \frac{\sum_{i \in \text{TC}}^{Y} E_i X_i}{\sum_{i \in \text{TC}}^{Y} E_i} \tag{D.1}$$

where X is any variable, TC indicates that the energy of separate trigger cells is summed, while Y could be either "2D" or "tot" if the sum is over single 2D layers or the total cluster, respectively. Consequently, $\langle X \rangle_Y$ is the energy-weighted mean of the variable X of the Y cluster. Similarly:

$$\sigma_{XX,Y} = \sqrt{\frac{\sum_{i \in \text{TC}}^{Y} E_i \left(X_i - \langle X \rangle_Y\right)^2}{\sum_{i \in \text{TC}}^{Y} E_i}} \tag{D.2}$$

$$\sigma_{XX,\text{max}} = \max_{\text{2D layers}} \left(\sqrt{\frac{\sum_{i \in \text{TC}}^{2D} E_i \left(X_i - \langle X \rangle_{\text{2D}}\right)^2}{\sum_{i \in \text{TC}}^{2D} E_i}} \right) \tag{D.3}$$

$$\sigma_{XX,\text{mean}} = \langle \sigma_{XX,\text{tot}} \rangle_{\text{tot}} \tag{D.4}$$

where, in the internal sum, only TCs within 5 cm of the TC with maximum energy are taken into account.

Appendix E

Deregionizer Logic Snapshots

This appendix is dedicated in explaining the logic of the deregionizer module by means of step-by-step snapshot of processing an event (Figs. E.1–E.30):

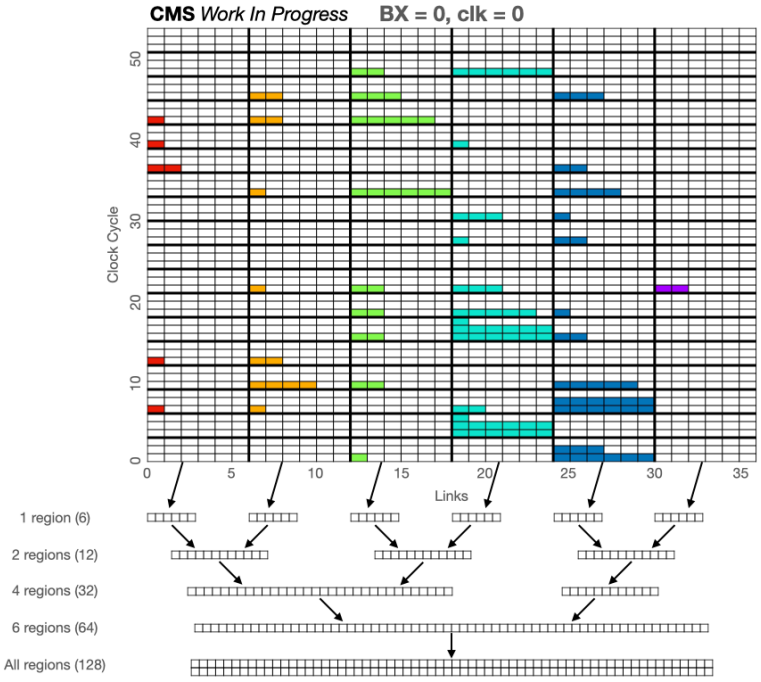


Figure E.1: Deregionizer snapshot, clock 0

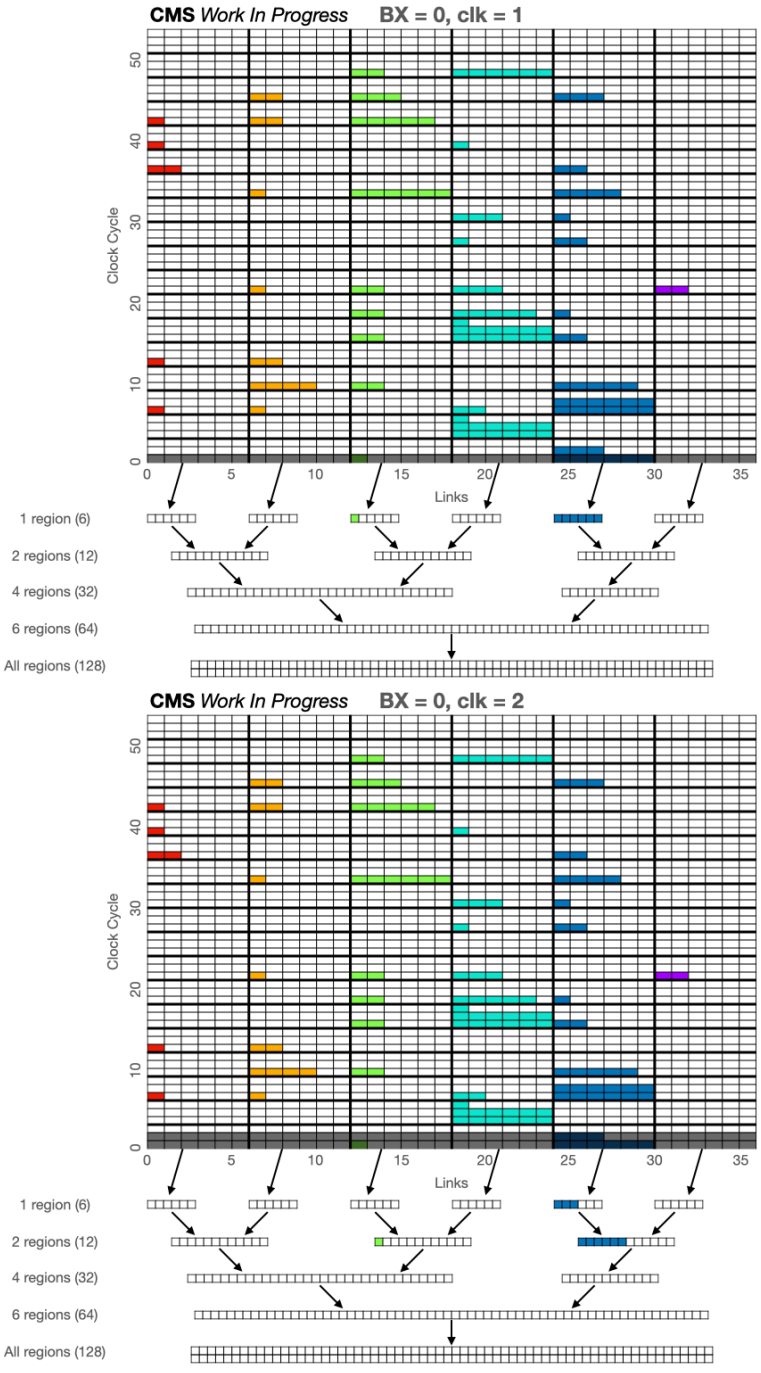


Figure E.2: Deregionizer snapshots, clocks 1–2

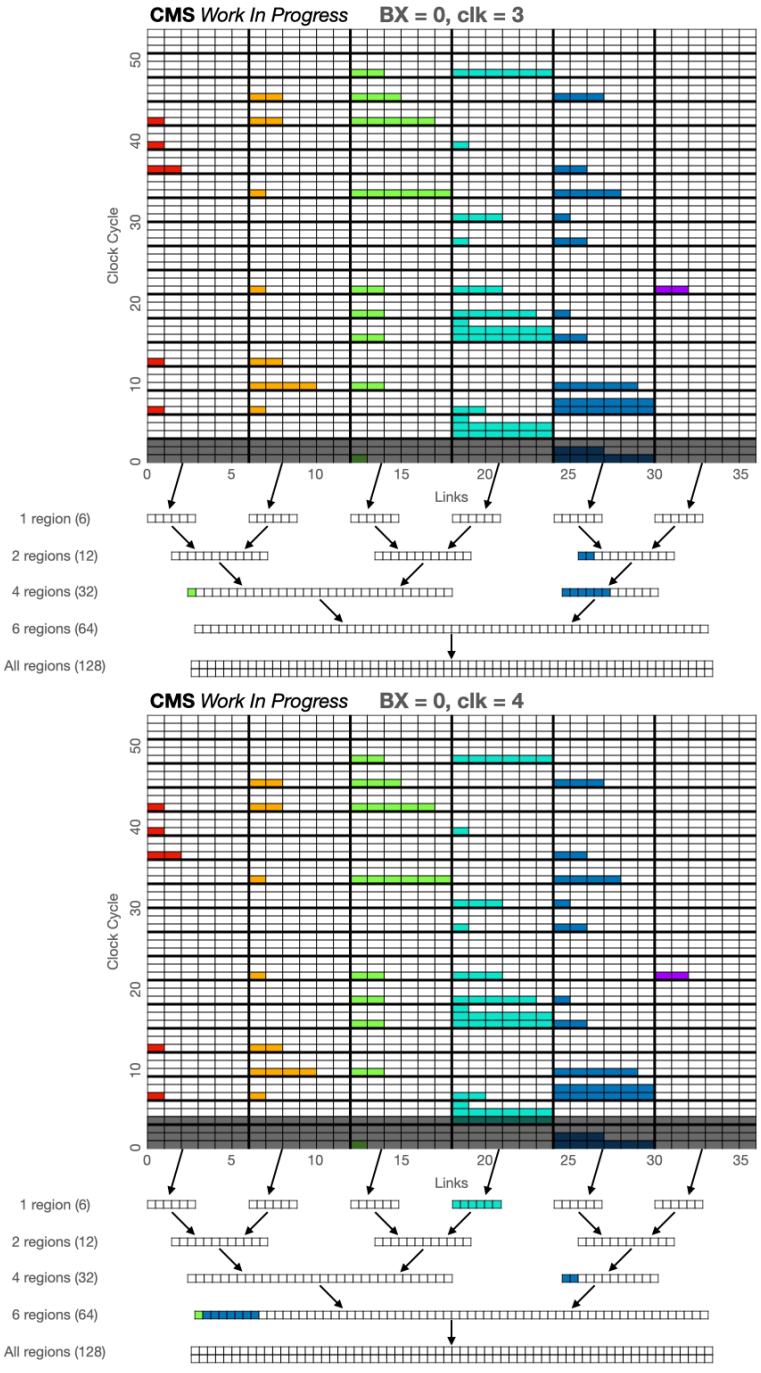


Figure E.3: Deregionizer snapshots, clocks 3–4

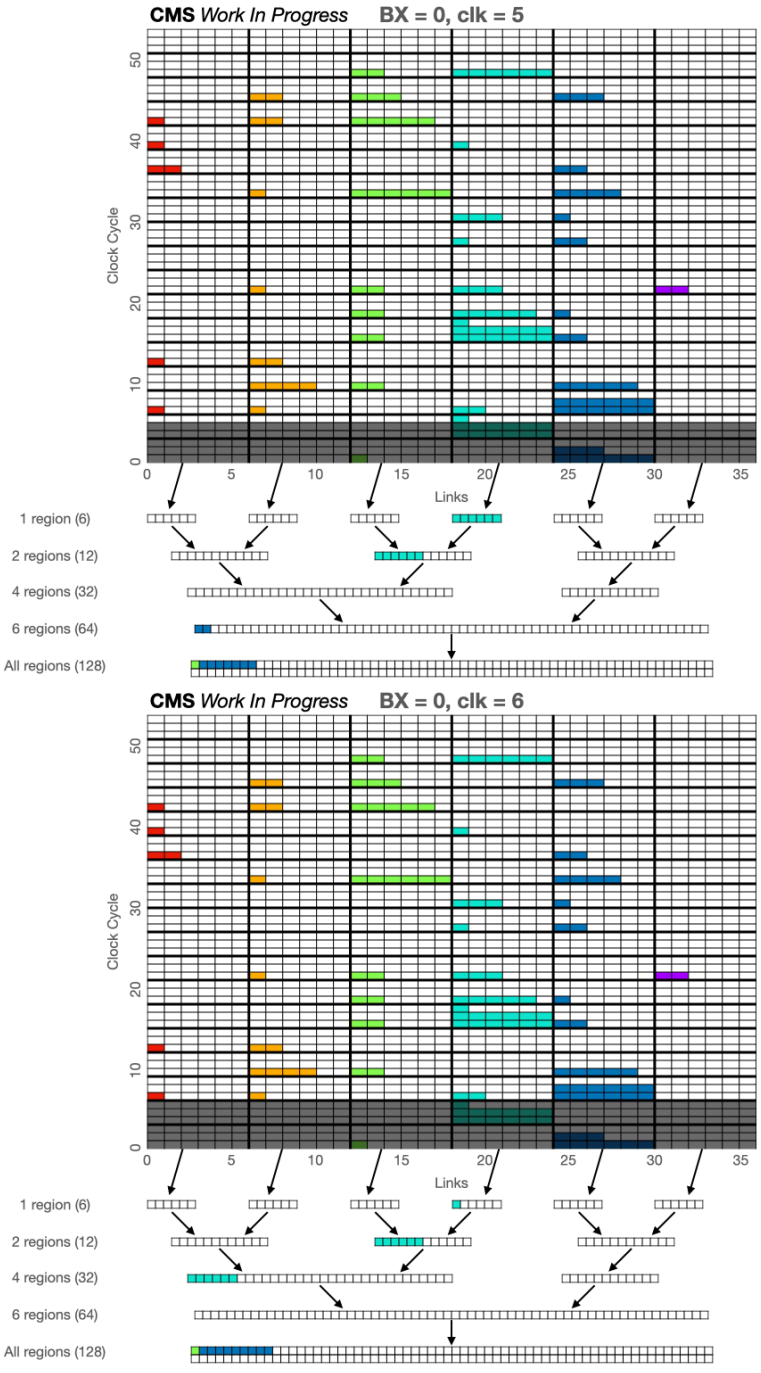


Figure E.4: Deregionizer snapshots, clocks 5–6

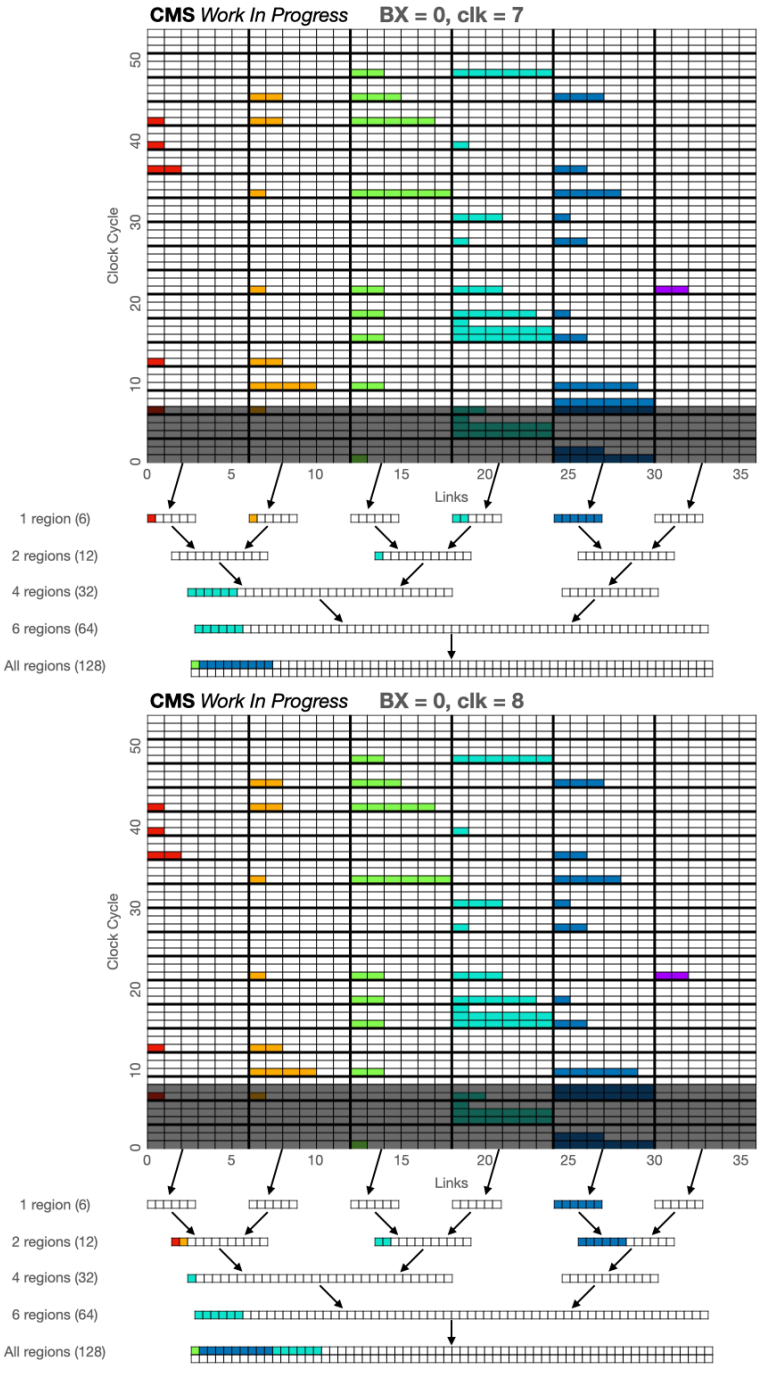


Figure E.5: Deregionizer snapshots, clocks 7–8

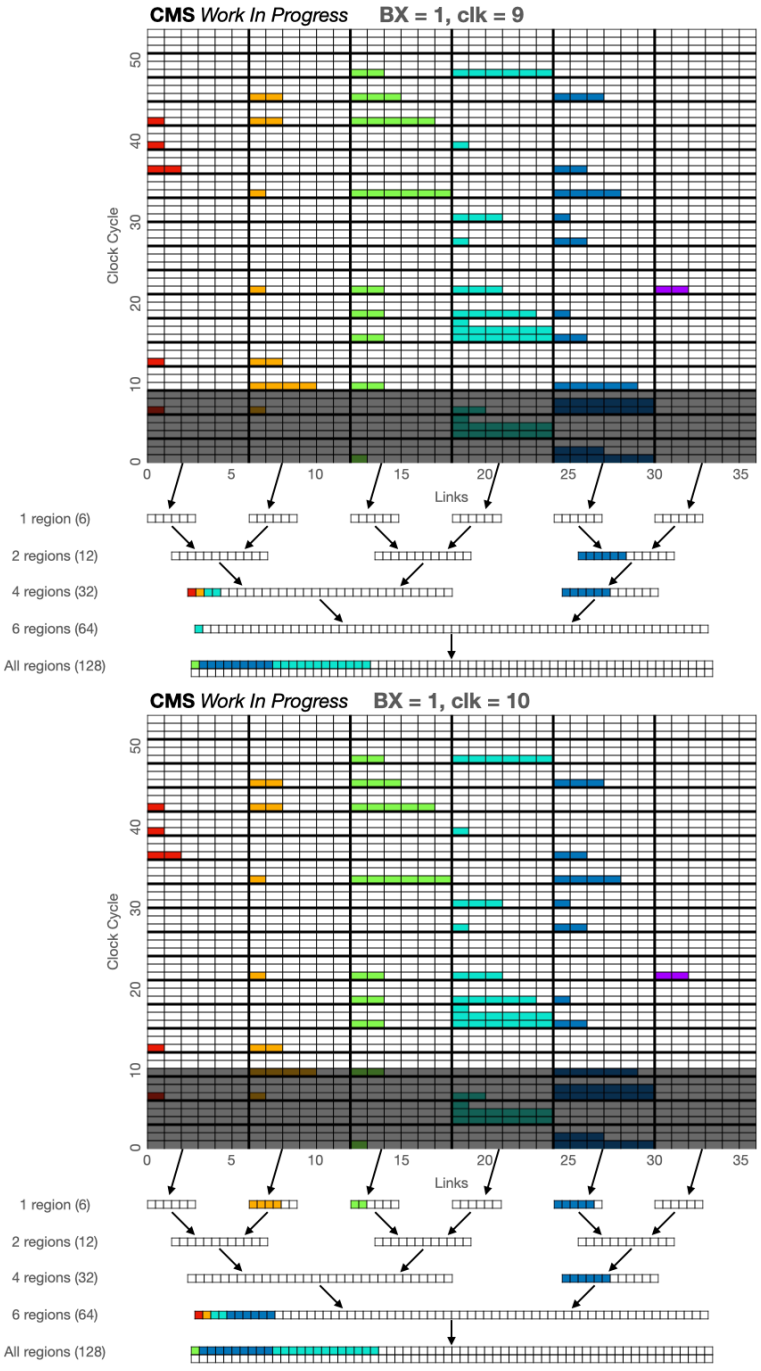


Figure E.6: Deregionizer snapshots, clocks 9–10

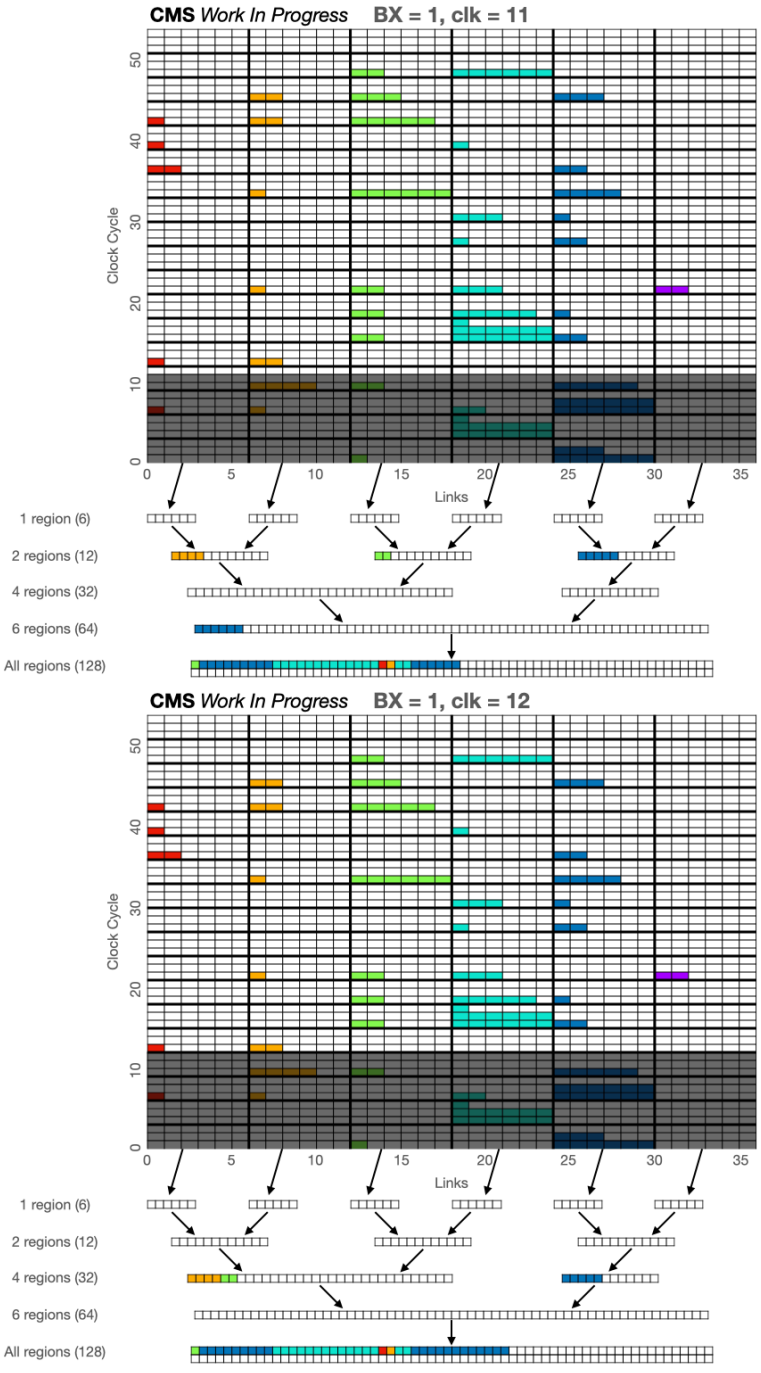


Figure E.7: Deregionizer snapshots, clocks 11–12

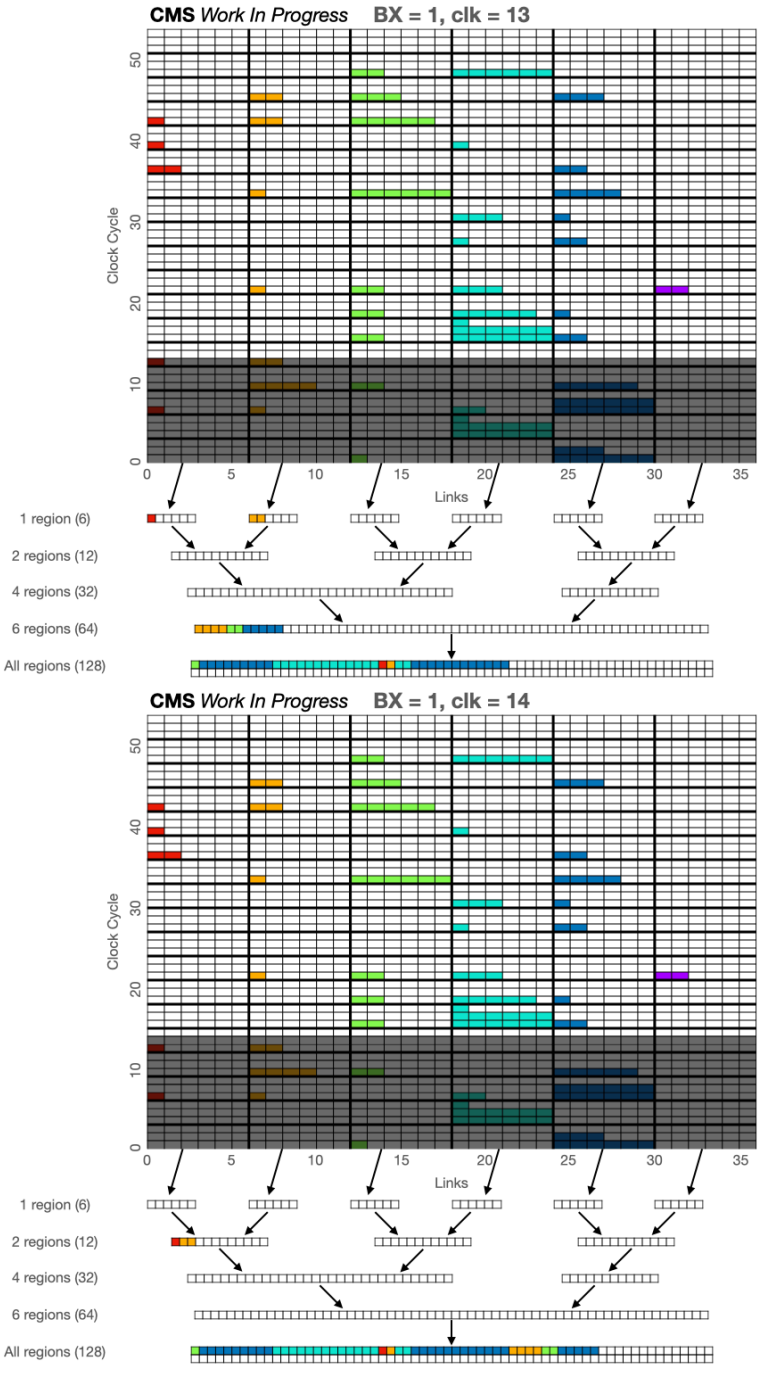


Figure E.8: Deregionizer snapshots, clocks 13–14

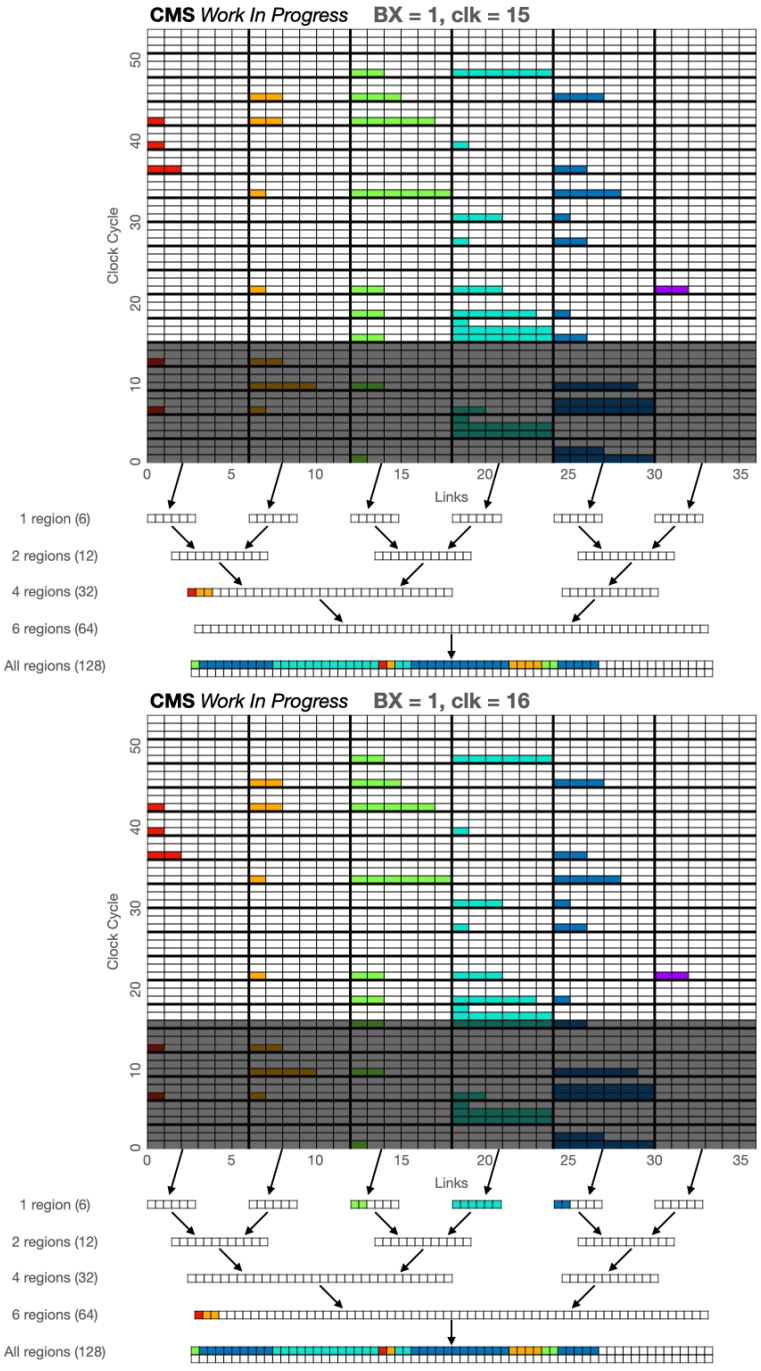


Figure E.9: Deregionizer snapshots, clocks 15–16

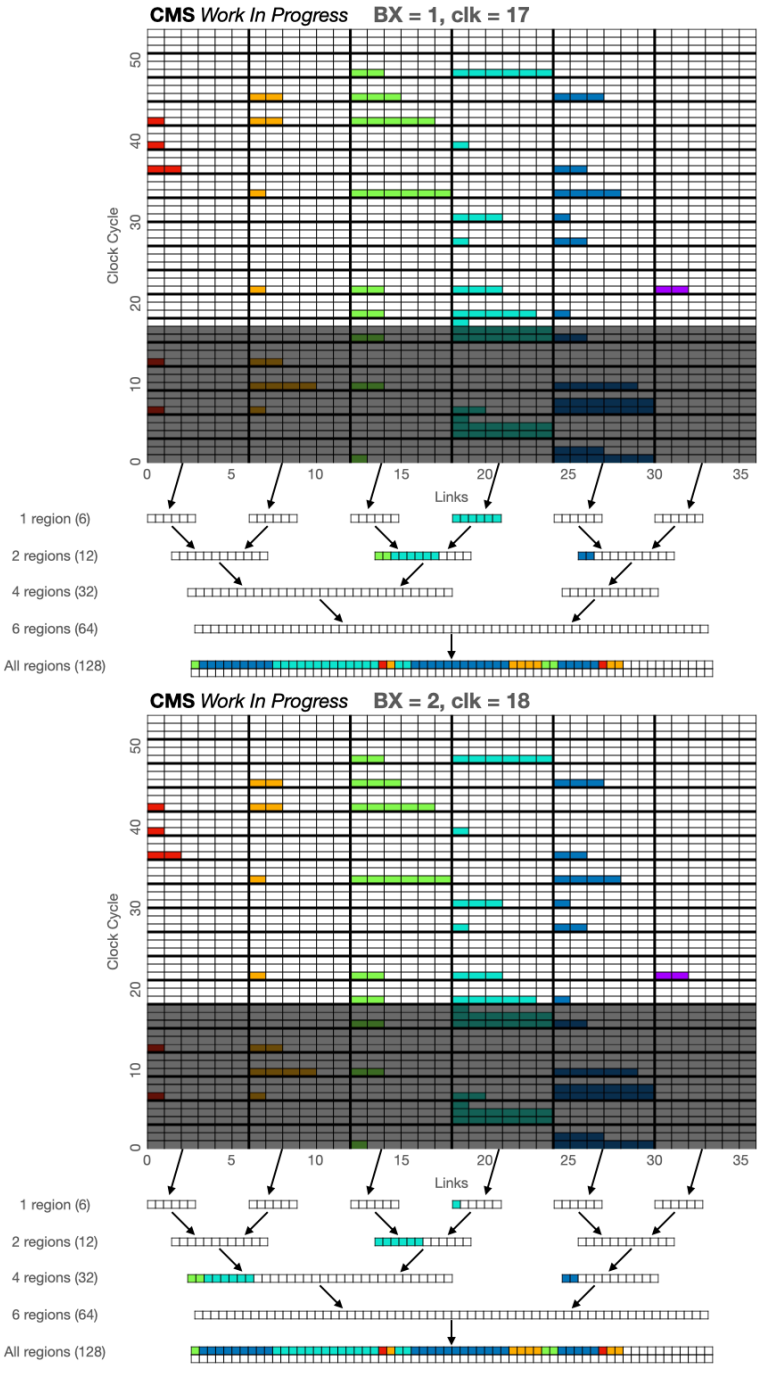


Figure E.10: Deregionizer snapshots, clocks 17–18

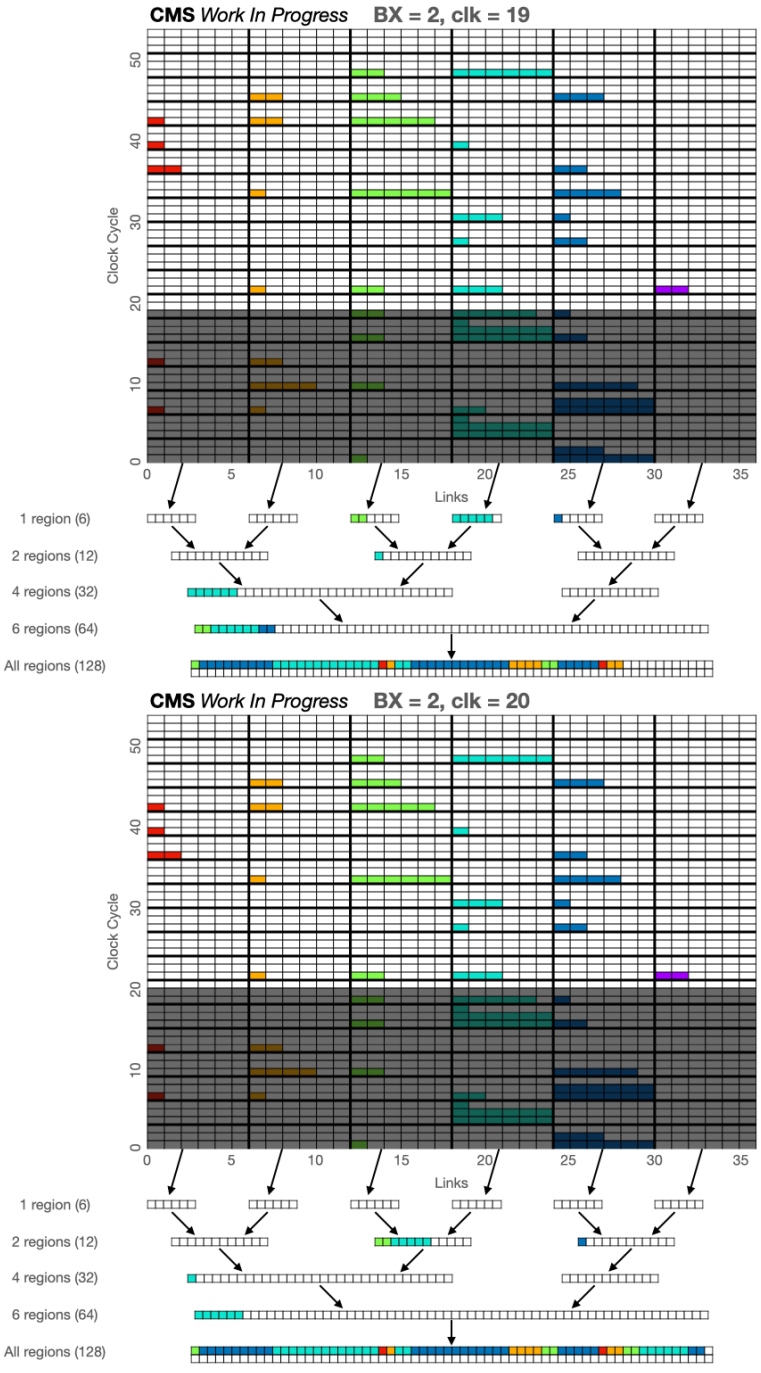


Figure E.11: Deregionizer snapshots, clocks 19–20

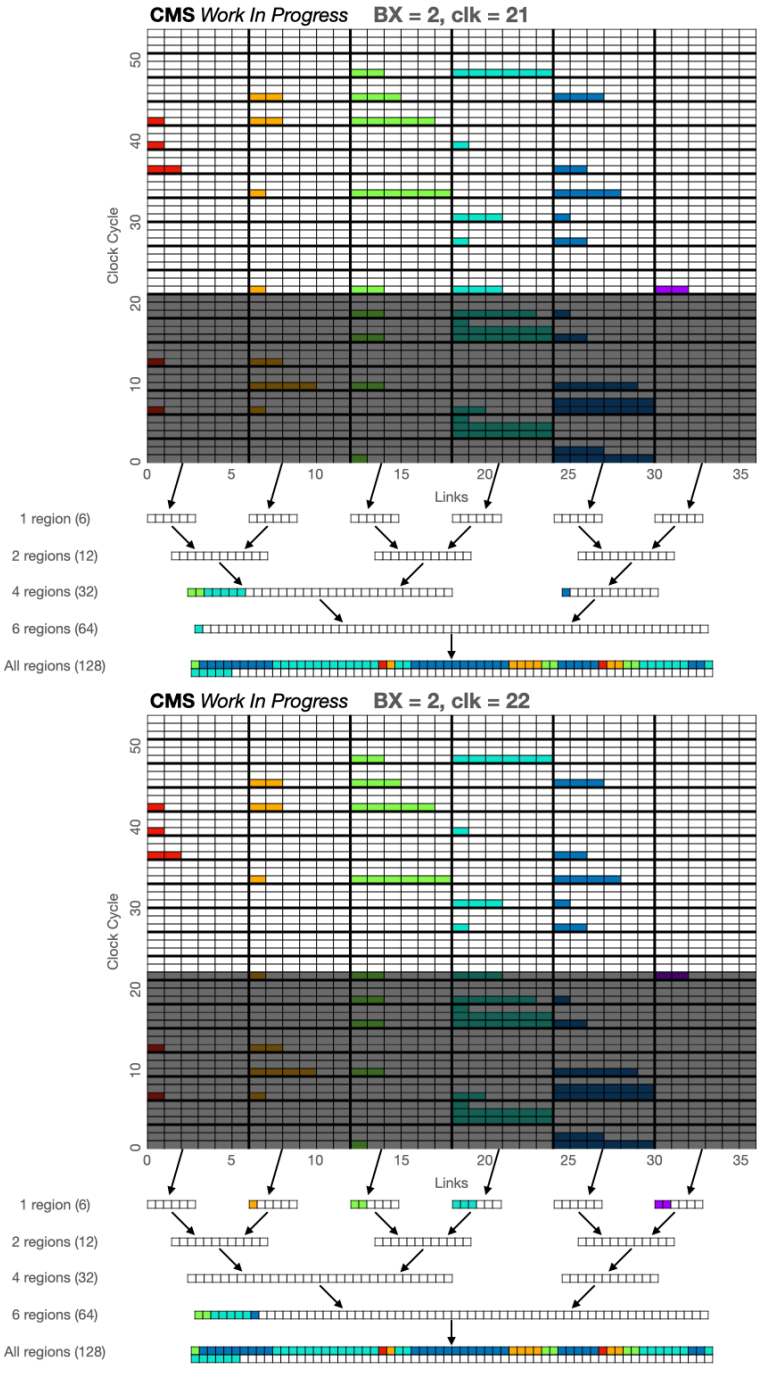


Figure E.12: Deregionizer snapshots, clocks 21–22

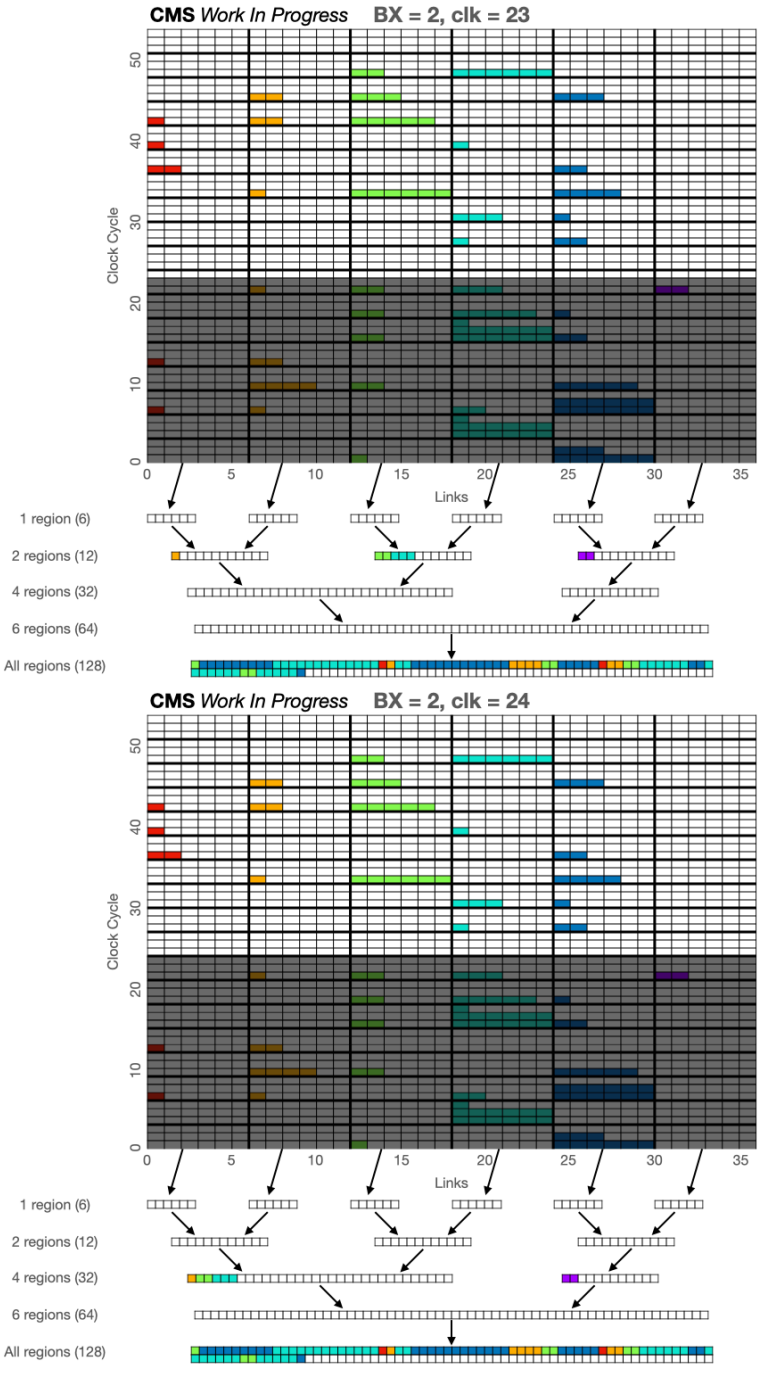


Figure E.13: Deregionizer snapshots, clocks 23–24

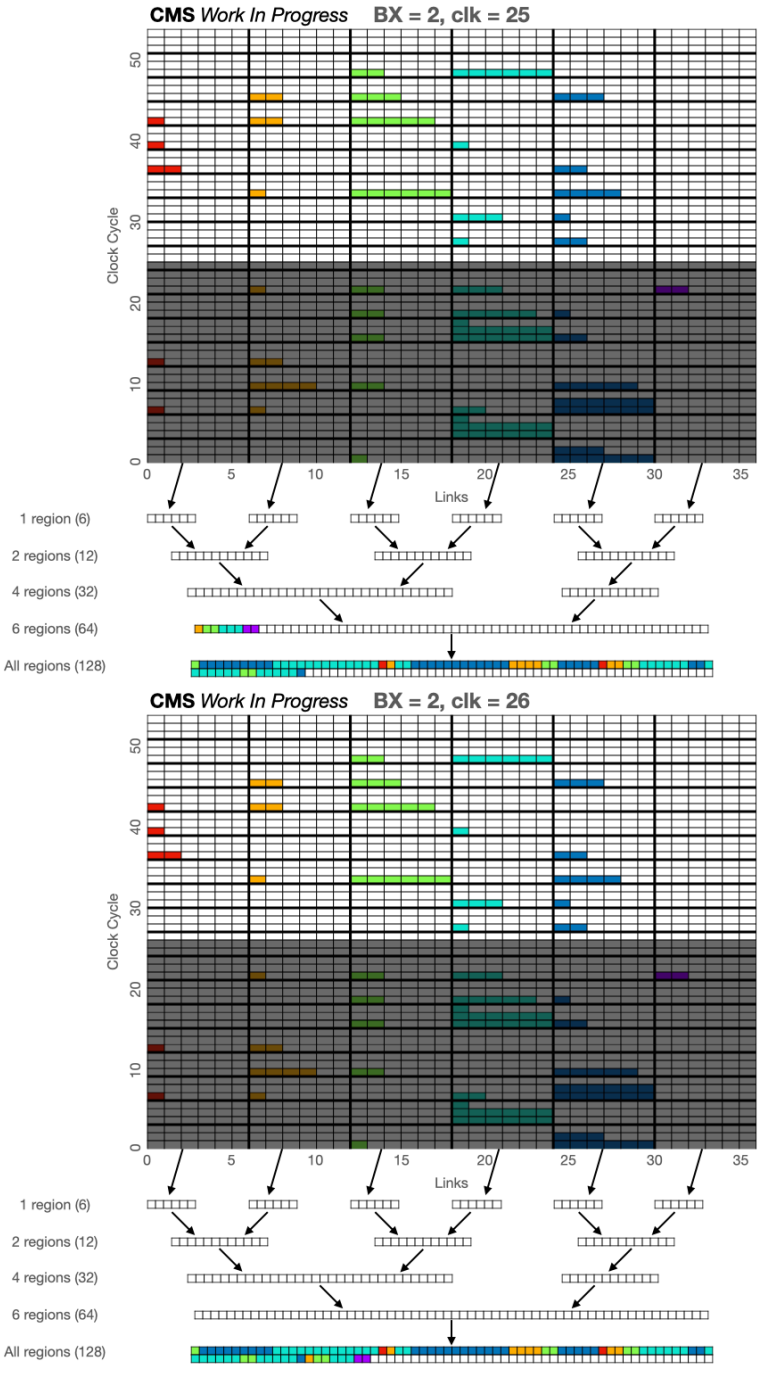


Figure E.14: Deregionizer snapshots, clocks 25–26

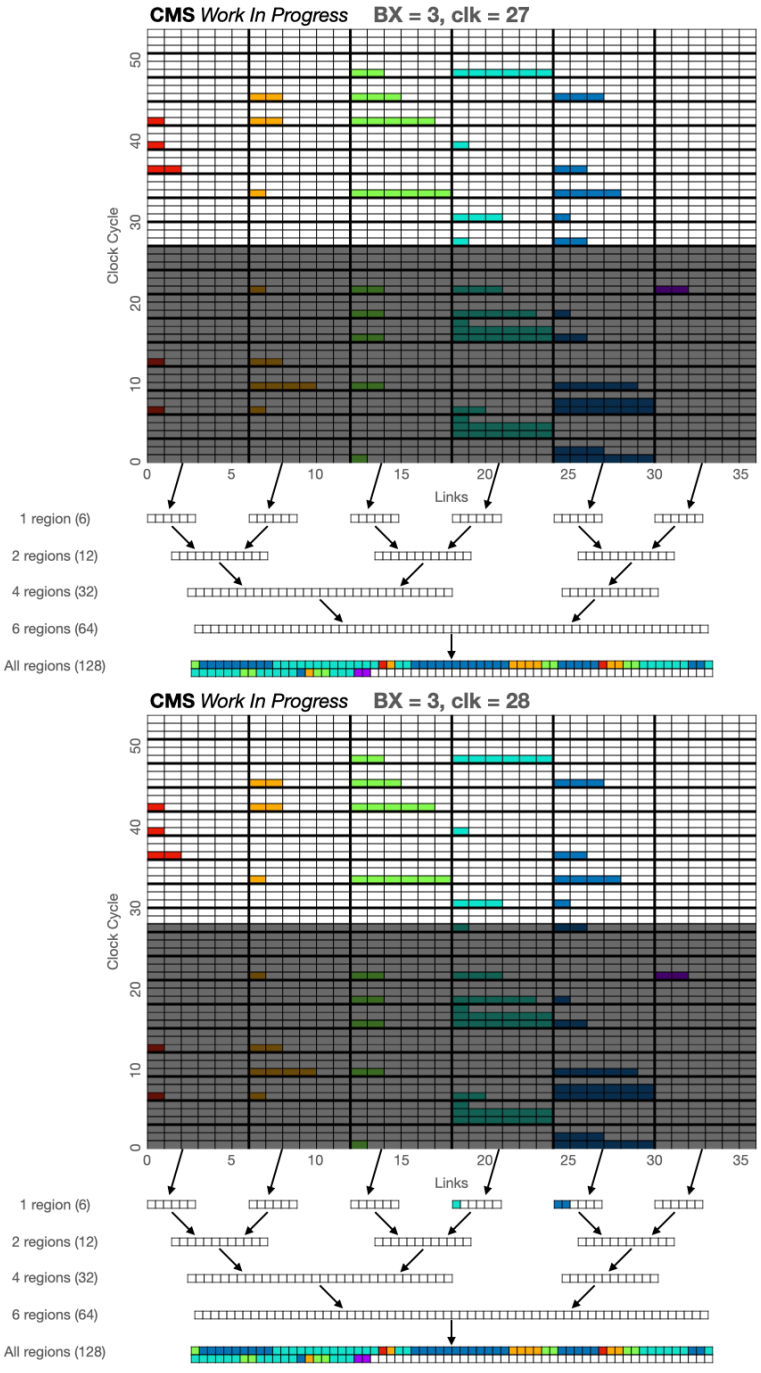


Figure E.15: Deregionizer snapshots, clocks 27–28

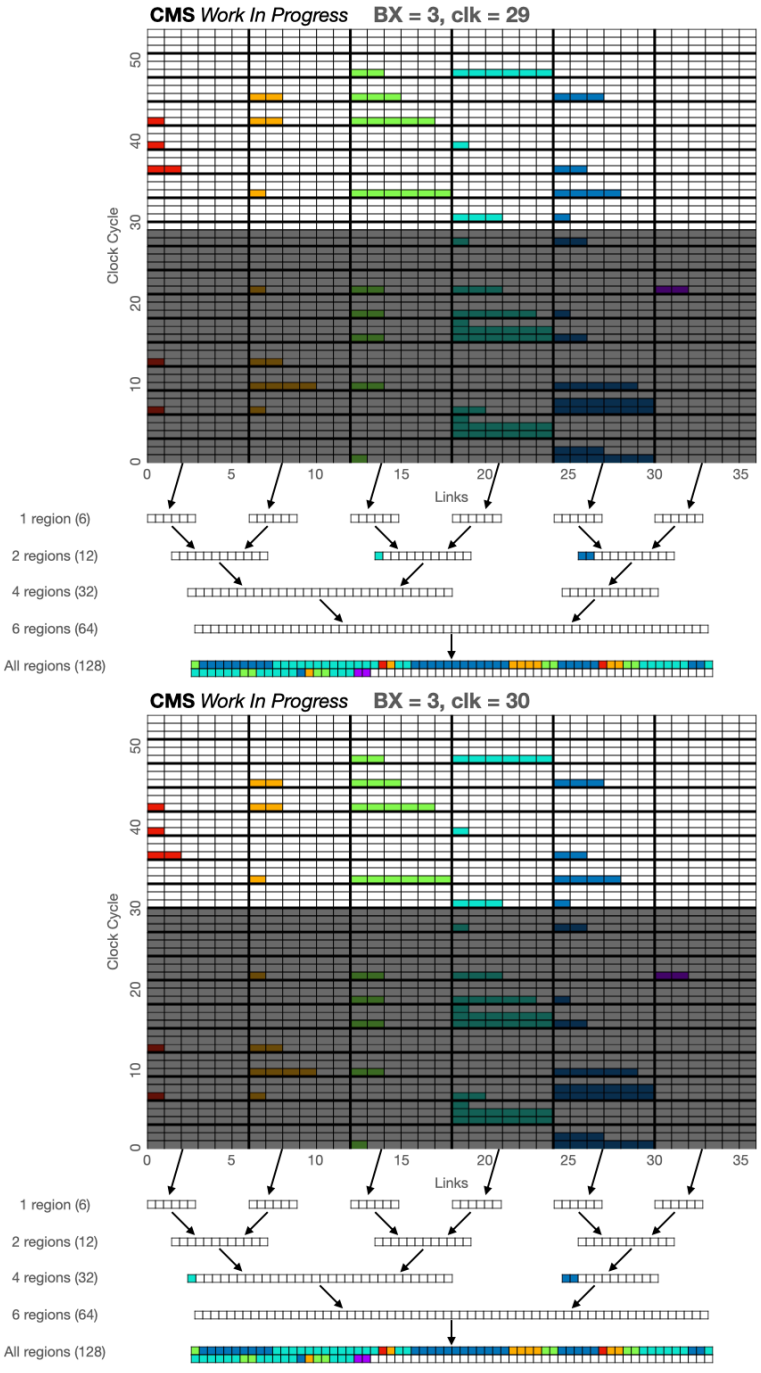


Figure E.16: Deregionizer snapshots, clocks 29–30

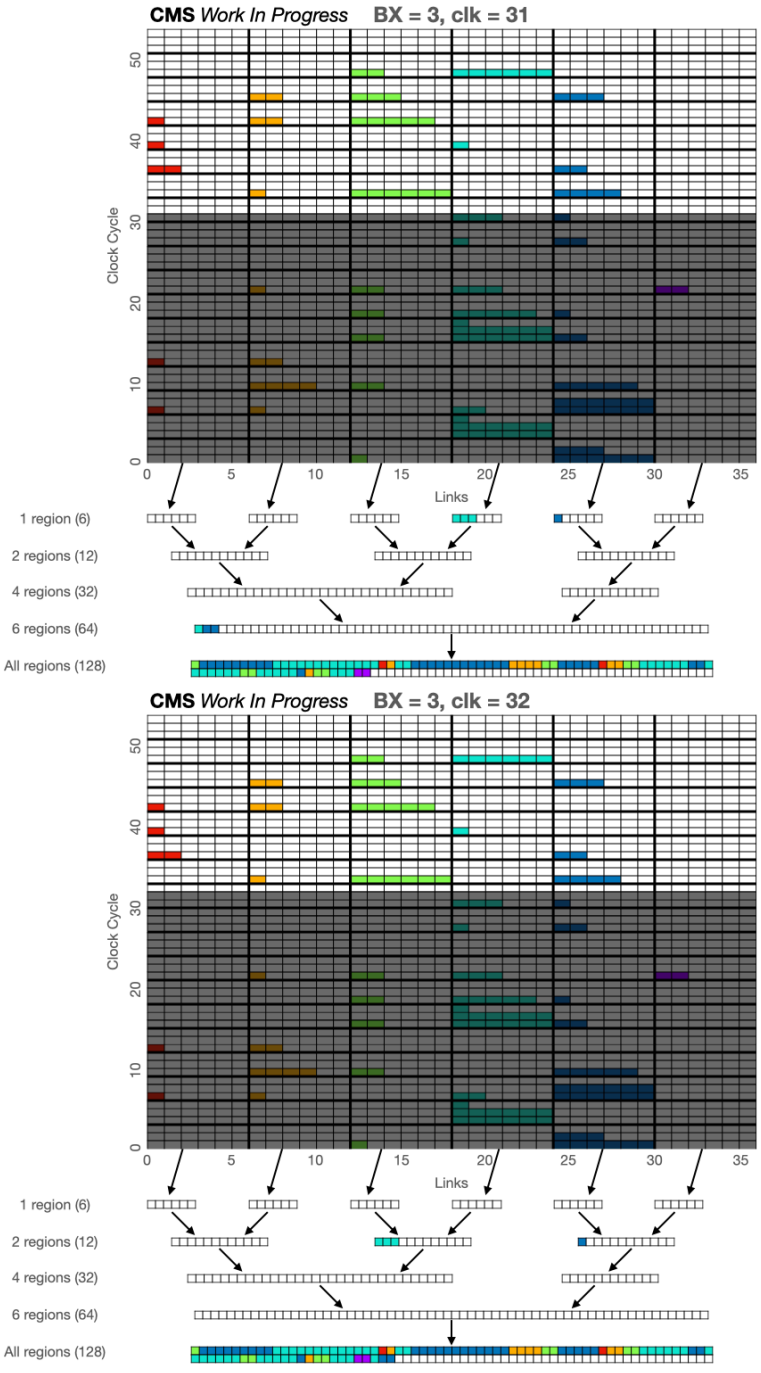


Figure E.17: Deregionizer snapshots, clocks 31–32

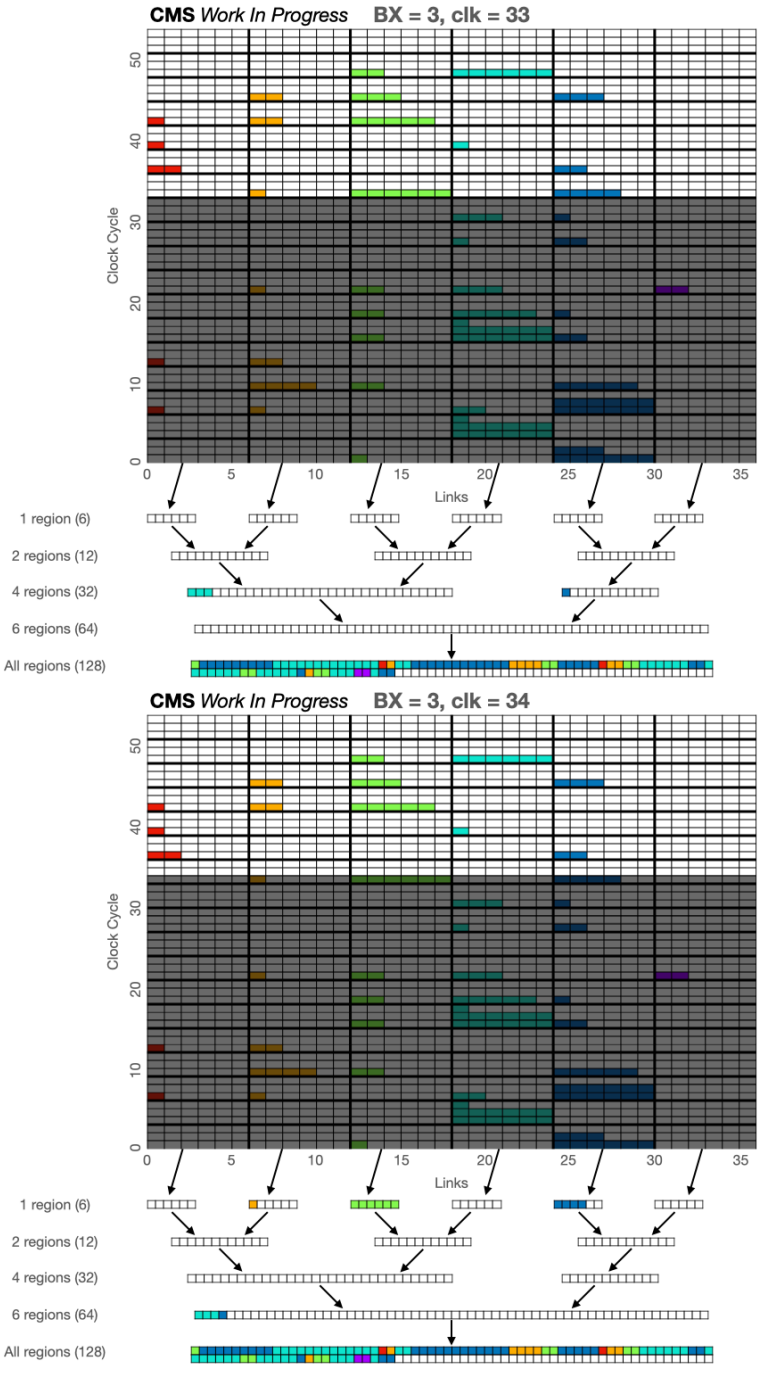


Figure E.18: Deregionizer snapshots, clocks 33–34

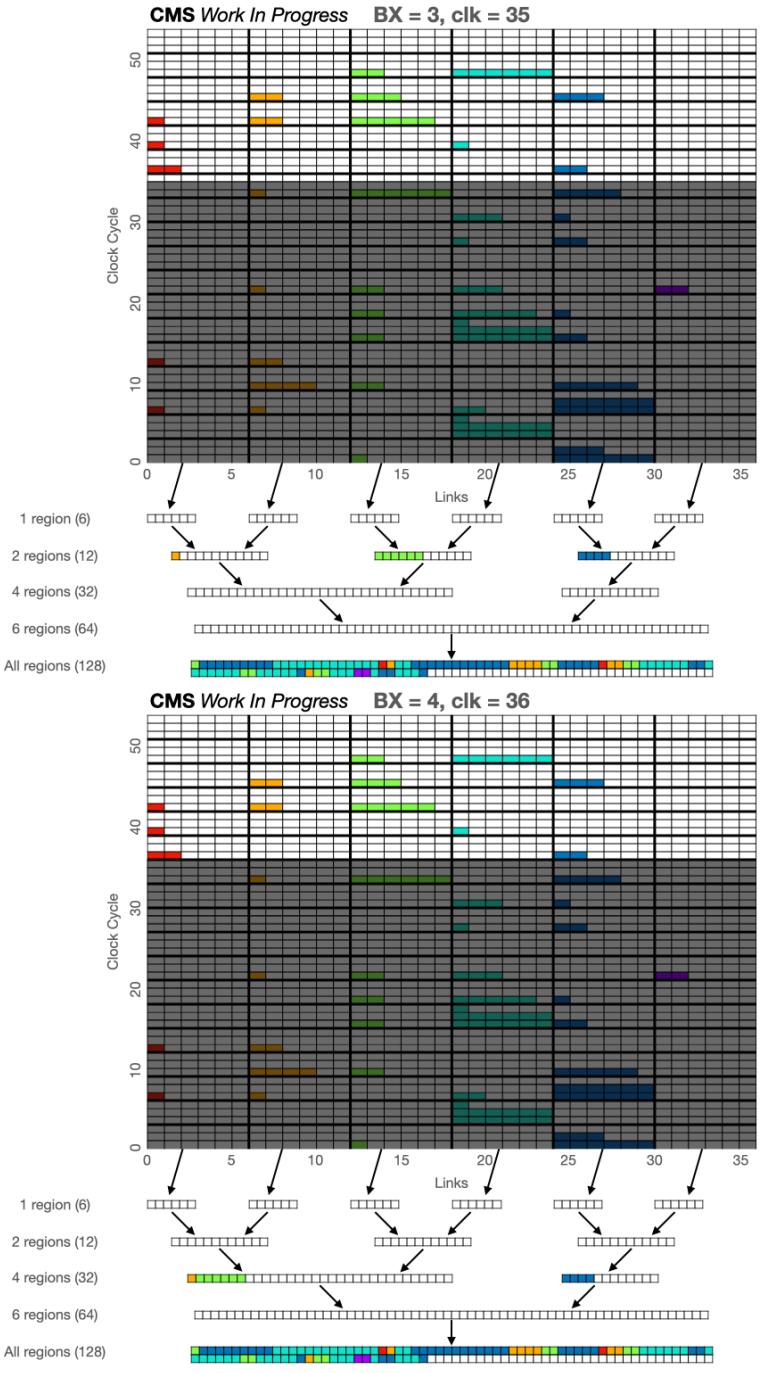


Figure E.19: Deregionizer snapshots, clocks 35–36

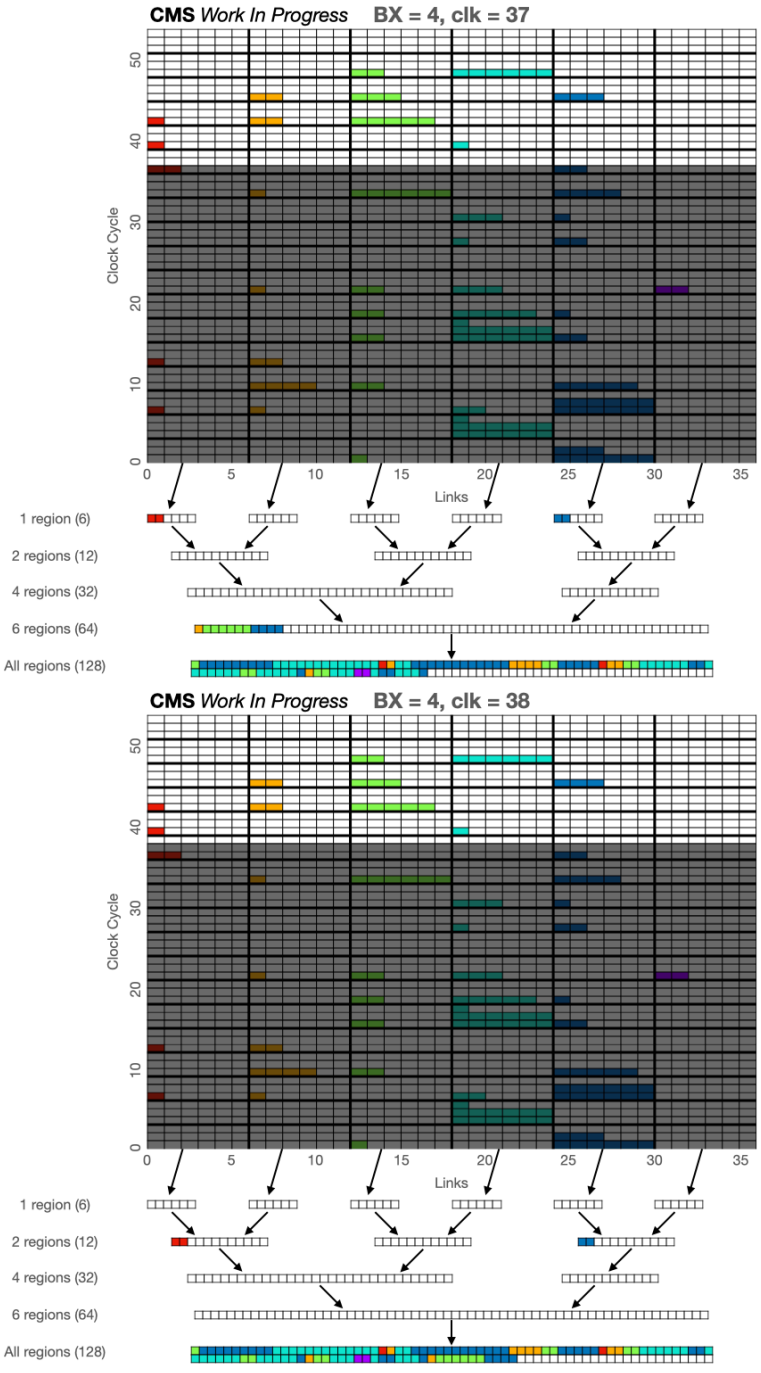


Figure E.20: Deregionizer snapshots, clocks 37–38

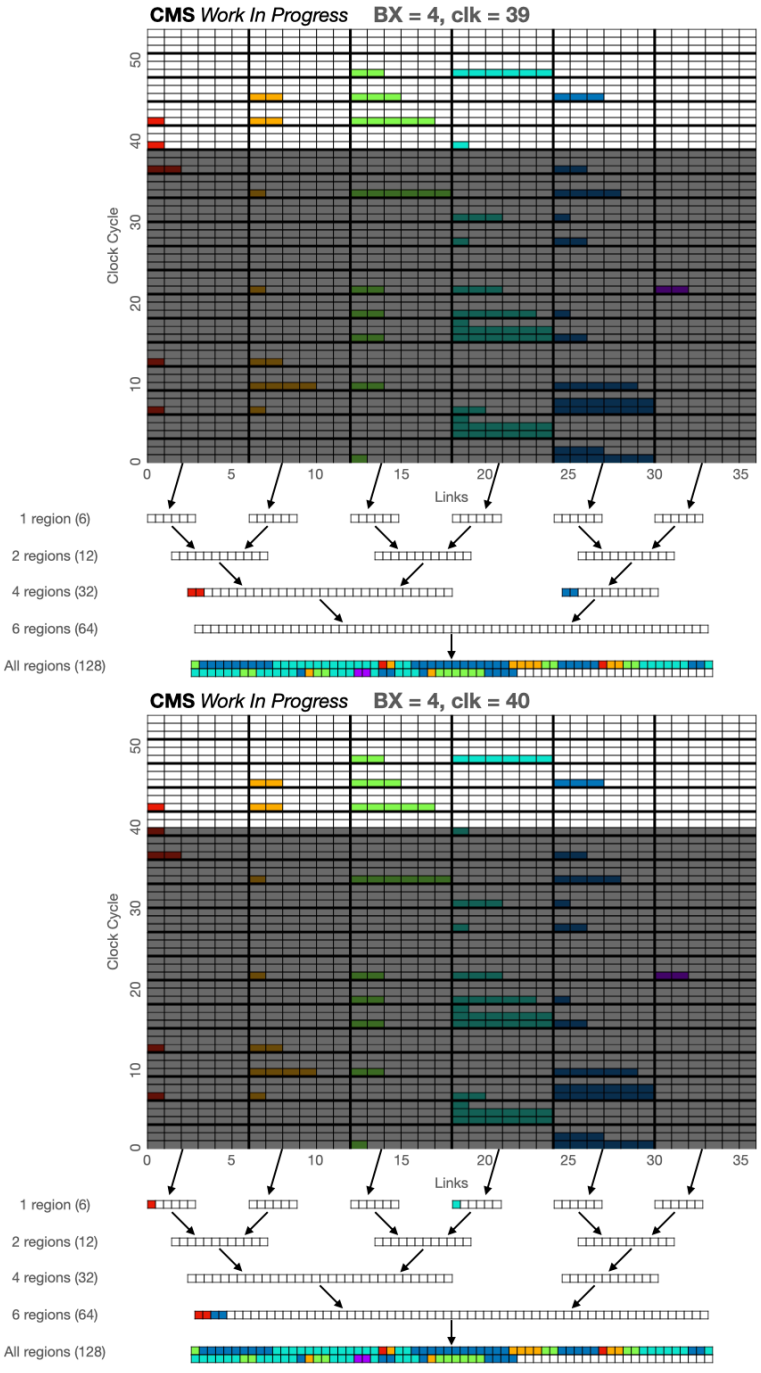


Figure E.21: Deregionizer snapshots, clocks 39–40

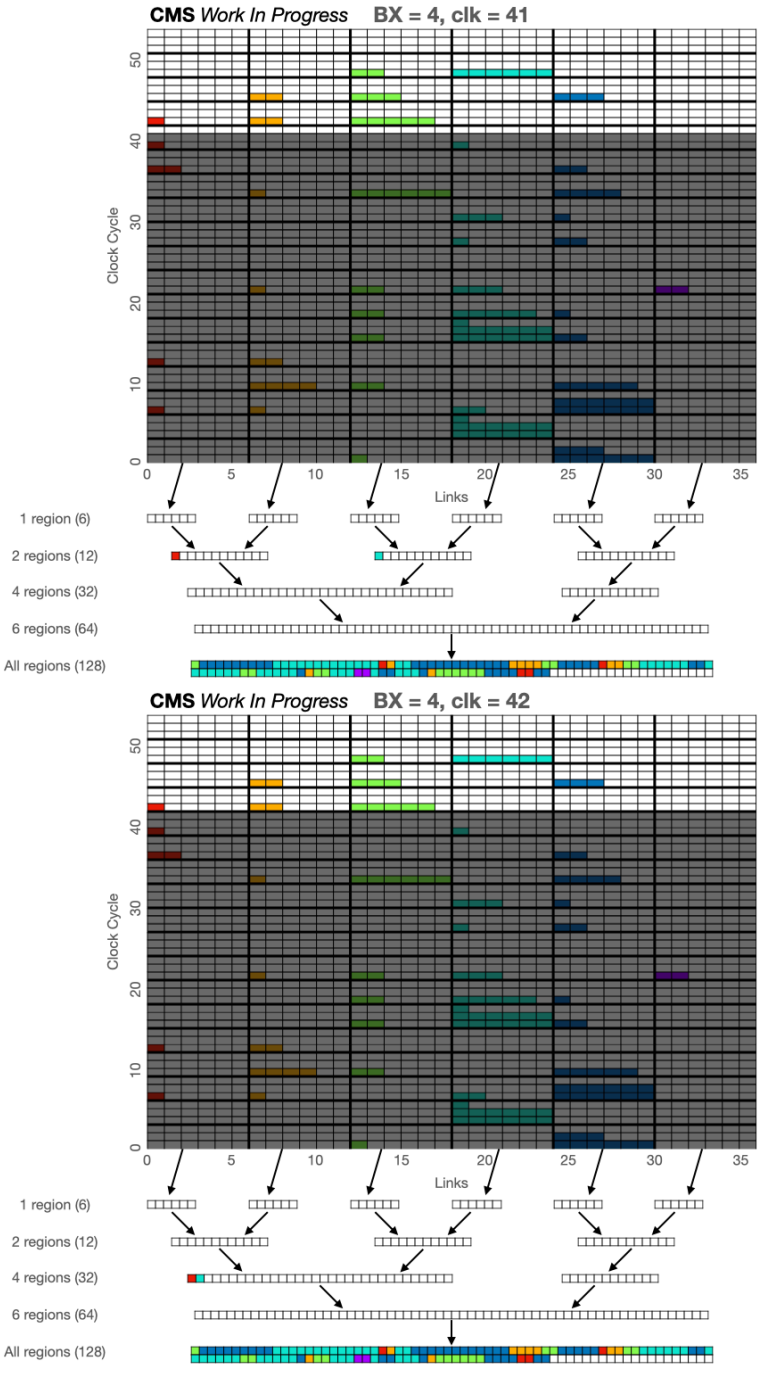


Figure E.22: Deregionizer snapshots, clocks 41–42

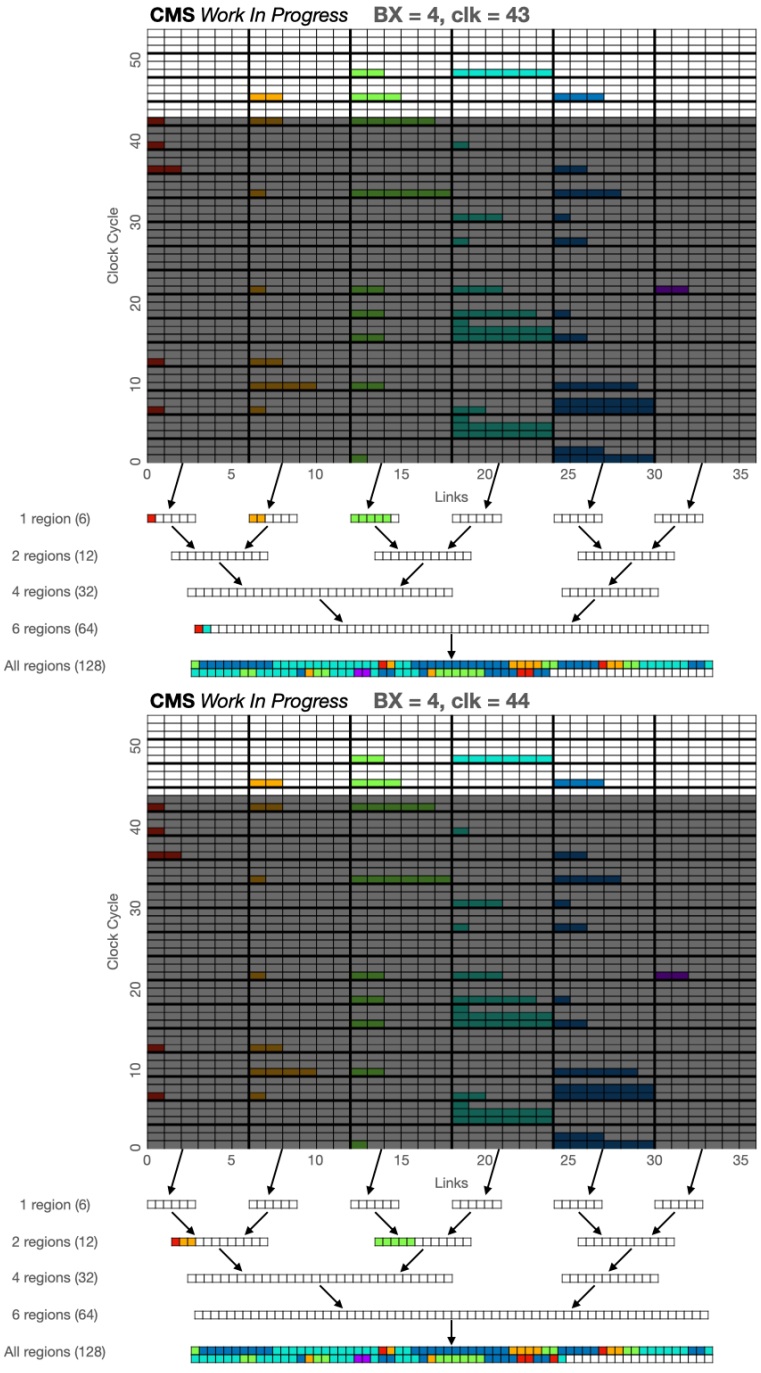


Figure E.23: Deregionizer snapshots, clocks 43–44

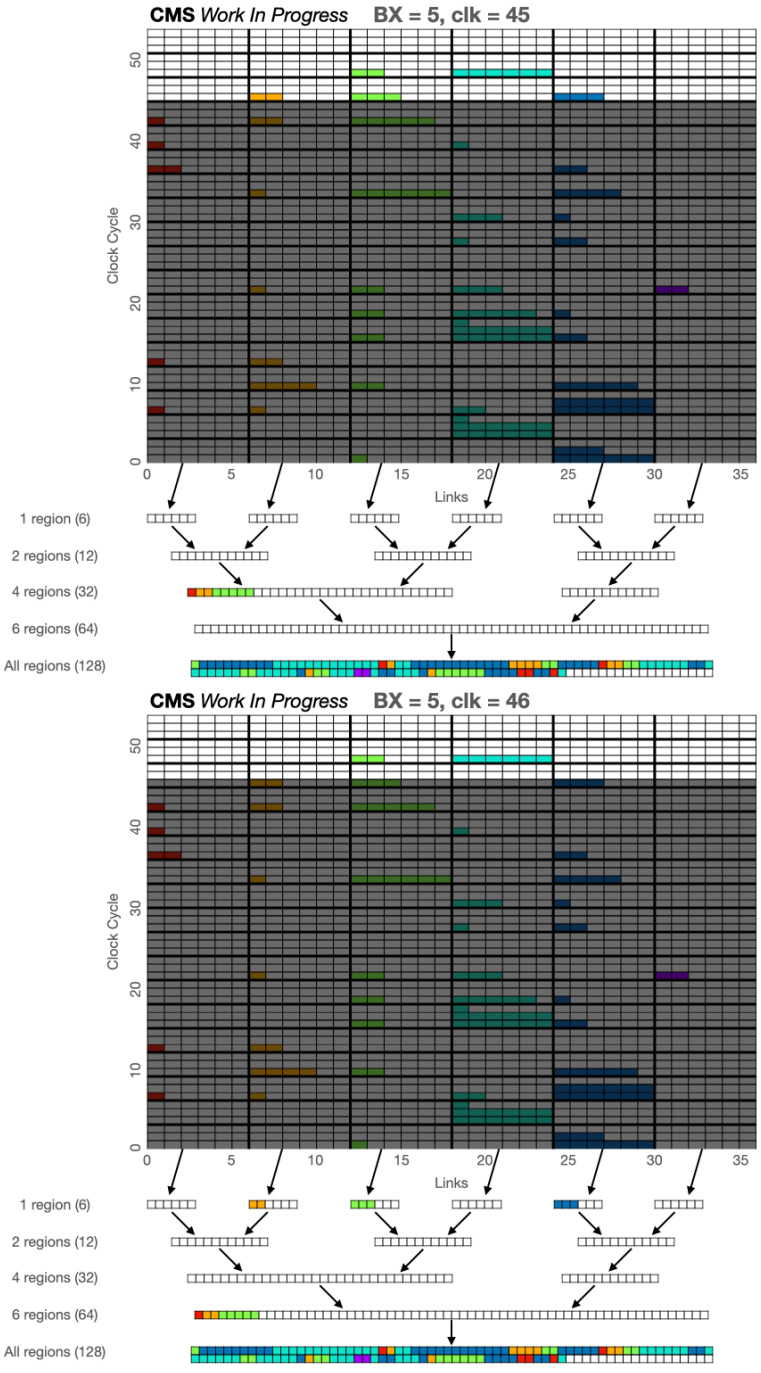


Figure E.24: Deregionizer snapshots, clocks 45–46

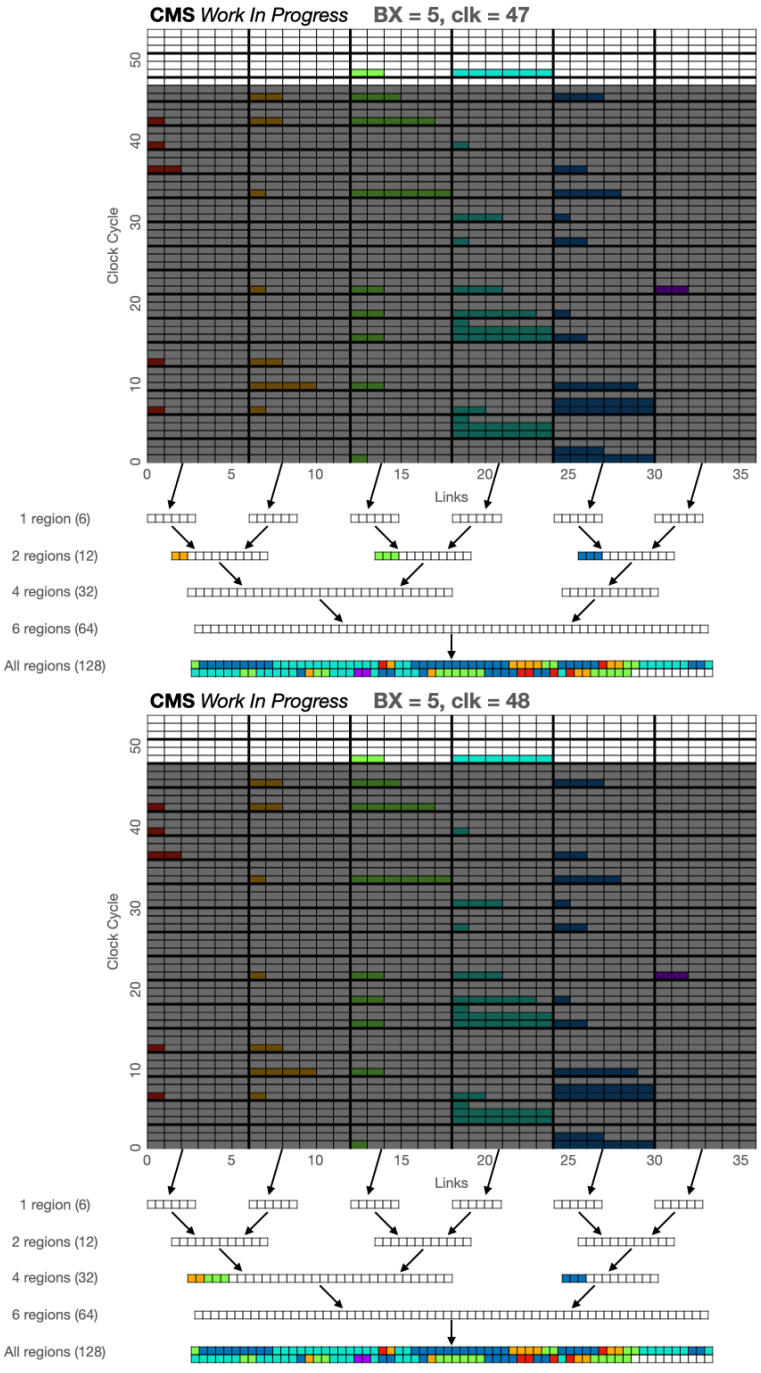


Figure E.25: Deregionizer snapshots, clocks 47–48

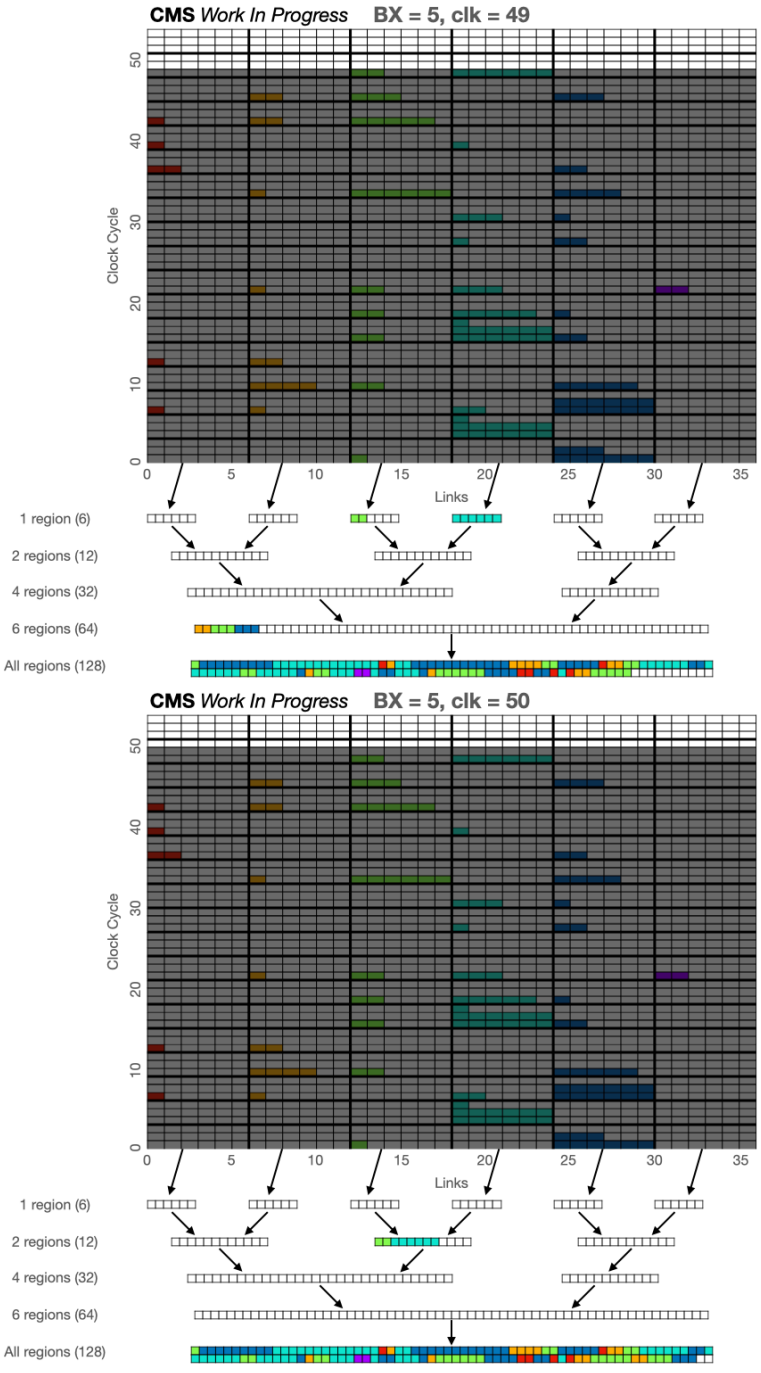


Figure E.26: Deregionizer snapshots, clocks 49–50

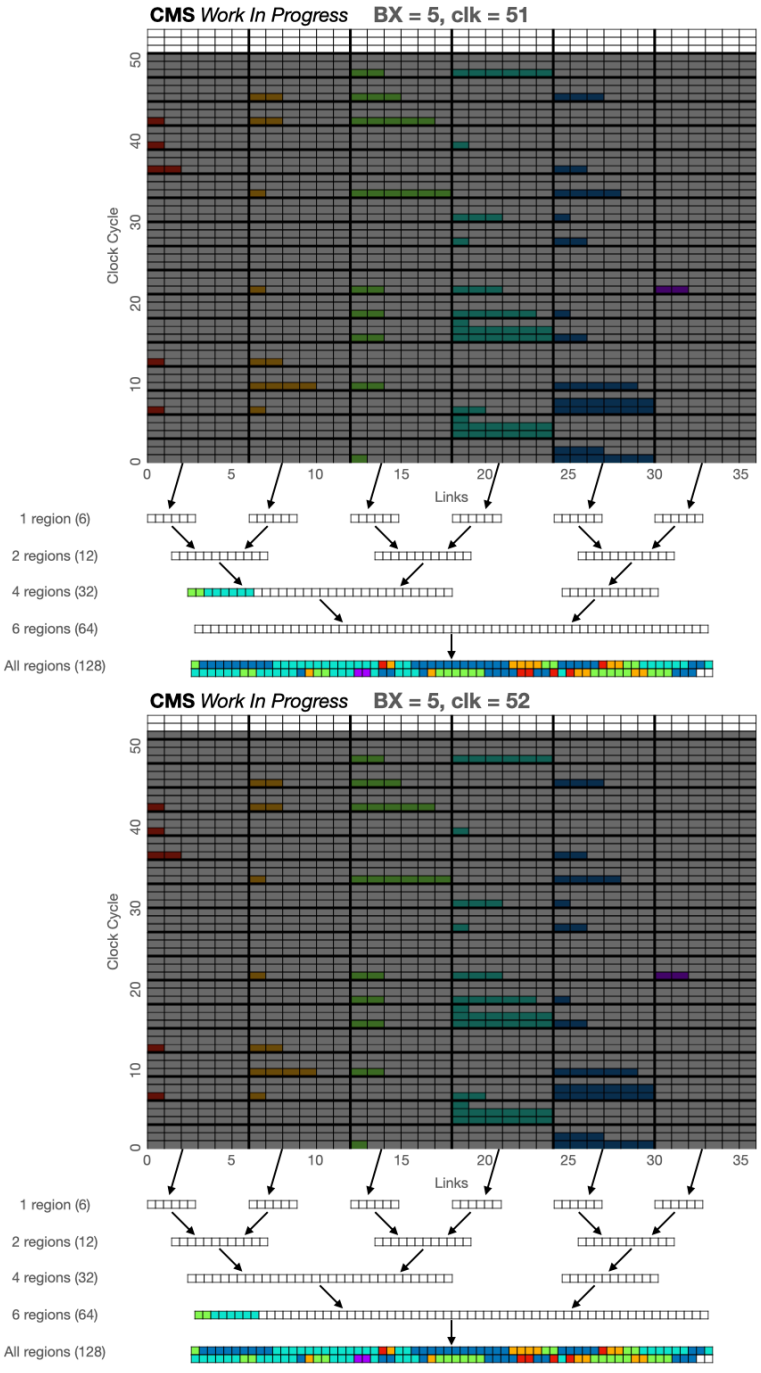


Figure E.27: Deregionizer snapshots, clocks 51–52

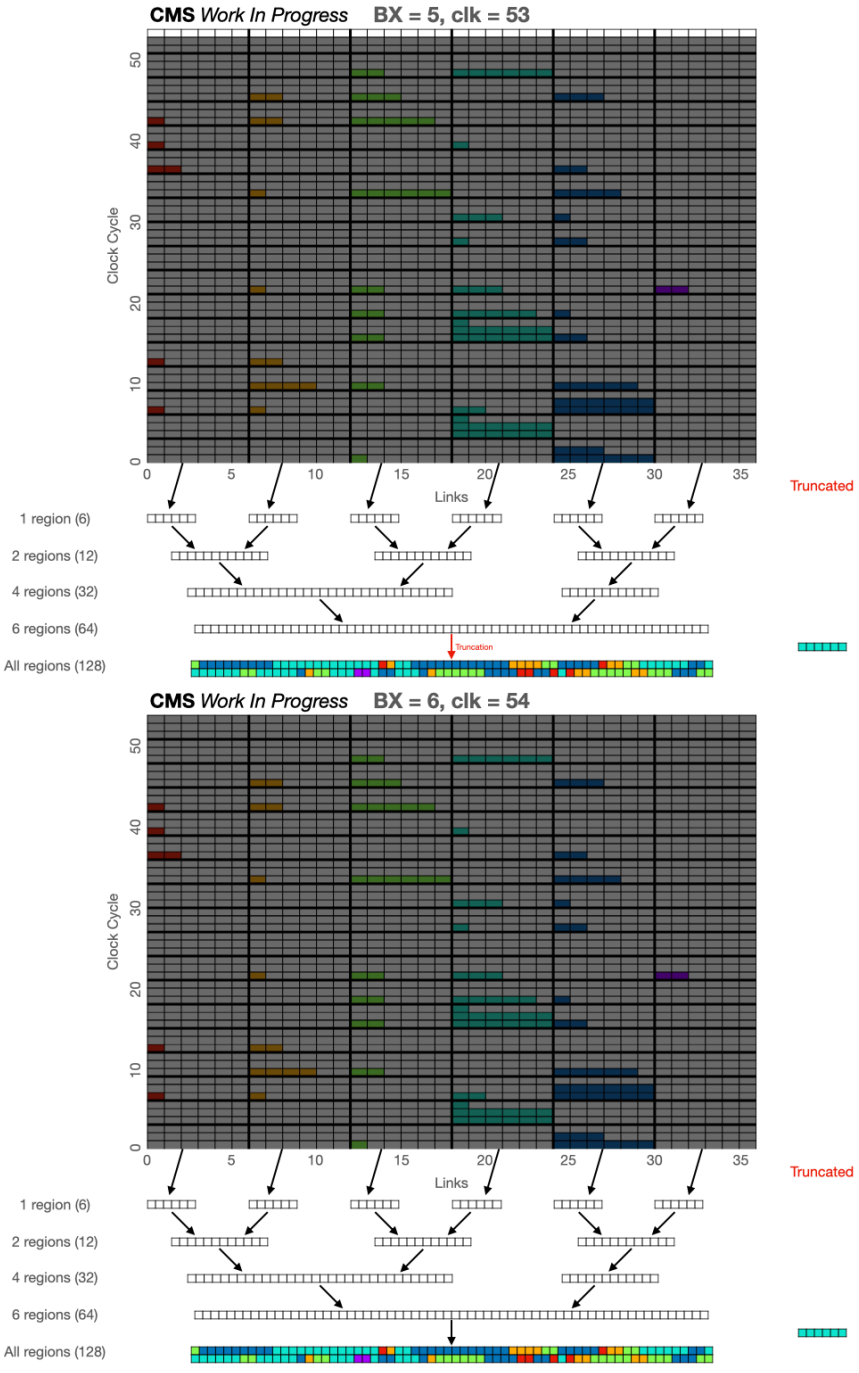


Figure E.28: Deregionizer snapshots, clocks 53–54

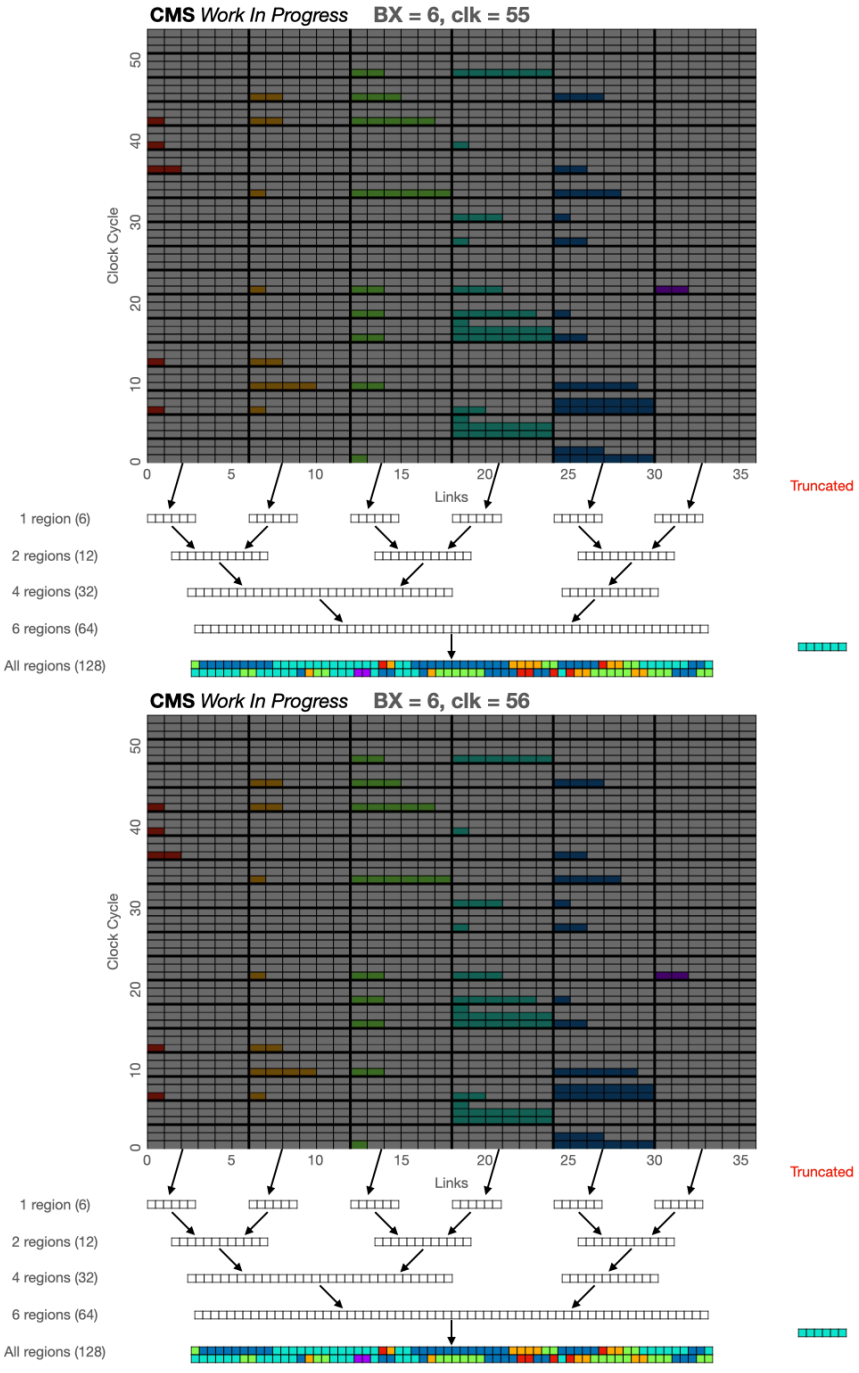


Figure E.29: Deregionizer snapshots, clocks 55–56

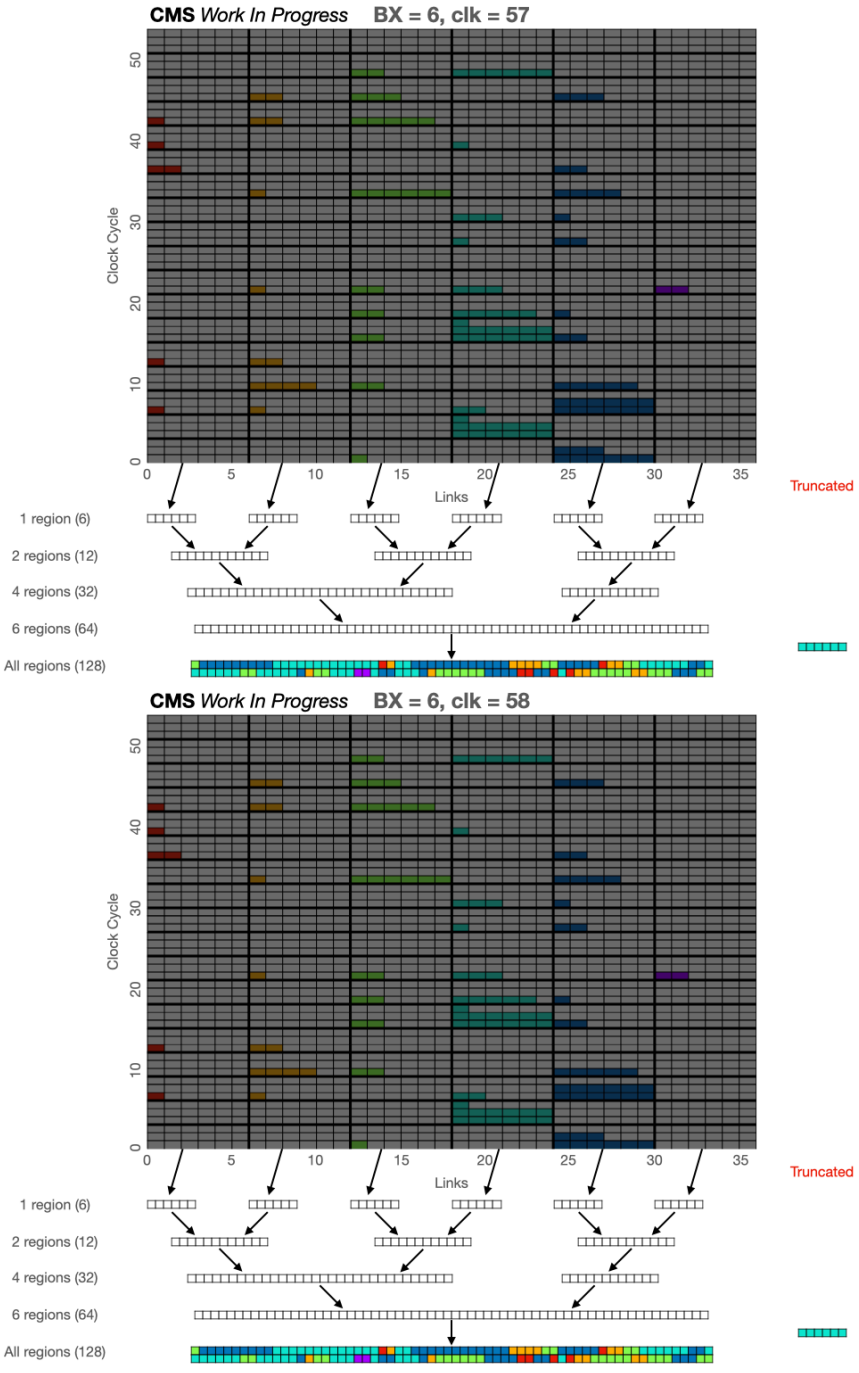


Figure E.30: Deregionizer snapshots, clocks 57–58

Bibliography

1. Zyla, P. *et al.* Review of Particle Physics. *PTEP* **2020** (2020).
2. CMS Collaboration. *CMS public results twiki* `https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsCombined`.
3. Garrett, K. & Duda, G. Dark Matter: A Primer. *Advances in Astronomy* **2011** (2011).
4. Ahmad, Q. R. *et al.* Measurement of the Rate of $\nu_e + d \rightarrow p + p + e^-$ Interactions Produced by ^{8}B Solar Neutrinos at the Sudbury Neutrino Observatory. *Phys. Review Lett.* **87** (2001).
5. Martin, S. P. A Supersymmetry Primer. *Advanced Series on Directions in High Energy Physics.* arXiv: `9709356 [hep-ph]` (1998).
6. Aitchison, I. J. Supersymmetry and the MSSM: An Elementary Introduction. arXiv: `0505105 [hep-ph]` (2005).
7. Fierz, M. Zur Fermischen Theorie des β-Zerfalls. *Zeitschrift für Physik* (1937).
8. Pal, P. B. Representation-independent manipulations with Dirac matrices and spinors. arXiv: `0703214 [physics.ed-ph]` (2007).
9. LHC SUSY Cross Section Working Group. *LHC cross sections twiki* `https://twiki.cern.ch/twiki/bin/view/LHCPhysics/SUSYCrossSections`.
10. Gröber, R., Mühlleitner, M. M., Popenda, E. & Wlotzka, A. Light stop decays: implications for LHC searches. *Eur. Phys. J. C* **75** (2015).
11. Baer, H., Chen, C.-h., Drees, M., Paige, F. & Tata, X. Collider Phenomenology for Supersymmetry with Large $\tan\beta$. *Phys. Rev. Lett.* **79** (1997).

12. Baer, H., Chen, C.-h., Drees, M., Paige, F. & Tata, X. Searching for supersymmetry at e+ e− supercolliders. *Int. J. Mod. Phys. A* **4** (1989).

13. Alwall, J., Schuster, P. & Toro, N. Simplified Models for a First Characterization of New Physics at the LHC. *Phys. Rev. D* **79.** arXiv: 0810.3921 [hep-ph] (2009).

14. Alves, D. *et al.* Simplified Models for LHC New Physics Searches. *J. Phys. G* **39.** arXiv: 1105.2838 [hep-ph] (2012).

15. Chatrchyan, S. *et al.* Interpretation of Searches for Supersymmetry with Simplified Models. *Phys. Rev. D* **88.** arXiv: 1301.2175 [hep-ex] (2013).

16. Hinchliffe, I., Paige, F. E., Shapiro, M. D., Söderqvist, J. & Yao, W. Precision SUSY measurements at CERN LHC. *Phys Rev. D* **55** (1997).

17. Randall, L. & Tucker-Smith, D. Dijet Searches for Supersymmetry at the Large Hadron Collider. *Phys Rev. Lett.* **101** (2008).

18. Lester, C. & Summers, D. Measuring masses of semi-invisibly decaying particle pairs produced at hadron colliders. *Phys. Lett. B* **463** (1999).

19. Tovey, D. R. On measuring the masses of pair-produced semi-invisibly decaying particles at hadron colliders. *JHEP* **2008** (2008).

20. CMS Collaboration. Inclusive search for squarks and gluinos in pp collisions at $\sqrt{s} = 7\,\text{TeV}$. *Phys. Rev. D* **85** (2012).

21. CMS Collaboration. *CMS public supersymmetry results twiki* https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsSUS.

22. CMS Collaboration. *Machine learning-based identification of highly Lorentz-boosted hadronically decaying particles at the CMS experiment* tech. rep. CMS-PAS-JME-18-002 (2019). https://cds.cern.ch/record/2683870.

23. CMS Collaboration. Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV. *JINST* **13.** arXiv: 1712.07158 (2018).

24. CMS Collaboration. Search for direct pair production of supersymmetric partners to the $\tilde{\tau}$ lepton in proton–proton collisions at $\sqrt{s} = 13\,\text{TeV}$. *Eur. Phys. J. C* **80** (2020).

25. CMS Collaboration. Search for supersymmetry in final states with two oppositely charged same-flavor leptons and missing transverse momentum in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$. *JHEP* **2021** (2021).

26. ALICE Collaboration. The ALICE experiment at the CERN LHC. *JINST* **3** (2008).

27. ATLAS Collaboration. The ATLAS Experiment at the CERN Large Hadron Collider. *JINST* **3** (2008).

28. CMS Collaboration. The CMS experiment at the CERN LHC. *JINST* **3** (2008).

29. LHCb Collaboration. The LHCb Detector at the LHC. *JINST* **3** (2008).

30. ATLAS Collaboration. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys. Lett. B* **716** (2012).

31. CMS Collaboration. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Phys. Lett. B* **716** (2012).

32. Mobs, E. The CERN accelerator complex - August 2018. Complexe des accélérateurs du CERN - Août 2018. `https://cds.cern.ch/record/2636343` (2018).

33. Brüning, O. S. *et al. LHC Design Report* `https://cds.cern.ch/record/782076` (2004).

34. CMS Collaboration. *CMS public luminosity results twiki* `https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults`.

35. Sakuma, T. Cutaway diagrams of CMS detector. `https://cds.cern.ch/record/2665537` (2019).

36. Universität Zürich. *How to draw diagrams in LaTeX with TikZ* `https://wiki.physik.uzh.ch/cms/latex:tikz`.

37. CMS Collaboration. *The CMS magnet project: Technical Design Report* `https://cds.cern.ch/record/331056` (1997).

38. CMS Collaboration. The performance of the CMS muon detector in proton-proton collisions at $\sqrt{s} = 7$ TeV at the LHC. *JINST* **8.** arXiv: `1306.6905` (2013).

39. Karimäki, V. *et al. The CMS tracker system project: Technical Design Report* https://cds.cern.ch/record/368412 (1997).

40. CMS Collaboration. *The CMS tracker: addendum to the Technical Design Report* https://cds.cern.ch/record/490194 (2000).

41. Adam, W. *et al.* Alignment of the CMS silicon strip tracker during stand-alone commissioning. *JINST* **4** (2009).

42. Dominguez, A. *et al. CMS Technical Design Report for the Pixel Detector Upgrade* tech. rep. CERN-LHCC-2012-016, CMS-TDR-11 (2012). https://cds.cern.ch/record/1481838.

43. CMS Collaboration. *The CMS electromagnetic calorimeter project: Technical Design Report* https://cds.cern.ch/record/349375 (1997).

44. Benaglia, A. The CMS ECAL performance with examples. *JINST* **9** (2014).

45. CMS Collaboration. *The CMS hadron calorimeter project: Technical Design Report* https://cds.cern.ch/record/357153 (1997).

46. Chatrchyan, S. and Khachatryan, V. and Sirunyan, A. and Adam, W. and Arnold, B. and Bergauer, H. and Bergauer, Th. and Dragicevic, M. and Eichberger, M. and Erö, J. and Friedl, M. and Frühwirth, R. and Ghete, V. and Hammer, J. and Hänsel, S. and Hoch, M. and Hörmann, N. and Hrubec, J. and Jeitler, M. and White, A. Identification and Filtering of Uncharacteristic Noise in the CMS Hadron Calorimeter. *JINST* **5** (2010).

47. CMS Collaboration. *The CMS muon project: Technical Design Report* https://cds.cern.ch/record/343814 (1997).

48. Colaleo, A., Safonov, A., Sharma, A. & Tytgat, M. *CMS Technical Design Report for the Muon Endcap GEM Upgrade* tech. rep. CERN-LHCC-2015-012, CMS-TDR-013 (2015). https://cds.cern.ch/record/2021453.

49. CMS Collaboration. Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$. *JINST* **13** (2018).

50. CMS Collaboration. Performance of the CMS Level-1 trigger in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$. *JINST* **15.** arXiv: 2006.10165 (2020).

51. CMS Collaboration. The CMS trigger system. *JINST* **12.** arXiv: 1609.02366 (2017).

52. CMS Collaboration. Particle-flow reconstruction and global event description with the CMS detector. *JINST* **12** (2017).

53. CMS Collaboration. Description and performance of track and primary vertex reconstruction with the CMS tracker. *JINST* **9** (2014).

54. CMS Collaboration. Alignment of the CMS silicon tracker during commissioning with cosmic rays. *JINST* **5** (2010).

55. Billoir, P. Progressive track recognition with a Kalman-like fitting procedure. *Computer Physics Communications* **57** (1989).

56. Billoir, P. & Qian, S. Simultaneous pattern recognition and track fitting by the Kalman filtering method. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **294** (1990).

57. Mankel, R. A concurrent track evolution algorithm for pattern recognition in the HERA-B main tracking system. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **395** (1997).

58. Rose, K. Deterministic annealing for clustering, compression, classification, regression, and related optimization problems. *Proceedings of the IEEE* **86** (1998).

59. CMS Collaboration. Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at $\sqrt{s} = 8\,\text{TeV}$. *JINST* **10** (2015).

60. CMS Collaboration. Electron and Photon performance in CMS with the full 2017 data sample and additional 2016 highlights for the CALOR 2018 Conference. https://cds.cern.ch/record/2320638 (2018).

61. CMS Collaboration. Performance of photon reconstruction and identification with the CMS detector in proton-proton collisions at $\sqrt{s} = 8\,\text{TeV}$. *JINST* **10** (2015).

62. Cacciari, M., Salam, G. P. & Soyez, G. The anti-k_t jet clustering algorithm. *JHEP* **04.** arXiv: 0802.1189 [hep-ex] (2008).

63. Cacciari, M., Salam, G. P. & Soyez, G. FastJet user manual. *Eur. Phys. J. C* **72.** arXiv: 1111.6097 [hep-ph] (2012).

64. CMS Collaboration. Pileup mitigation at CMS in 13 TeV data. *JINST* **15.** arXiv: 2003.00503 [hep-ex] (2020).

65. CMS Collaboration. Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV. *JINST* **12.** arXiv: 1607.03663 [hep-ex] (2017).

66. CMS Collaboration. *Heavy flavor identification at CMS with deep neural networks* https://twiki.cern.ch/twiki/bin/view/CMSPublic/BTV13TeVDPDeepCSV.

67. CMS Collaboration. Performance of reconstruction and identification of τ leptons decaying to hadrons and ν_τ in pp collisions at $\sqrt{s} =$ 13 TeV. *JINST* **13** (2018).

68. CMS Collaboration. Reconstruction and identification of τ lepton decays to hadrons and ν_τ at CMS. *JINST* **11** (2016).

69. CMS Collaboration. Performance of the DeepTau algorithm for the discrimination of taus against jets, electron, and muons. https://cds.cern.ch/record/2694158 (2019).

70. CMS Collaboration. Performance of missing transverse momentum reconstruction in proton-proton collisions at $\sqrt{s} = 13$ TeV using the CMS detector. *JINST* **14.** arXiv: 1903.06078 [hep-ex] (2019).

71. CMS Collaboration. Search for new physics in events with two soft oppositely charged leptons and missing transverse momentum in proton–proton collisions at $\sqrt{s} = 13$ TeV. *Phys. Lett. B* **782.** arXiv: 1801.01846 (2018).

72. ATLAS Collaboration. Searches for electroweak production of supersymmetric particles with compressed mass spectra in $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector. *Phys. Rev. D* **101.** arXiv: 1911.12606 (2020).

73. ATLAS Collaboration. Search for chargino–neutralino pair production in final states with three leptons and missing transverse momentum in $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector. *Eur. Phys. J. C* **81.** arXiv: {2106.01676} (hep-ex) (2021).

74. CMS Collaboration. Search for supersymmetry in final states with two or three soft leptons and missing transverse momentum in proton-proton collisions at
$$\sqrt{s}$$
= 13 TeV. *JHEP* **2022.** arXiv: 2111.06296 [hep-ex] (2022).

75. Beenakker, W. *et al.* Production of Charginos, Neutralinos, and Sleptons at Hadron Colliders. *Phys. Rev. Lett.* **83.** [Erratum: doi:10.1103/PhysRevLett.100.029901]. arXiv: hep-ph/9906298 [hep-ph] (1999).

76. Fuks, B., Klasen, M., Lamprea, D. R. & Rothering, M. Gaugino production in proton-proton collisions at a center-of-mass energy of 8 TeV. *JHEP* **10.** arXiv: 1207.2159 [hep-ph] (2012).

77. Fuks, B., Klasen, M., Lamprea, D. R. & Rothering, M. Precision predictions for electroweak superpartner production at hadron colliders with Resummino. *Eur. Phys. J. C* **73.** arXiv: 1304.0790 [hep-ph] (2013).

78. Edsjo, J. & Gondolo, P. Neutralino relic density including coannihilations. *Phys. Rev. D* **56.** arXiv: hep-ph/9704361 (1997).

79. Chen, C.-H., Drees, M. & Gunion, J. F. Nonstandard string-SUSY scenario and its phenomenological implications. *Phys. Rev. D* **55.** [Erratum: doi:10.1103/PhysRevD.60.039901] (1997).

80. Profumo, S., Stefaniak, T. & Stephenson-Haskins, L. Not-so-well-tempered neutralino. *Phys. Rev. D* **96.** arXiv: 1706.08537 [hep-ph] (2017).

81. Baer, H., Barger, V. & Huang, P. Hidden SUSY at the LHC: the light higgsino-world scenario and the role of a lepton collider. *JHEP* **11.** arXiv: 1107.5581 [hep-ph] (2011).

82. Baer, H., Mustafayev, A. & Tata, X. Monojet plus soft dilepton signal from light higgsino pair production at LHC14. *Phys. Rev. D* **90,** 115007. arXiv: 1409.7058 [hep-ph] (2014).

83. Han, C. *et al.* Probing light higgsinos in natural SUSY from monojet signals at the LHC. *JHEP* **02.** arXiv: 1310.4274 [hep-ph] (2014).

84. Han, Z., Kribs, G. D., Martin, A. & Menon, A. Hunting quasidegenerate Higgsinos. *Phys. Rev. D* **89.** arXiv: 1401.1235 [hep-ph] (2014).

85. MSSM Working Group. *The minimal supersymmetric standard model: Group summary report* 1998.

86. Beenakker, W., Hopker, R. & Spira, M. *PROSPINO: A program for the production of supersymmetric particles in next-to-leading order QCD* 1996.

87. Djouadi, A., Kneur, J.-L. & Moultaka, G. SuSpect: A Fortran code for the supersymmetric and Higgs particle spectrum in the MSSM. *Comput. Phys. Commun.* **176,** 426. arXiv: hep-ph/0211331 [hep-ph] (2007).

88. Mühlleitner, M., Djouadi, A. & Mambrini, Y. SDECAY: A Fortran code for the decays of the supersymmetric particles in the MSSM. *Comput. Phys. Commun.* **168,** 46. arXiv: hep-ph/0311167 [hep-ph] (2005).

89. Djouadi, A., Kalinowski, J. & Spira, M. HDECAY: A program for Higgs boson decays in the standard model and its supersymmetric extension. *Comput. Phys. Commun.* **108,** 56. arXiv: hep-ph/9704448 [hep-ph] (1998).

90. Mühlleitner, M. M., Djouadi, A. & Spira, M. *Decays of supersymmetric particles: The Program SUSY-HIT* in *Physics at LHC. Proc. 3rd Conf. Acta Phys. Polon. B* **38** (2007) 635 (2006), 635. arXiv: hep-ph/0609 292 [hep-ph].

91. Skands, P. Z. *et al.* SUSY Les Houches accord: Interfacing SUSY spectrum calculators, decay packages, and event generators. *JHEP* **07,** 036. arXiv: hep-ph/0311123 [hep-ph] (2004).

92. Balázs, C., Carena, M. & Wagner, C. E. M. Dark matter, light stops and electroweak baryogenesis. *Phys. Rev. D* **70.** arXiv: hep-ph/0403 224 [hep-ph] (2004).

93. De Sanctis, U., Lari, T., Montesano, S. & Troncon, C. Perspectives for the detection and measurement of supersymmetry in the focus point region of mSUGRA models with the ATLAS detector at LHC. *Eur. Phys. J. C* **52.** arXiv: 0704.2515 [hep-ex] (2007).

94. Sjöstrand, T. *et al.* An introduction to PYTHIA 8.2. *Comput. Phys. Commun.* **191,** 159. arXiv: 1410.3012 [hep-ph] (2015).

95. Chen, C., Drees, M. & Gunion, J. A Nonstandard string / SUSY scenario and its phenomenological implications. *Phys. Rev. D* **55.** arXiv: hep-ph/9607421 (1997).

96. Alwall, J. *et al.* The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *JHEP* **07.** arXiv: 1405.0301 [hep-ph] (2014).

97. Frederix, R. & Frixione, S. Merging meets matching in MC@NLO. *JHEP* **12.** arXiv: 1209.6215 [hep-ph] (2012).

98. Alwall, J. *et al.* Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions. *Eur. Phys. J. C* **53.** arXiv: 0706.2569 [hep-ph] (2008).

99. Nason, P. A new method for combining NLO QCD with shower Monte Carlo algorithms. *JHEP* **11.** arXiv: hep-ph/0409146 [hep-ph] (2004).

100. Frixione, S., Nason, P. & Oleari, C. Matching NLO QCD computations with parton shower simulations: the POWHEG method. *JHEP* **11.** arXiv: 0709.2092 [hep-ph] (2007).

101. Alioli, S., Nason, P., Oleari, C. & Re, E. A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX. *JHEP* **06.** arXiv: 1002.2581 [hep-ph] (2010).

102. Alioli, S., Nason, P., Oleari, C. & Re, E. NLO single-top production matched with shower in POWHEG: s- and t-channel contributions. *JHEP* **09.** [Erratum: *JHEP* **02** (2010) 011, doi:10.1007/JHEP02(2010)011]. arXiv: 0907.4076 [hep-ph] (2009).

103. Melia, T., Nason, P., Röntsch, R. & Zanderighi, G. W^+W^-, WZ and ZZ production in the POWHEG BOX. *JHEP* **11.** arXiv: 1107.5051 [hep-ph] (2011).

104. Nason, P. & Zanderighi, G. W^+W^-, WZ and ZZ production in the POWHEG-BOX-V2. *Eur. Phys. J. C* **74.** arXiv: 1311.1365 [hep-ph] (2014).

105. Re, E. Single-top Wt-channel production matched with parton showers using the POWHEG method. *Eur. Phys. J. C* **71,** 1547. arXiv: 1009.2450 [hep-ph] (2011).

106. Ball, R. D. *et al.* Parton distributions for the LHC Run II. *JHEP* **04.** arXiv: 1410.8849 [hep-ph] (2015).

107. Ball, R. D. *et al.* Parton distributions from high-precision collider data. *Eur. Phys. J. C* **77** (2017).

108. Skands, P., Carrazza, S. & Rojo, J. Tuning PYTHIA 8.1: the Monash 2013 tune. *Eur. Phys. J. C* **74.** arXiv: 1404.5630 [hep-ph] (2014).

109. CMS Collaboration. Event generator tunes obtained from underlying event and multiparton scattering measurements. *Eur. Phys. J. C* **76.** arXiv: 1512.00815 [hep-ex] (2016).

110. CMS Collaboration. Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements. *Eur. Phys. J. C* **80.** arXiv: 1903.12179 (2020).

111. Agostinelli, S. *et al.* GEANT4—a simulation toolkit. *Nucl. Instrum. Meth. A* **506** (2003).

112. Abdullin, S., Azzi, P., Beaudette, F., Janot, P. & Perrotta, A. The fast simulation of the CMS detector at LHC. *J. Phys. Conf. Ser.* **331** (2011).

113. Giammanco, A. The Fast Simulation of the CMS Experiment. *J. Phys. Conf. Ser.* **513** (2014).

114. CMS Collaboration. *CMS tag-and-probe twiki* https://twiki.cern.ch/twiki/bin/view/CMSPublic/TagAndProbe.

115. CMS Collaboration. *Jet algorithms performance in 13 TeV data* CMS Physics Analysis Summary CMS-PAS-JME-16-003 (2017). https://cds.cern.ch/record/2256875.

116. CMS Collaboration. Measurement of the inelastic proton-proton cross section at $\sqrt{s} = 13$ TeV. *JHEP* **07.** arXiv: 1802.02613 (2018).

117. Elagin, A., Murat, P., Pranko, A. & Safonov, A. A new mass reconstruction technique for resonances decaying to $\tau\tau$. *Nucl. Instrum. Meth. A* **654.** arXiv: 1012.4686 (2011).

118. CMS Collaboration. Search for new phenomena with multiple charged leptons in proton–proton collisions at $\sqrt{s} = 13\,\text{TeV}$. *Eur. Phys. J. C* **77.** arXiv: 1701.06940 [hep-ex] (2017).

119. CMS Collaboration. *CMS Luminosity Measurements for the 2016 Data Taking Period* CMS Physics Analysis Summary CMS-PAS-LUM-17-001 (2017). https://cds.cern.ch/record/2257069.

120. CMS Collaboration. *CMS luminosity measurement for the 2017 data-taking period at* $\sqrt{s} = 13$ *TeV* CMS Physics Analysis Summary CMS-PAS-LUM-17-004 (2018). https://cds.cern.ch/record/2621960.

121. CMS Collaboration. *CMS luminosity measurement for the 2018 data-taking period at* $\sqrt{s} = 13$ *TeV* CMS Physics Analysis Summary CMS-PAS-LUM-18-002 (2019). https://cds.cern.ch/record/2676164.

122. Junk, T. Confidence level computation for combining searches with small statistics. *Nucl. Instrum. Meth. A* **434.** arXiv: hep-ex/9902006 [hep-ex] (1999).

123. Read, A. L. Presentation of search results: the CL_s technique. *J. Phys. G* **28** (2002).

124. Cowan, G., Cranmer, K., Gross, E. & Vitells, O. Asymptotic formulae for likelihood-based tests of new physics. *Eur. Phys. J. C* **71.** [Erratum: doi:10.1140/epjc/s10052-013-2501-z]. arXiv: 1007.1727 [physics.data-an] (2011).

125. ATLAS and CMS Collaborations, LHC Higgs Combination Group. *Procedure for the LHC Higgs boson search combination in Summer 2011* tech. rep. ATL-PHYS-PUB/2011-11, CMS NOTE-2011/005 (2011). http://cdsweb.cern.ch/record/1379837.

126. Delaunay, B. N. Sur la sphère vide. À la mémoire de Georges Voronoï. *Bulletin de l'Académie des Sciences de l'URSS* **6.** https://zbmath.org/?q=an:0010.41101|60.0946.06 (1934).

127. Aberle, O. *et al. High-Luminosity Large Hadron Collider (HL-LHC): Technical design report* https://cds.cern.ch/record/2749422 (2020).

128. Contardo, D., Klute, M., Mans, J., Silvestris, L. & Butler, J. *Technical Proposal for the Phase-II Upgrade of the CMS Detector* tech. rep. (2015). https://cds.cern.ch/record/2020886.

129. CMS Collaboration. *The Phase-2 Upgrade of the CMS Tracker* tech. rep. (2017). https://cds.cern.ch/record/2272264.

130. CMS Collaboration. *The Phase-2 Upgrade of the CMS Barrel Calorimeters* tech. rep. (2017). https://cds.cern.ch/record/2283187.

131. CMS Collaboration. *The Phase-2 Upgrade of the CMS Endcap Calorimeter* tech. rep. (2017). https://cds.cern.ch/record/2293646.

132. CMS Collaboration. *The Phase-2 Upgrade of the CMS Muon Detectors* tech. rep. (2017). https://cds.cern.ch/record/2283189.

133. Colaleo, A., Safonov, A., Sharma, A. & Tytgat, M. *CMS Technical Design Report for the Muon Endcap GEM Upgrade* tech. rep. (2015). https://cds.cern.ch/record/2021453.

134. CMS Collaboration. *A MIP Timing Detector for the CMS Phase-2 Upgrade* tech. rep. (2019). https://cds.cern.ch/record/2667167.

135. CMS Collaboration. *The Phase-2 Upgrade of the CMS Level-1 Trigger* tech. rep. (2020). https://cds.cern.ch/record/2714892.

136. CMS Collaboration. *The Phase-2 Upgrade of the CMS Data Acquisition and High Level Trigger* tech. rep. (2021). https://cds.cern.ch/record/2759072.

137. Baber, M. *et al.* Development and testing of an upgrade to the CMS level-1 calorimeter trigger. *Journal of Instrumentation* **9** (2013).

138. Hoecker, A. *et al. TMVA - Toolkit for Multivariate Data Analysis* 2009. arXiv: physics/0703039 [physics.data-an].

139. Xilinx. *Vivado Design Suite User Guide: High-Level Synthesis* https://www.xilinx.com/support/documentation/sw_manuals/xilinx2019_2/ug902-vivado-high-level-synthesis.pdf.

140. Antcheva, I. *et al.* ROOT — A C++ framework for petabyte data storage, statistical analysis and visualization. *Computer Physics Communications* **180.** http://root.cern.ch (2009).

141. Broyden, C. G. The Convergence of a Class of Double-rank Minimization Algorithms 1. General Considerations. *IMA Journal of Applied Mathematics* **6.** eprint: https://academic.oup.com/imamat/article-pdf/6/1/76/2233756/6-1-76.pdf (1970).

142. Fletcher, R. A new approach to variable metric algorithms. *The Computer Journal* **13.** eprint: https://academic.oup.com/comjnl/article-pdf/13/3/317/988678/130317.pdf (1970).

143. Goldfarb, D. A family of variable-metric methods derived by variational means. *Math. Comp.* **24.** eprint: https://www.ams.org/journals/mcom/1970-24-109/S0025-5718-1970-0258249-6/S0025-5718-1970-0258249-6.pdf (1970).

144. Shanno, D. F. Conditioning of quasi-Newton methods for function minimization. *Math. Comp.* **24.** eprint: https://www.ams.org/journals/mcom/1970-24-111/S0025-5718-1970-0274029-X/S0025-5718-1970-0274029-X.pdf (1970).