Search for a resonance decaying into a quark top and a Higgs boson in the same sign dilepton final state using Run 2 data with the CMS experiment at LHC

Benjamin Alexandre Blancon

To cite this version:

Benjamin Alexandre Blancon. Search for a resonance decaying into a quark top and a Higgs boson in the same sign dilepton final state using Run 2 data with the CMS experiment at LHC. Physique Atomique [physics.atom-ph]. Université Claude Bernard - Lyon I, 2024. Français. NNT : 2024LYO10229 . tel-04894782

HAL Id: tel-04894782
https://theses.hal.science/tel-04894782v1

Submitted on 17 Jan 2025

Université Claude Bernard Lyon 1

THÈSE de DOCTORAT DE L'UNIVERSITÉ DE LYON
Opérée au sein de :
l'Université Claude Bernard Lyon 1

Ecole Doctorale 52
Physique et Astrophysique

Spécialité de doctorat : Physique des particules

Soutenue publiquement le 21/11/2024, par :
Benjamin Alexandre Blancon

Search for a resonance decaying into a top quark and a Higgs boson

in the same sign dilepton final state using Run 2 data with the CMS experiment at the LHC

Devant le jury composé de :

Anne Ealet Directrice de recherche, Institut de Physique des Deux infinis de Lyon	**Présidente de jury**
Elisabetta Gallo Professor, University of Hamburg	**Rapporteure**
Pascal Vanlaer Professor, Université Libre de Bruxelles	**Rapporteur**
Éric Chabert Maître de conférences Habilitation à Diriger des Recherches, Université de Strasbourg	**Examinateur**
Aldo Deandrea Professeur des universités, Université Claude Bernard Lyon 1	**Examinateur**
Frédéric Déliot Directeur de recherche, Institut de recherche sur les lois fondamentales de l'Univers	**Examinateur**
Abideh Jafari Assistant Professor, Isfahan University of Technology	**Examinatrice**
Nicolas Chanon Chargé de recherche Habilitation à Diriger des Recherches, Institut de Physique des Deux infinis de Lyon	**Directeur de thèse**

Remerciements

Ça y est, c'est la fin. Après tant d'années de labeur à résister, à tenir pour faire ce que j'aime, je vois enfin le bout du tunnel. J'aimerais dire que tout a été merveilleux, mais ce fut loin d'être le cas. Tant de moments ont été difficiles, tant de fois l'envie d'abandonner a empli mon cerveau, mais je me suis toujours accroché à la branche face au vent si violent et cinglant que peut être la vie. J'ai tellement souffert pour en arriver là, mais Dieu que ça valait le coup. Je suis tellement heureux, tellement fier, tellement content de moi, de tout ce que j'ai pu entreprendre dans cette aventure, de toutes ces émotions, de toutes ces rencontres. J'ai vécu des choses tellement fortes que malgré toutes les difficultées associées, je souhaite que tout le monde puisse vivre ce que j'ai vécu. Mais comme cette section s'appelle les remerciements, je dois dire que tout cela ne serait jamais arrivé sans les personnes dont je vais parler dès le prochain paragraphe. Enlever une seule de ses personnes équivaudrait à ne pas aller jusqu'au bout de cette thèse, et ce peu importe quand notre relation a débuté, c'est pourquoi je me dois de parler de vous tous. Spoiler : ça va être long.

J'aimerais tout d'abord remercier Stéphanie pour m'avoir proposé ce sujet. J'ai passé trois années fantastiques à être plongé au cœur du monde de la recherche et de la physique des particules grâce à toi, et je ne l'oublierai pas. Je n'oublierai pas aussi que grâce à toi, j'ai pu voir ce que c'était que d'être un piètre chercheur, une personne qui regroupe tout ce que je déteste dans l'être humain, et comment cette rencontre m'a permis d'un côté de perdre de l'espérance de vie en te côtoyant, mais aussi de gagner du temps en voyant que ce monde de la recherche n'était pas fait pour moi s'il n'était rempli que de gens comme toi (ce qui est fort heureusement n'est pas le cas). Je suppose donc que je peux aussi te remercier pour cela.

Passons donc maintenant aux vrais remerciements ! Je dois évidemment commencer par mon directeur de thèse, Nicolas. Tu as vraiment repris la barre de ce bateau sur lequel était écrit "Thèse de Benjamin Blancon" qui était en train de couler, écopé de toute ton énergie, colmaté lers brèches, et redressé la structure entière pour en faire un magnifique navire. Le début fut long à se mettre en place car ni toi ni moi n'avions prévu cela, mais tu m'as permis de reprendre confiance en moi, me calmer aux moments où j'avais besoin d'être rassuré, m'aider dès que tu le pouvais, tout cela en gardant ta gentillesse, ta bienveillance et ton professionnalisme qui te caractérisent si bien. Pour tout cela, j'aimerais te dire merci mille fois, et merci encore, car je n'y serais jamais arrivé sans toi le premier, toi qui a toujours été aux avant-postes de ma réussite.

Continuons avec les gens du laboratoire ! Je dois commencer par mon trio de l'amour, mon trio de cœur qui m'a soutenu et suivi partout pendant ces trois années, à savoir Elsa, Florencia et Grégoire. Vous avez été présents de la première à la dernière seconde, malgré mon caractère de cochon, mes blagues de merde et mon attitude globalement insupportable. Vous avez toujours été les premier supports et je ne saurais quoi dire de plus à part merci. J'ai une pensée toute particulière pour Jieun également, qui a été pendant deux ans dans la même galère que moi, et sans qui je crois que ni l'un ni l'autre n'aurions tenu, alors 감사합니다. Je pense bien évidemment aussi à Christopher qui a paratagé mon bureau pendant ces trois années, Sarah, Jad (شكراً جزيلاً I hope this is not wrong), David qui a partagé mon bureau pendant deux ans de son côté, Élise, Téo, Yann, Alfred, Charles, Inès, François (même s'il était dans le labo d'à côté), Elisa (Grazie mille), Anna (Thank you!), Yasmine, Paul, Tatiana, Wanda, Mohamad (شكراً جزيلاً I hope this is not wrong) et Maxime (même s'il est lui aussi dans le labo d'à côté). Vous êtes des gens fabuleux que je garderai au fond de mon âme pour un très long moment. J'aimerais aussi remercier l'équipe CMS du labo : Candan (Teşekkür ederim) et Arnab (আপনাকে ধন্যবাদ I hope this is not wrong) qui m'ont sauvé les fesses à peu près quatre fois par jour avec le code, Jie (谢谢 I hope this is not wrong), Suzanne, Muriel et Gaëlle toujours à l'écoute, vous avez tous répondu présents à toutes mes interrogations et questions en tout genre, alors merci. Merci aussi à Luc qui a toujours été présent pour répondre à toutes mes questions les plus inutiles sur la partie théorique de cette thèse. Une pensée spéciale à Paul également, qui j'espère sera reconnu le plus rapidement possible comme un artiste incroyablement talentueux, fasciné, et curieux de tout, merci à toi.

Au-delà du groupe et des doctorants, j'ai énormément de monde à remercier pour la gestion globale de la thèse. Tout d'abord, j'aimerais remercier infiniment Anne, Dany, Corinne (la queen qu'elle pense être) et Antoine pour leur implication infinie dans mes problèmes de thèse. Vous avez toujours été disponible et à l'écoute, vous n'avez jamais remis en cause mes dires, m'avaient soutenu instantanément et agi rapidement et efficacement, et putain qu'est-ce que ça fait du bien d'être bien entouré au niveau de la direction. Vous êtes l'une des principales raisons pour laquelle je n'ai pas explosé en plein vol en deuxième année, alors merci du fond du cœur. J'ai une énorme pensée pour le personnel administratif, à savoir Céline qui m'a sauvé les fesses pour absolument toutes les missions que j'ai faites, Florence, Déborah que j'ai harcelée en permanence avec une quantité astronomique de mails, Sandra et Sybil, qui est l'une des personnes les plus formidables et humaines de ce laboratoire et une source de bonheur constante. Merci à Bruno et Denis pour tout ce qui est gestion informatique de cette thèse qui est plus que primordiale, je pense que je ne vais pas tarder à ériger des statues à votre effigie en signe de ma gratitude.

Passons à la collaboration CMS ! C'était un immense plaisir d'échanger avec autant de monde pendant ces trois années sur autant de sujets différents. J'ai appris une quantité astronomique d'informations sur le fonctionnement du CERN et du LHC en général, ainsi que de la recherche au niveau national et international. Ce fut un chemin semé d'embûches par moment, mais rattrapé par de multiples rencontres avec des scientifiques venant du monde entier. J'ai une pensée particulière pour l'équipe de Hambourg avec Di (谢谢 I hope this is not wrong), Elisabetta (Grazie mille) et Rainer (Danke schön) qui ont toujours eu une écoute très attentive sur mon travail et avec qui j'ai hâte d'aller jusqu'à la publication, ainsi que l'équipe de Strasbourg avec Anne-Catherine qui a été un soutien permanent durant ces trois ans, Raphaël que j'apprécie de tout mon cœur, Océane, Paul, Dylan, Caroline, Jérémy et Éric. Une pensée particulière pour Deborah (Grazie mille), qui fut un support crucial à bien des moments durant cette thèse, j'espère qu'on pourra rester en contact le plus possible ! Je fais un tour du LHC pour aller remercier Alberto (Grazie mille), qui fut une rencontre formdiable et un ami sans faille depuis notre rencontre en première année, merci à toi.

Avant de passer aux proches d'un point de vue personnel, je dois bien évidemment parler de mon jury, et le remercier pour avoir relu et commenté de manière très pertinente un manuscrit bien trop long et barbant pour ce qu'il est. J'aimerais remercier Anne d'avoir accepté de présider ce jury, Elisabetta et Pascal d'avoir très gentiment accepté d'être les rapporteurs de cette thèse, ainsi qu'Aldo, Éric, Frédéric qui a été présent dans ce jury et bien plus durant toute la thèse, et Abijeh (خیلی سپاسگزارم I hope this is not wrong) pour avoir accepté d'être examinateurs. J'espère que vous avez pris autant de plaisir en lisant cette thèse que j'ai eue en l'écrivant.

La première partie des remerciements étant terminée, il est temps pour moi de passer aux remerciements pour ceux qui ont été là bien avant la thèse, et qui continuent toujours d'être présents jusqu'à aujourd'hui. Comment ne pas commencer par maman et papa ? Que de soutien depuis toujours et encore, d'inquiétude plus ou moins cachée, mais toujours présents malgré vents et marées. Tu vois maman, je pense qu'on pourra dire à tous ceux qui t'ont dit que tes enfants feraient rien de leur vie d'aller bien se faire voir. Je continue bien sûr avec mes quatre reines, Christelle, Sandra, Fanny et Delphine. Toutes différentes, toutes exceptionnelles, toutes indispensables dans mon équilibre tendu que constitue ma vie. Je vous aime infiniment. J'ai une pensée bien sûr pour ceux qui partagent vos vies, que ce soit les compagnons ou les enfants qui grandissent bien trop vite quand on est adultes. Je pense également bien sûr à tous ceux qui gravitent autour même s'ils ne sont pas frères ou sœur de sang, notamment Mickaël qui a été une source d'inspiration phénoménale pour moi, Armand et Guillaume aussi taquins qu'humains, et Manon qui fait toujours autant de bien autour d'elle.

Léa. Léa, Léa, Léa. J'aimerais écrire dix pages rien qu'avec ton prénom pour que les gens se rendent compte de l'importance extraordinaire que tu as dans ma vie. Mon soleil, mon modèle, mon épaule, mon ange gardien, ma psychologue, mon humoriste préférée, ma conseillère attitrée. Ma vie tout simplement. Tu es la perfection de l'être humain à mes yeux, tu es la raison pour laquelle j'ai compris que la vie méritait d'être vécue, tu es tout pour moi et bien plus encore. Je t'aime comme je n'ai jamais aimé quelqu'un avant, et comme je n'aimerais sûrement jamais quelqu'un d'autre. Tu me permets de me transcender et de devenir la meilleure version de moi-même à chaque fois que tu es là. Quand je suis avec toi, toutes mes peurs s'évaporent et je ne suis qu'un être sans crainte ni stress. Tout est si simple quand je te parle, les problèmes se résolvent en un rien de temps comme s'ils n'avaient jamais existé. Et malgré tout ça, la chose la plus belle dans notre relation est que je sais à quel point ce constat est réciproque (sauf peut-être avec Antony, mais bon ça compte pas (et je t'embrasse bien sûr également)). Tu es mon autre, je suis ton autre. Je souhaite à tout le monde de trouver sa propre Léa dans la vie, et de la garder au plus profond de son cœur et de son corps. Que j'ai hâte de poursuivre toute ma vie à tes côtés, en respectant cette promesse faite le 28 février 2015. Tout ce qui gravite autour de toi baigne dans cette même perfection, comme l'est toute ta famille. Michel, Sarah, Maëlle et Lilian : la famille James est à la famille Potter ce que Sirius est à moi, un deuxième havre de paix. Un lieu pur, sain, sans jugement, avec la porte toujours ouverte. Je sais que je fais partie de plusieurs familles sur ce monde, mais vous êtes très clairement mon deuxième foyer, celui que j'ai choisi, et rien ni personne ne pourra me l'enlever. Je vous aime divinement, et Dieu que vous me le rendez bien.

Concernant le troisième foyer, je pense que le nom est tout trouvé pour parler de celui du lycée Jeanne d'Arc. Au-delà de Léa que je pense avoir saucé sur assez de lignes juste au-dessus, ce fut le lieu de tellement de rencontres fantastiques que c'est avec un plaisir non dissimulé que je peux parler de vous tous. Commençons par Étienne, avec qui j'aurais partagé ces trois années de la seconde à la terminale, et qui est la première personne avec qui j'aurais discuté en ce premier cours de SVT sans jamais le quitter : merci pour ton intelligence, ta folie et ta candeur incarnée. Merci également à Heloïse, fougueux et impétueux petit bout de femme avec qui j'adore passer de longues après-midi à échanger sur des sujets aussi larges que ne l'est ta curiosité. Doriane, que dire de toi, à part que tu es l'une des personnes les plus marquantes que je n'ai jamais rencontrées et ce dès le premier regard, une bombe d'humour et de sincérité qui fait toujours du bien au moral et au cœur. Tu me rappelles en bien des points Manon, arrivée deux ans après dans ce même lycée, avec la même énergie et la même joie de vivre : vous êtes toutes les deux de formidables êtres à garder près de soi. Comment ne pas parler de Charlotte également, my Baby Girl, des cours d'italien à des heures beaucoup trop tardives pour que ce soit légal à des conversations animées sur nos vies respectives ? Ta présence est toujours synonyme de folie, mais aussi de bienveillance et de soutien sans aucun jugement. Le constat est un peu le même que pour Laurie finalement, une future folle qui l'est déjà bien assez vu sa passion pour l'anglais, mais qui saura toujours être présente dans les moments les plus simples comme les plus durs à surmonter. C'est un réel honneur de vous compter parmi mes amis les plus proches. C'est aussi grâce à toi Laurie, que j'ai pu rencontrer Camille, avec qui j'ai partagé tant de discussions sur la vie, et qui aura toujours garder son honnêteté qui fait tant de bien, puis Hugo plus tard, qui même si je ne le vois que peu, me marque toujours par sa gentillesse et son calme dans n'importe quelle situation. Je ne peux pas non plus ne pas parler d'Esther, rencontrée avant même le lycée, et qui aura été une personne si importante au lycée et même au-delà, toujours présente pour discuter autour d'un verre ou simplement en face à face, avec sa pétillance et sa fraîcheur que je lui connais depuis le premier jour. Tant de personnes auront été précieuses durant cette période si charnière pour mon existence (Pauline, Hannah, Naelwenn, Laëtitia, Laura, François, Jérémy, Emma, Rémi, Nathan, Hanna, Charlène, Coralie, Julie... tous se reconnaîtront) : merci pour tous ces moments partagés ensemble, ils ne disparaîtront jamais. Des remerciements tous spéciaux à Julien (oui, je n'ai rien écouté à ce cours sur les conquêtes espagnoles où vous étiez en chemise blanche à moitié transparente et non, je ne regrette rien) et Gaëla qui m'auront conforté dans mon choix de venir professeur : vous êtes la raison même pour laquelle j'ai fait ce métier, une source d'inspiration énorme pour mon futur moi et des futurs collègues qui seront très chers à mon cœur.

Après le lycée, les deux premières années d'université à Clermont furent laborieuses, mais ô combien riche en rencontres pérennes. Tout d'abord, j'aimerais remercier Adèle d'avoir été présente en de si nombreux moments et de m'avoir permis d'échapper à mes problèmes si lourds à porter durant cette période. J'aimerais remercier également Jeanne pour ce même investissement qui fut plus que nécessaire. De manière générale, j'aimerais remercier tout ce groupe de nanas (Marie, Yana, Aurélie, Eugénie et Margot) qui m'ont permis de tenir la barre durant la galère qu'a représenté le début de cette licence.

Puis arrive l'ENS. Pas grand-chose de positif à retenir de ces années malheureusement, sauf quelques personnes qui m'ont permis de tenir durant toutes ces années. Je commence é-vi-dem-ment par le seul, l'unique, le plus grand parmi tous les plus grands, sa majesté, son altesse Vincent, seul personne à pouvoir prononcer un ornithopiccoloelatybus à gorge dorée sans mourir en plein milieu. Définitivement la personne la plus drôle qui n'ait jamais foulée cette Terre, ainsi que l'une des personnes les plus belles, bienveillantes et incroyables que je n'ai jamais rencontrées. Nous avons passé tellement d'heures à discuter d'absolument tout ce dont il est possible, même sur des sujets qui t'emmerdent profondément. Tu es une personne fondamentalement exceptionnelle, un rappel quotidien que oui, une personne aussi inspirante peut exister, et qu'on peut avoir la chance de la côtoyer tous les jours. Je pense que je ne te dirai jamais assez à quel point tu es important pour moi, et je souhaite être à tes côtés le plus loin possible, notamment lorsque tu seras directeur de l'opéra de Paris en 2046. Je pense également à Valérian, hilarant en permanence et surtout sans le faire exprès, qui fut un soutien à bien des moments lorsque j'avais besoin de péter un câble entre deux cours. Enfin, je me dois de parler de toi, Corentin. Je ne sais pas où tu es, je ne sais pas si tu liras cela un jour, mais sache que tu es l'une des deux personnes qui m'a permis de me sortir de la dépression, et je ne sais honnêtement pas si je serais encore de ce monde sinon. Je pense à toi tous les jours encore, je suis triste de voir comment tout cela s'est arrêté, mais je pense que c'était la meilleure solution pour nous deux. Sache simplement que je t'aime de toute mon âme également, et que tu auras toujours une place dans mon cœur pour avoir été aussi important à un moment si dur de ma vie. Tu auras été la personne qui m'aura permis de reprendre le goût à la vie, et je ne l'oublierai jamais.

Et puis, Trivia arriva. Tu auras été cette deuxième personne qui m'aura sorti de la dépression, lorsque les nuits furent si propices à plonger dedans. Grâce à toi, j'ai découvert une quantité astronomique de personnes toutes plus formidables les unes que les autres, à commencer par toi. Tant de moments passés en ta compagnie à distance, de rires, de joie, de larmes. J'ai réussi à remonter la pente avec le temps, et te rencontrer fut l'une des choses les plus fabuleuses qui me soit arrivé. Tu es une personne admirable, drôle, sincère, tellement équilibrée, avec une bienveillance et une bonté sans pareil. J'espère être toujours à tes côtés et auprès de toute les baleines. À commencer par toi, mon Gurvan. J'ai tellement grandi et compris de choses grâce à toi. Qu'est-ce que tu es chiant, mais qu'est-ce que j'aime être à tes côtés. Ta présence ces dernières années vaut bien tous les médicaments du monde, et je ne serai jamais assez reconnaissant de cela lorsque les moments furent compliqués pour toi comme pour moi. Je t'aime infiniment, et j'ai hâte de partager le reste de ma vie avec toi (dans le jeu). J'aimerais remercier Fanny aussi, qui partage (subit) la discussion de crackhead avec nous deux, même si elle en tient une bonne couche elle aussi. Merci pour tous ces moments partagés ensemble, tu es ce type de personnes dont on sait instantanément qu'elle est fabuleuse, et qu'il faut les garder près de soi. Merci également à LVV avec tes avis désastreux, Val pour être le goat, Amyrith pour être aussi le goat, Dibou pour être aussi le goat (bon courage pour la DA), Nazu que j'aime d'amour, Daxou pour nos conversations bizarres, Bruck aussi pour nos conversations bizarres, Axel pour être le goat (oui je l'ai déjà dit), William pour me faire rire simplement en respirant, Indo pour être folle, Latyfre pour être un goat lui aussi, Loyso pour tes takes désastreuses sur ta ville, Krokro pour ton aigreur légendaire, Jeanne pour être toujours d'aussi bonne compagnie, Bénédicte pour ta vivacité d'esprit si précieuse, Zzaelde pour être le goat (l'ai-je déjà dit ?), Igers mon virtual DADDY, Klikli avec tes takes désastreuses sur ton domaine, Thiby pour ton humour toujours on point, Dicen pour être le goat (on va arrêter de compter hein), Truite pour être le goat, Salia pour être la goat, Marmiza pour ne parler que d'histoire et de politique avec un ton désabusé permanent, Marion ma viiiiiie, Poof le fou du bus, Sam pour m'accompagner dans tous les plans ciné foireux, Bouah pour être la goat, Julien 1 et 2 pour toujours vous confondre, Amped pour être notre requin préféré, Bestlaf pour être le goat aigri, Karlemen pour être si cool sans le faire exprès, Parox pour être mon futur mari, Saso pour être le BG ultime, Rammstein pour toutes ces conversations, Touracoq pour tes takes désastreuses sur ton domaine aussi, Lyes pour être la force tranquille, Morghall pour être le goat, Tarmil pour pouvoir dire que les vrais savent que c'est Node, Baguette pour tous tes conseils, Xeno pour être le goat, Sangmaitre pour être le goat, Uncon pour être le fou du bus originel, Luigi pour toutes ces fois où j'ai dormi chez toi, Loony pour toutes ces crises de fou rire, Ant même si on a loupé le coche pour se voir, Dagnir pour être le goat, Corvo pour être le goat, Chou-fleur parce que j'aime beaucoup trop dire ça, Elsa pour être le goat, Iosa pour être le goat et Le maître (du désastre sinon c'est bizarre) pour être le goat final. J'ai un peu la flemme de faire un paragraphe sur chacun, et que le but de ce manuscrit est quand même de parler de physique à un moment, mais sachez que je vous porte tous à un endroit très très particulier dans mon cœur (oui, il ne va pas tarder à exploser). Vous êtes tous acteurs dans la réussite de cette thèse, et croyez-moi que vous êtes bien plus importants que vous pensez l'être. Je vous aime d'amour. Un remerciement spécial à Galba, pour être le père que tout le monde veut avoir, et pour avoir été le premier à me faire plonger dans cette océan de bonheur dont je ne suis jamais sorti, merci à vous pour tout ce que vous êtes et ce que vous faites.

Enfin (c'est le dernier paragraphe promis), je vais terminer avec toutes ces personnes qui suivent mon quotidien de loin, sur Twitch, et qui permettent de rendre chaque journée bien plus agréable que sans leur compagnie. Tout d'abord, je tiens à remercier JDG, et Fred en particulier. Je ne sais pas si tu liras ça un jour par un hasard improbable, mais sache que tu as toujours été un modèle incroyable pour moi, une source d'inspiration inépuisable et une preuve qu'on peut être en adéquation avec soi-même dès le plus jeune âge en faisant asbtraction de tous les abrutis qui nous entourent. Je pense que si jamais j'apparais un jour dans un JDG, je risque de faire une syncope avant, mais sachez que c'est l'un de mes plus grands rêves pour pouvoir vous remercier de m'accompagner dans ma vie depuis plus de 13 ans. Merci également à toute la team du lundi : MV pour être l'une des personnes les plus inspirantes sur le net, Antoine pour sa bonne humeur perpétuelle, Florence pour être mon Auvergnate préférée même si tu es vraiment nulle à n'importe quelle jeu, Baghera et Horty pour être les folles du bus qui rendent les soirées si mémorables, Zera pour être si impressionnant dans tout ce que tu entreprends et Mynthos pour être toujours l'aigri fatigué qui fait vriller chaque discussion. Je ne pense pas que vous vous rendez compte à quel point vous êtes devenus un rendez-vous incontournable qui fait autant de bien aux gens qui le regardent, alors merci infiniment. J'aimerais également remercier Damdam pour ta vie et ta personne si passionnante à découvrir et soutenir, Farore pour être si exceptionnelle et vivace sur le monde qui t'entoure, Margot pour être la personne la plus drôle de tout Twitch (et une personne incroyable au passage), Shisheyu pour me faire rire en permanence, Ultia pour être la queen que tu penses être, Riv pour m'avoir donné beaucoup d'idées pour le futur, les Dingus, et plein d'autres. J'ai une pensée pour l'équipe d'OOTR, notamment le quatuor de folie qu'est Papy, Zorro, Camille et Menou. J'aurais passé bien des soirées et des nuits à rattraper tous les tournois lorsque le moral n'était pas au beau fixe, et ce ne fut clairement pas de trop. Enfin, merci à Zul pour m'avoir permis de comprendre que ça n'allait pas dans ma vie et de voir comment je pouvais remonter la pente et devenir ce que je suis aujourd'hui. J'arrête peut-être la physique des particules ici en ayant tant été impliqué dans ce domaine de l'intérieur, mais ce n'est pas la dernière fois que vous entendrez parler de moi et de recherche !

Merci à tous. Merci pour tout. Merci.

I'm the fury in your head, I'm the fury in your bed, I'm the ghost in the back of your head.

Résumé

Cette thèse s'intéresse à la recherche d'une nouvelle particule à ajouter au Modèle Standard pour résoudre le problème de hiérarchie dans le secteur du boson de Higgs. Le Modèle Standard est la théorie la plus précise de la physique des particules, classifiant toutes les particules fondamentales ainsi que leurs interactions. L'interaction des particules avec le champ de Higgs, caractérisé par le boson de Higgs, leur donne notamment leur masse. Cependant, la masse du boson de Higgs n'est pas expliquée par le Modèle Standard. Ce problème, communément appelé le problème de hiérarchie, peut être résolu en ajoutant des particules au Modèle Standard.

Nous nous focalisons sur l'ajout d'un hypothétique "Vector-Like Quark" T' au Modèle Standard. Nous nous intéressons en particulier à la désintégration de la particule T' en un boson de Higgs et un quark top dans un état final dileptonique même signe. La recherche de cette particule s'effectue avec les données de la période dite du "Run 2" (2016-2018) au sein de l'expérience CMS, l'une des quatre expériences principales du LHC. L'énergie du centre de masse étudiée ici de $\sqrt{s} = 13$ TeV pour une luminosité intégrée totale de $L_{int} = 137.6$ fb^{-1}. L'étude de cette désintégration n'a jamais été effectuée auparavant, c'est pourquoi une stratégie d'analyse complète a dû être mise en place pour identifier la présence de la particule T'.

La thèse présente l'intégralité de la stratégie d'analyse en cinq étapes. La première étape consiste à identifier les différentes particules qui constituent l'état final. La deuxième étape définit la région de signal dans laquelle la signature de la désintégration de la particule T' doit être identifiable. De nombreuses études ont été menées concernant la liste des coupures, le choix de la variable discriminante, ainsi que l'optimisation de la sélection pour conserver la forme de la distribution du bruit de fond coupure après coupure. La troisième étape définit les régions de contrôle et de mesure pour estimer la distribution des bruits de fond principaux de la région de signal. Une région de contrôle et trois régions de mesure sont définies ici. La quatrième étape implémente les différentes incertitudes, divisées en trois sources : les incertitudes systématiques expérimentales, les incertitudes systématiques théoriques et les incertitudes statistiques.

La dernière étape évalue les limites d'exclusion attendues sur la section efficace en l'absence de signal avec un niveau de confiance de 95%, ainsi que la significance attendue en fonction de la masse nominale de la particule T'. Les résultats présentés ont une sensibilité similaires à ceux obtenus dans des états finaux parallèles à celui étudié ici, avec une section efficace attendue variant de 1400 à 400 pb pour une masse nominale de la particule T' allant de 600 à 1200 GeV. À la suite de cette thèse, les résultats obtenus avec les données simulées seront comparés aux vraies données enregistrées pour évaluer la potentielle existence du Vector-Like Quark T'.

En parallèle de cette analyse, un nouveau chemin du système de déclenchement du détecteur CMS a été caractérisé pour la désintégration du Vector-Like Quark T' en un boson de Higgs et un quark top dans un état final tout hadronique. Ces études s'effectuent dans le cadre du Run 3 (2022-2025). Les résultats préliminaires obtenus avec les données de 2022 et 2023 ont permis une amélioration dans les propriétés du chemin, permettant une prise de données optimale jusqu'en 2025.

Summary

This thesis is centered on the search for a potential new particle which could be integrated into the Standard Model to address the hierarchy problem in the Higgs boson sector. The Standard Model is the most precise theories in particle physics, classifying elementary particles and their interactions. These particles acquire mass through the interaction with the Higgs field, characterized by the Higgs boson. However, the Higgs boson mass cannot be explained with the Standard Model. This problem, known as the hierarchy problem, can be solved by adding particles to the Standard Model.

We propose to incorporate a hypothetical 'Vector-Like Quark' T' into the Standard Model. Specifically, we will investigate the decay of the particle T' into a Higgs boson accompanied with a top quark, focusing on the same sign dilepton final state. This search uses data collected during the Run 2 period (2016-2018) by the CMS detector, one of the four main experiments at the LHC. The center-of-mass energy considered here is $\sqrt{s} = 13$ TeV, and the integrated luminosity is $L_{int} = 137.6$ fb^{-1}. As this final state has never been explored before, a new analysis strategy was developed to identify the presence of the particle T'.

The thesis presents the comprehensive analysis strategy in five key steps. The first step involves the identification of the different particles which constitute the final step. The second step focuses on the definition of the signal region, where the signature of the particle T' must be detected. This includes extensive studies on the cut-based selection, the determination of the discriminant observable, and the optimization of the selection process to maintain control over the background shape after the full selection. The third step establishes the control and measurement regions for estimating the background shape of each major background process within the signal region. This includes one control region and three measurement regions. The fourth step integrates the uncertainties, categorized into three types: the experimental systematic uncertainties, the theoretical systematic uncertainties, and the statistical uncertainties.

The last step evaluates the expected exclusion limits on the cross section without signal with a 95% Confidence Level, and the expected significance as a function of the nominal mass of the particle T'. The results have a similar sensitivity to that of other final states, with an expected cross section ranging from 1400 to 400 pb for a nominal mass value of the particle T' ranging from 600 to 1200 GeV. Following this thesis, these simulation results will be compared with real data to assess the potential existence of the Vector-Like Quark T'.

In addition to this analysis, a new trigger path within the CMS collaboration has been characterized for the Vector-Like Quark T' decay into a Higgs boson accompanied with a top quark in the all-hadronic resolved final state. These studies concern the Run 3 period (2022-2025). Preliminary results with 2022 and 2023 data have refined the properties of the trigger path, enabling optimal data acquisition through 2025.

Contents

Introduction

The Standard Model of particle physics, developed throughout the 20th century, has emerged as one of the most significant and coherent theories in the field of physics. This model classifies elementary particles and their interactions at the fundamental level, as it will be detailed in Chapter 1. The Standard Model has been extensively validated and refined over decades through dedicated experiments which have confirmed its theoretical predictions with remarkable precision. Despite its success, the Standard Model has known limitations, suggesting that it may be an effective low-energy approximation of a more comprehensive theory. Efforts to extend the Standard Model are known as 'Beyond the Standard Model' (BSM) theories. Searching for BSM physics is the goal of this thesis.

Among the observations not accounted for by the Standard Model, the mass of the Higgs boson remains inexplicable. The Higgs boson, the final particle predicted by the Standard Model, is responsible for the particles to acquire mass through their interaction with the Higgs field. Experimentally discovered in 2012 by both the ATLAS and CMS experiments at CERN, the Higgs boson mass has been measured at 125.25 ± 0.17 GeV/c^2. However, the Standard Model predicts that the Higgs boson mass should diverge at large energy scale due to quantum corrections. To address this inconsistency, numerous theories beyond the Standard Model have been proposed. Among these, some assume the existence of an additional particle within the Standard Model, known as a Vector-Like Quark. This hypothetical particle could resolve the hierarchy problem by interacting with the Higgs boson, a topic that will be further explored in Chapter 1.

To confirm this prediction, the experimental discovery of the Vector-Like Quark remains necessary. Accordingly, this thesis has been performed within the CMS experiment, which will be presented in Chapter 2. The CMS experiment is one of the four primary experiments of the Large Hadron Collider, based at CERN in Geneva. It is a high-energy particle detector with two main objectives: the precision measurement of particles properties as predicted by the Standard Model, and the search for new signatures as predicted by theories beyond the Standard Model. In such cases, the analysis typically focuses on the decay products of the particle under study, as will be done in this investigation of the Vector-Like Quark.

The search for Vector-Like Quarks has been ongoing since the Run 1 period in 2008 in both the ATLAS and CMS experiments. In 2016, the search for a specific Vector-Like Quark T' presented a deviation from the Standard Model predictions at the level of 3 standard deviations in the CMS experiment, for a nominal mass of the particle $m_{T'} = 680$ GeV. This analysis focused on the Vector-Like Quark decay into a Higgs boson accompanied with a top quark in a final state only composed of quarks, and known as the all-hadronic final state. Although subsequent studies aimed to identify the same signature in other final states, no significant findings were reported, as it will be discussed in Chapter 4. The objective of this thesis is to search for the signature of the Vector-Like Quark T' decay into a new final state, which has not been previously studied in either the CMS or ATLAS experiments.

We focus on the decay of the Vector-Like Quark T' into a Higgs boson accompanied with a top quark in a final state composed of two leptons which have the same electric charge, known as the same sign dilepton final state. The final state also includes three quarks, among which a bottom quark, and two neutrinos which are not directly identified in the CMS detector, contributing to the missing transverse energy observed. The targeted nominal masses of the Vector-Like Quark range from 600 GeV to 1200 GeV. A dedicated analysis strategy has been designed, with associated studies detailed in Chapters 4 and 5. Chapter 4 discusses the identification of the event selection criteria, the signature of the Vector-Like Quark decay, and the definition of the most discriminating observable among those tested to separate the signal from the background. Chapter 5 discusses the background estimation, and presents the results as the expected limits with a 95% confidence level on the cross section of the VLQ, along with the significance as a function of the nominal mass of the Vector-Like Quark. All the results are presented using simulated samples, as data remain 'blinded' before a future comparison between data and simulation. This final step will serve as one of the final cross-checks of the 2016 CMS result for the Run 2 period.

Concurrently with the Vector-Like Quark analysis, I undertook service tasks within the CMS collaboration for the Run 3 period, as discussed in Chapter 3. These tasks involved the development of a High Lever Trigger path designed to select particles produced in the decay of the Vector-Like Quark T' in the tH all-hadronic resolved final state at the trigger level, which represents the first step of any analysis selection within the CMS collaboration. The existing High Level Trigger path during the Run 2 period was not fully in harmony with this analysis because the event selection required 6 jets, while the studied final state provides 5 jets, necessitating the design of a new path. My work concentrated on evaluating the impact of the jet reconstruction algorithm, and the algorithm for identifying jets originating from the decay of a bottom quark. This work was conducted in collaboration with different groups of the CMS collaboration, and was refined throughout my thesis.

Chapter 1

The Standard Model of particle physics and beyond

The concept of matter has been introduced several thousand years ago in various parts of the world such as ancient India or Greece. Various philosophers claimed that matter consisted of fundamental elements like air, water, fire, earth, or even aether. Democritus was the first one around 400 B.C. to consider matter as a combination of tiny bodies called atoms (from ancient Greek $\alpha\tau o\mu o\nu$, *atomon*, i.e. 'indivisible, unbreakable'). This theory, known as atomism, assumed that atoms were the elementary bricks of everything. However, Aristotle's theory that matter is a universal substance capable of transforming from one form to another prevailed within the scientific community until the 17th century, with the work of physicists like Galileo [1]. But it was not until the 19th century that scientists could explain the behavior of gases and chemical reactions thanks to the understanding of atomic structure, and the Mendeleev's invention of the periodic table in 1869. The discovery of the electron in 1897 by Thomson, and the Rutherford's gold foil experiment in 1911 provided a more concrete model of the atom, depicting electrons orbiting a positively charged nucleus.

The birth of particle physics is inherently linked to the quantum mechanics revolution at the beginning of the 20th century, when Planck and Einstein postulated that light emission or absorption corresponds to discrete amounts of energy called *quanta*. Following the Bohr model which implemented the quantum nature of the electron within the atom in 1913, quantum atomic models began to multiply thanks to numerous discoveries in subatomic physics: discoveries of the neutron and positron in 1932, nuclear fission in 1938... All these results propelled physicists to classify particles according to their properties in what will be called the Standard Model (SM) in the 1960s. The final particle, the Higgs boson, was added in 1964 in the model with the Brout-Englert-Higgs (BEH) mechanism, which gives the particles their mass. After the final implementations of the SM at the end of the 20th century, the discovery of the Higgs boson in 2012 at the Large Hadron Collider (LHC), thanks to the A Toroidal LHC ApparatuS (ATLAS) and the Compact Muon Solenoid (CMS) experiments, completed one of the most precise theories in physics. However, many unknowns remain such as explaining the large asymmetry between matter and antimatter in the universe, or the nature of dark matter. The Higgs boson mass especially poses an issue for the SM, as the measured value at the LHC cannot be explained within the SM without a large amount of fine tuning. This is known as the hierarchy problem. New theories Beyond the Standard Model (BSM) propose adding new particles to the SM, such as Vector-Like Quarks (VLQs), as a possible solution to the hierarchy problem.

This chapter will be divided into two sections. The first section will focus on the SM, and the characteristics and differences between the two types of particles, known as the bosons and the fermions. Bosons are responsible for the transfer of energy between different particles as an expression of the fundamental forces included in the SM (weak, strong, and electromagnetic interactions), while fermions are elementary particles forming matter. All of these properties can be summarized in a Lagrangian as a mathematical formalism of the SM. The second section will explain the hierarchy problem in more detail, and how it is one of the most challenging problems for the SM. We will finally discuss about the VLQs, the current theories predicting their production and properties, and how they could solve the hierarchy problem.

1.1 The Standard Model

1.1.1 The bosons, mediators of the fundamental forces

Bosons are particles whose spin quantum number has an integer value. They obey Bose-Einstein statistics, allowing multiple bosons to have exactly the same quantum numbers [2].

The photon and Quantum Electrodynamics

Einstein's prediction of the photon in 1905 [3] appears to contradict Maxwell's electromagnetism theory, developed in 1860, which unified electric and magnetic phenomena. New corpuscular properties cannot be explained by a wave-based interpretation of electromagnetism. For instance, the fact that the energy of a light wave depends only on its intensity and not on its frequency cannot be accounted for. Dirac was the first one to describe this wave-particle duality for the electromagnetism, formulating the electromagnetic Lagrangian in 1927 (see section 1.1.3) [4]. The Lagrangian must respect the local symmetry of the electromagnetism, meaning it remains unchanged under spacetime-dependent transformations. This property is known as the gauge invariance of the Lagrangian.

The concept of gauge was initially introduced by Maxwell and Lorentz in the 19th century within the theory of electromagnetism. Electric and magnetic fields derive from an electromagnetic potential which is defined up to a constant. The choice of this constant provides a unambiguous definition of the potential, known as the gauge [5]. Gauge invariance was later introduced by Weyl in 1918 when he tried to unify electromagnetism with the theory of general relativity [6], then used by Dirac for the first formulation of the electromagnetic Lagrangian. However, when Pauli extended the invariance of the Lagrangian by implementing quantum field theory with Dirac and others [7], the Lagrangian must respect an additional symmetry, known as an internal symmetry. The combination of both local and internal symmetries defines the electromagnetic gauge of the Lagrangian.

Mathematically speaking, Quantum ElectroDynamics (QED) is described using the continuous unitary group $U(1)_{\text{em}}$, which is a Lie group. The transformation can be then interpreted as a phase change of the unit circle with a unique generator, making the group abelian by definition. We introduce the gauge field A_μ, which describes the photon γ. The photon is a gauge boson with a spin equal to 1. It has no electric charge Q_{em} and no mass, thus propagating at the speed of light $c = 299,792,458$ m.s^{-1} [8]. Its energy E is equal to $E = p \cdot c$ where p represents the norm of the momentum vector $\vec{p}$ of the photon.

The W and Z bosons and weak interaction

Weak isospin I_W^3 is a quantum number which differs by one unit between the neutron and the proton, and should be conserved during β decay experiments. Unfortunately, the QED is not sufficient to explain it. Hence, Yang and Mills extended the notion of gauge invariance to non-abelian groups in 1954 [9]. The group describing weak interaction is the continuous special unitary group $SU(2)_\mathrm{L}$, known as the weak isospin group. The three generators T_i ($i = 1, 2, 3$) are related to the usual Pauli matrices σ_i as follows:

$$T_i = \frac{\sigma_i}{2}, \;\text{ with } [\sigma_i, \sigma_j] = 2i\epsilon_{ijk}\sigma^k \tag{1.1}$$

as the commutative law for $SU(2)_\mathrm{L}$. The three gauge fields W_μ^i will describe the W^+, W^-, and Z bosons after the ElectroWeak (EW) symmetry breaking explained later [10]. These bosons are gauge bosons with a spin equal to 1. They have an electric charge equal of 1, -1, and 0 respectively. They are also the only massive gauge bosons as we will discuss with the BEH mechanism, leading to a lifetime of under 1.10^{-25} s [8].

The gluons and Quantum Chromodynamics

The theory of Quantum ChromoDynamics (QCD) was developed in the 1950s and 1960s to address the Coulomb repulsion between protons within atomic nuclei. If matter is bond together, it implies the existence of an additional force stronger than the electromagnetic force at short ranges. The concept of gluon was introduced by Gell-Mann in 1962 [11] as a 'glue' which keeps the quarks within hadrons together (see section 1.1.2). Building on the work of Yang and Mills in the weak interaction, Greenberg [12] and Han-Nambu [13] proposed a similar theory in QCD in 1965 which could be described by the continuous group $SU(3)_\mathrm{c}$. The eight generators t_A ($A = 1, ..., 8$) are related to the Gell-Mann matrices λ_i as follows:

$$t_A = \frac{\lambda_A}{2}, \;\text{ with } [\lambda_A, \lambda_B] = 2if_{ABC}\lambda^C \tag{1.2}$$

as the commutative law for $SU(3)_\mathrm{c}$, and f_{ABC} the structure constants generalizing the Pauli matrices for $SU(3)_\mathrm{c}$ [14]. The eight gauge fields A_μ^A describe the gluons g. Gluons are gauge bosons with a spin equal to 1. They have no electric charge and no mass.

Gluons can be described by states, known as the color states, and their respective color charge as an alternative interpretation of the Gell-Mann matrices [15]. Color charge is, like electric charge in QED and weak isospin in weak interaction, the quantum number which must be conserved in QCD. It is not related to the usual usage of color and charge, but simply becomes popular because of the analogy to primary colors. Each gluon carries one type of color ('blue', 'yellow', or 'red') and one type of anticolor ('antiblue' or 'orange', 'antiyellow' or 'purple', or 'antired' or 'green'), representing the color states.

Magnitude of the forces and β-function

Each force involves the coupling of a gauge boson with fermions or other bosons through a coupling constant, or coupling parameter g. But referring to g as a constant is inaccurate, as its value varies depending on the energy scale at which the force is evaluated. Higher-order corrections can lead to divergence unless they cancel each other in a gauge invariant theory thanks to the Ward-Takahashi identity [16, 17]. In that case, the Lagrangian must undergo renormalization, wherein the divergence is removed by introducing an energy-dependent coupling parameter. This dependence is typically represented in a β-function as follows:

$$\beta(g) = \frac{\partial g}{\partial \log \Lambda}, \tag{1.3}$$

where Λ is the energy scale.

The QED, weak interaction, and QCD exhibit distinct behaviors regarding the energy scale. In QED, the coupling parameter increases with the energy. Its value is typically around $g \approx \frac{1}{137}$ at low energies, and reaches approximately $g \approx \frac{1}{127}$ at the Z boson energy scale [18]. In QCD, the coupling parameter decreases with energy, and becomes sufficiently low at high energies to consider quarks as nearly free particles in deep inelastic scattering (DIS) (see the parton model in section 1.1.2). This is the asymptotic freedom discovered by Gross, Wilczek [19], and Politzer [20] in 1973. In weak interaction, the coupling parameter increases with energy but remains lower than those of QED and QCD at low energies, hence its name.

The Brout-Englert-Higgs mechanism

The gauge bosons discussed earlier represent the fundamental forces, excluding gravity, and the different theories are well understood within the scientific community. But one problem still remains: why are the W and Z bosons massive? The $SU(2)_{\mathrm{L}}$ group which explains the weak interaction assumed that the gauge bosons must be massless. However, they are found to be massive. On top of that, gauge bosons mass terms themselves would not be gauge invariant (see section 1.1.3), and the Lagrangian is not renormalizable.

In 1957, The Bardeen-Cooper-Schrieffer (BCS) theory introduced the idea of spontaneous symmetry breaking to elucidate superconductivity in condensed matter [21]. A spontaneous broken symmetry occurs when a symmetric state ends up in an asymmetric state, while the overall symmetry is preserved in the system [22]. In 1962, Anderson demonstrated that superconductivity is actually a state where electromagnetic gauge invariance is spontaneously broken [23]. Nambu and Jona-Lasinio applied this idea to particle physics to explain the chiral symmetry breaking in QCD [24, 25]. A spontaneous symmetry breaking in the QCD sector leads to the mass generation of hadrons far exceeding the masses of the quarks. Goldstone generalized this notion of spontaneous symmetry breaking in 1961 [26] through what is known as the Goldstone's theorem: 'the spontaneous breaking of a continuous global symmetry is always accompanied by the appearance of one or more massless scalar (spin 0) particles, the Goldstone bosons' [15]. For his proof, Goldstone added a scalar field which would have a spontaneous broken symmetry to the theory. However, this does not lead to the mass generation of the W and Z bosons. Work from Higgs [27], Brout, Englert [28], Hagen, Guralnik, and Kibble [29] finally solved this problem in 1964 by introducing the Brout–Englert-Higgs–Hagen-Guralnik–Kibble mechanism, or BEH mechanism for short.

Let us consider a complex scalar field $\phi = \frac{1}{\sqrt{2}}(\phi_1 + i\phi_2)$ with the following potential:

$$V(\phi) = \mu^2 \phi^* \phi + \lambda(\phi^* \phi)^2. \tag{1.4}$$

The shape of the potential is illustrated in Figure 1.1. On the left side, the potential symmetry is conserved such that the minimum is located at $\phi_1 = \phi_2 = 0$ (case for $\mu^2 > 0$ and $\lambda > 0$). This is the general case before any symmetry breaking. On the right side, the minima are located at $\phi_1^2 + \phi_2^2 = -\frac{\mu^2}{\lambda} = v^2$ (case for $\mu^2 < 0$ and $\lambda > 0$), with v the vacuum expected value (vev).

Since all the minima are equivalent, we can choose $(\phi_1, \phi_2) = (v, 0)$ as the ground state, so that the potential around the minimum is:

$$\begin{cases} \phi_1(x) & = v + \eta(x). \\ \phi_2(x) & = \xi(x) \end{cases} \tag{1.5}$$

The choice of the minimum breaks the continuous symmetry of the potential: this is the Goldstone theorem. The BEH mechanism can be viewed as an extension of the Goldstone theorem, where Higgs & colleagues assumed the existence of an additional gauge group with a corresponding gauge invariance in the system. By choosing the gauge to be the unitary gauge [30], the degree of freedom $\xi(x)$ characterizing the Goldstone boson disappears, and the gauge boson gains mass. $\eta(x)$ can be then interpreted as the Higgs field $H(x)$ characterized by a new particle, the Higgs boson. It has no electric charge and no spin, but is massive. It is important to note that setting the gauge does not modify the physical properties of the system, but rather the different terms in the Lagrangian now correspond to physical observables.

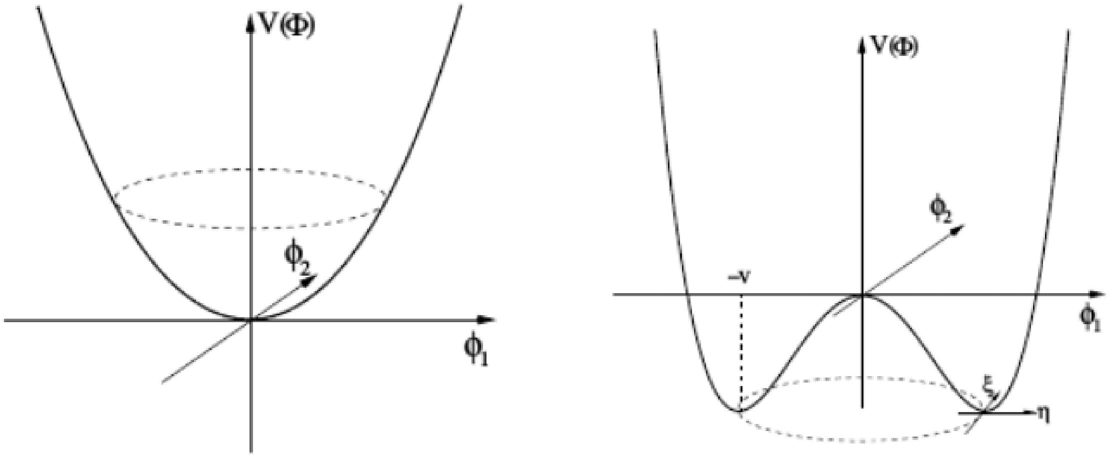


Figure 1.1: Potential of the scalar field ϕ before and after the continuous symmetry breaking [31].

Application of the Brout-Englert-Higgs mechanism: the electroweak theory

Glashow was the first to propose the EW theory as a combination of the $SU(2)_\mathrm{L}$ group in weak interaction and the $U(1)_\mathrm{em}$ group in QED [32]. The $SU(2)_\mathrm{L} \otimes U(1)_\mathrm{em}$ gauge group preserves weak interaction with $SU(2)_\mathrm{L}$ and electromagnetism with $U(1)_\mathrm{em}$, while predicting neutral-current interaction via a new gauge boson Z. This theory was however inefficient to explain the mass of the W and Z gauge bosons, until Salam [33] and Weinberg [34] applied the BEH mechanism to the $SU(2)_\mathrm{L} \otimes \mathrm{U}(1)$ group in 1967. The idea was to replace the $U(1)_\mathrm{em}$ group in QED with a new $U(1)_\mathrm{Y}$ group. This constitutes the Glashow-Salam-Weinberg (GSW) model. Similar to QED, this model introduces a new conserved quantum number Y called the weak hypercharge, and the corresponding gauge field B_μ. Weak hypercharge is related to the electric charge and weak isospin through the equation $Y = 2(Q_{em} - I_W^3)$. The gauge fields related to the photon and Z boson are linear combinations of those related to $SU(2)_\mathrm{L}$ and $U(1)_\mathrm{Y}$ as follows:

$$\begin{cases} A_\mu &= B_\mu \cos\theta_W + W_\mu^3 \sin\theta_W, \\ Z_\mu &= -B_\mu \sin\theta_W + W_\mu^3 \cos\theta_W \end{cases} \tag{1.6}$$

where W_μ^3 is the longitudinal component of the gauge field of the $SU(2)_\mathrm{L}$ group, θ_W the Weinberg angle, and Z_μ the gauge field associated to the Z boson.

The $SU(2)_\mathrm{L} \otimes U(1)_\mathrm{Y}$ gauge group provides a symmetry which can be spontaneously broken thanks to the BEH mechanism. We need three Goldstone bosons for the W and Z bosons to gain mass. The simplest model is a complex field doublet as follows:

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix}. \tag{1.7}$$

After the broken symmetry which occurs at the EW scale $\Lambda_{EW} \sim 10^2$ GeV [35], two Goldstone bosons must have an electric charge while one Goldstone boson must be neutral. In that case, the field can be rewritten in the unitary gauge such as:

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix}. \tag{1.8}$$

Similar to the previous example, the Goldstone bosons disappear, and the GSW model allows the W and Z bosons to gain mass while the photon remains massless (see section 1.1.3).

1.1.2 The fermions, elementary particles of matter

Fermions are particles whose spin quantum number has a half-integer value. They obey Fermi-Dirac statistics, forbidding two fermions to have exactly the same quantum numbers [36, 37].

Dirac spinors, helicity and chirality

All elementary fermions are mathematically described as Dirac spinors u and v, for particles and antiparticles respectively. They represent the solutions to the Dirac equation for spin-$\frac{1}{2}$ massive particles [38]. The helicity of a particle is the projection of its spin onto the direction of momentum. It is positive or 'right-handed' if the direction of its spin is the same as the direction of its motion, and negative or 'left-handed' if the directions of spin and motion are opposite. We usually use the ↑ and ↓ notation for right-handed and left-handed particles respectively.

Chirality is a more abstract concept. Chirality eigenstates are the eigenstates of the operator $\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3$, where γ^i are the gamma matrices. Chirality eigenstates are either right (R) or left (L). We define the projectors P_R and P_L as follows:

$$\begin{cases} P_R &= \frac{1}{2}(1+\gamma^5) \\ P_L &= \frac{1}{2}(1-\gamma^5) \end{cases} \Rightarrow \begin{cases} P_R u_R = u_R, P_R u_L = 0, P_R v_R = 0, P_R v_L = u_L. \\ P_L u_R = 0, P_L u_L = u_L, P_L v_R = v_R, P_L v_L = 0 \end{cases} \tag{1.9}$$

It is worth noting that in the ultrarelativistic (UR) limit, chirality and helicity eigenstates are equivalent, meaning $u_R = u_\uparrow, u_L = u_\downarrow, v_R = v_\uparrow$, and $v_L = v_\downarrow$. This always holds true for massless particles.

P-violation in weak sector

In 1957, Wu demonstrated that weak interaction violates parity (P) symmetry in β decay of cobalt-60, unlike QED and QCD [39]. Marshak, Sudarshan, followed by Feynman and Gell-Mann [40] proposed a theory in weak interaction where vectorial and axial vectorial terms combine, and break P-symmetry. This is known as the V-A theory. In fact, maximal P-violation occurs for either V+A or V-A terms, but experiments confirm the V-A coupling. It is worth noting that Bludman [41] was the first one to add an intermediate particle, now known as the W boson, in a 4-fermions interaction to explain weak interaction. Marshak and colleagues used a direct 4-fermions vertex, known as the Fermi theory.

P-violation holds significant consequences for fermions interacting via W and Z bosons. Left-handed particles and right-handed antiparticles only contribute in weak interaction. In the UR limit, left particles and right antiparticles only contribute in weak interaction. We will consider the UR limit for the rest of the chapter. Note that weak-neutral currents do not represent a pure but a partial V-A coupling.

The leptons

Leptons are divided into three generations, or flavors, with one particle and its corresponding neutrino. The first three particles are the electron, the muon, and the tau. They have a spin equal to $\frac{1}{2}$ and an electric charge of -1. They are massive, with the tau mass approximately 3,500 times greater than the electron mass. They interact with both electromagnetic and weak forces. The corresponding neutrinos are the electron, muon, and tau neutrinos. They have a spin equal to $\frac{1}{2}$ and no electric charge. They also have no mass in the SM, even though experiments proved that neutrinos can change flavor after propagating through space, and thus gain mass, a phenomenon known as neutrino oscillations [42–45]. They only interact with weak force. Due to their low or null masses (511 keV for the electron in natural units [8]), the electron and the neutrinos are the only stable leptons.

Weak interaction couples left leptons with right antiparticles. Weak-charge currents change the electric charge of the leptons by one unit of electric charge, while weak-neutral currents preserve it. Left particles and right antiparticles form $SU(2)_\mathrm{L}$ doublets as follows:

$$\begin{pmatrix} \nu_e \\ e^- \end{pmatrix}_L \begin{pmatrix} \nu_\mu \\ \mu^- \end{pmatrix}_L \begin{pmatrix} \nu_\tau \\ e^\tau \end{pmatrix}_L \begin{pmatrix} e^+ \\ \overline{\nu_e} \end{pmatrix}_L \begin{pmatrix} \mu^+ \\ \overline{\nu_\mu} \end{pmatrix}_L \begin{pmatrix} \tau^+ \\ \overline{\nu_\tau} \end{pmatrix}_L, \quad \text{with } I_W^3 = \begin{cases} +\frac{1}{2} \\ -\frac{1}{2} \end{cases} . \tag{1.10}$$

Right particles and left antiparticles form $SU(2)_\mathrm{L}$ singlets as follows:

$$e_R^- \; \mu_R^- \; \tau_R^- \; e_L^+ \; \mu_L^+ \; \tau_L^+, \quad \text{with } I_W^3 = 0. \tag{1.11}$$

It is noteworthy that right neutrinos and left antineutrinos are not included in the SM.

The quarks and hadrons

Similar to leptons, quarks are divided into three generations of two quarks. The first half are the up, charm, and top quarks ('up-quarks'), with an electric charge of $\frac{2}{3}$, and the second half are the down, strange, and bottom quarks ('down-quarks'), with an electric charge of $-\frac{1}{3}$. They have a spin equal to $\frac{1}{2}$. They are massive, with the top quark mass approximately 90,000 times greater than the up quark mass. Due to their low masses (2.16 MeV for the up quark and 4.67

MeV for the down quark [8]), up and down quarks are the only stable quarks. They interact with electromagnetic, weak, and strong forces. Left quarks and right antiquarks form $SU(2)_\mathrm{L}$ doublets as follows:

$$\begin{pmatrix} u \\ d \end{pmatrix}_L \begin{pmatrix} c \\ s \end{pmatrix}_L \begin{pmatrix} t \\ b \end{pmatrix}_L \begin{pmatrix} \overline{d} \\ \overline{u} \end{pmatrix}_L \begin{pmatrix} \overline{s} \\ \overline{c} \end{pmatrix}_L \begin{pmatrix} \overline{b} \\ \overline{t} \end{pmatrix}_L, \quad \text{with } I_W^3 = \begin{cases} +\frac{1}{2} \\ -\frac{1}{2} \end{cases} . \tag{1.12}$$

Right quarks and left antiquarks form $SU(2)_\mathrm{L}$ singlets as follows:

$$u_R\; d_R\; c_R\; s_R\; t_R\; b_R\; \overline{u}_L\; \overline{d}_L\; \overline{c}_L\; \overline{s}_L\; \overline{t}_L\; \overline{b}_L, \quad \text{with } I_W^3 = 0. \tag{1.13}$$

Quarks are the only fermions which respond to QCD. Each quark carries one color charge, such that the gluon carries the color and anticolor charges of both the quark and antiquark in a quark-antiquark-gluon vertex. However, quarks and gluons cannot be observed individually under normal conditions. This phenomenon is known as the color confinement: observable states must form $SU(3)_\mathrm{c}$ singlets or 'colorless', 'white' states. Quarks and gluons must clump together to form composite particles of two or more quarks, bond together by gluons, and named hadrons [46]. Although analytically unproven, lattice QCD approach, where spacetime is discretized into a lattice, confirms the color confinement model [47]. Quarks can indeed interact with one another due to their color charges, and form a flux tube between them. As quarks move apart, the tube elongates until it ruptures, and creates a quark-antiquark pair. This process repeats until the quarks' energy is sufficiently low to create colorless hadrons. This is the hadronization process [48] illustrated in Figure 1.2. It must be pointed out that the hadronization process generates so many hadrons that particles are experimentally clustered in a cone of hadrons, known as a 'jet'.

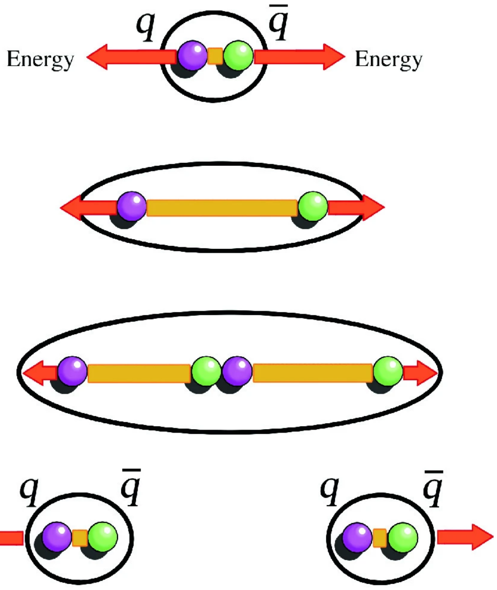


Figure 1.2: Hadronization process. The energy transferred to the two quarks q and $\overline{q}$ creates instead a new quark-antiquark pair [49].

Hadrons are categorized into baryons with an odd number of quarks (usually three), and mesons with an even number of quarks (usually two with one quark and one antiquark), named the valence quarks. Baryons, like protons and neutrons, are fermions while mesons, like pions, are actually bosons. Protons stand out as the only free stable hadron, and highly unstable hadrons, like tetraquarks and pentaquarks, have been observed only recently [50–52]. In 1969, Feynman proposed a new model to characterize Deep Inelastic Scattering (DIS) between hadrons at high energies [53]. Hadrons are composed of point-like particles named partons, later identified as quarks and gluons, which interact with an incident particle while carrying a fraction x of the hadron's momentum. In addition to the valence quarks, these partons are made of a mix of quarks, antiquarks, and gluons, known as the quark sea, which give the hadrons their properties without changing their quantum numbers [54]. The distributions of the quarks' momentum x within a hadron, regarding its interaction with an incident particle, are referred to as the Parton Density Functions (PDFs). These functions are crucial for calculating cross sections in proton-proton colliders like the LHC (see section 2.2.2).

CP-violation in weak sector for quarks

In 1964, Cronin and Fitch demonstrated that weak interaction violates both Charge and Parity (CP) symmetries in β decay of neutral kaons K^0 ($d\bar{s}$) and $\overline{K^0}$ ($\bar{d}s$) [55]. Neutral kaons can decay into either two or three pions. According to CP-symmetry, kaons must only decay into three pions but experimental results show the opposite results. In 1973, Kobayashi and Maskawa predicted a third generation of quarks previously unknown to account for this violation [56]. The idea was built on Cabibbo's work in 1963 to preserve the weak interaction universality [57]. He demonstrated that weak interaction eigenstates differ from mass, or Hamiltonian eigenstates, but are connected by a 2×2 rotation matrix. It is to note that Glashow, Iliopoulos, and Maiani also predicted a fourth quark in 1970 then unknown to confirm Cabibbo's interpretation, through the Glashow-Iliopoulos-Maiani (GIM) mechanism [58]. The Cabibbo-Kobayashi-Maskawa (CKM) matrix extends this concept to three generations of quarks. This matrix is a unitary, almost diagonal 3×3 matrix represented by three real parameters, and one complex phase violating the CP-symmetry. These parameters are not predicted by the theory but are derived from experimental observations [8].

Yukawa mechanism and fermion mass

Similar to gauge bosons, fermion mass terms, added by hand in the Lagrangian, are not gauge invariant, and the Lagrangian is not renormalizable. One approach to resolve this issue is to introduce new terms coupling fermions with a scalar field. These are known as Yukawa terms, where the scalar component corresponds to the Higgs field in the SM [34]. Following spontaneous symmetry breaking in the EW sector, both left and right fermions interact with the Higgs field to acquire mass, as for gauge bosons. However, quarks can change flavor through the CKM matrix as it will be explained in the section 1.1.3. It is worth mentioning that Yukawa interaction is consistent with experimental results but is still an input to the theory. A more comprehensive theory is needed to understand the origin of the Yukawa terms [59, 60].

1.1.3 The Lagrangian of the Standard Model

The Lagrangian of the SM can be divided into four sectors [61, 62] as follows:

$$L_{SM} = L_{YM} + L_D + L_{Higgs} + L_{Yukawa}. \tag{1.14}$$

These are the Yang-Mills, Dirac, Higgs, and Yukawa sectors respectively.

Yang-Mills sector

The Yang-Mills part of the Lagrangian corresponds to the kinematic terms of the gauge fields as follows:

$$\begin{aligned} L_{YM} &= -\frac{1}{4}B_{\mu\nu}B^{\mu\nu} - \frac{1}{4}W^i_{\mu\nu}W_i^{\mu\nu} - \frac{1}{4}G^A_{\mu\nu}G_A^{\mu\nu}, \text{ with} \\ B_{\mu\nu} &= \partial_\mu B_\nu - \partial_\nu B_\mu \\ W^i_{\mu\nu} &= \partial_\mu W^i_\nu - \partial_\nu W^i_\mu + g\epsilon^{ijk}W_{\mu j}W_{\nu k} \\ G^A_{\mu\nu} &= \partial_\mu A^A_\nu - \partial_\nu A^A_\mu + g" f^{ABC}A_{\mu B}A_{\nu C}. \end{aligned} \tag{1.15}$$

The symbols g and $g"$ are the coupling parameters of the weak interaction and QCD respectively. The tensor $B_{\mu\nu}$ is the hypercharge tensor connected to the $U(1)_\mathrm{Y}$ gauge field. The tensor $W^i_{\mu\nu}$ is the weak interaction tensor connected to the $SU(2)_\mathrm{L}$ gauge fields. The tensor $G^A_{\mu\nu}$ is the gluon field strength tensor connected to the $SU(3)_\mathrm{c}$ gauge fields. The tensors $W^i_{\mu\nu}$ and $G^A_{\mu\nu}$ transform with an extra term to ensure the invariance gauge under $SU(2)_\mathrm{L}$, $U(1)_\mathrm{Y}$, and $SU(3)_\mathrm{c}$ transformations such as:

$$\begin{aligned} SU(2)_\mathrm{L} : W^i_\mu &\to W'^i_\mu = e^{-i\alpha^i(x)T_i}W^i_\mu \\ U(1)_\mathrm{Y} : B_\mu &\to B'_\mu = e^{-\frac{i}{2}\beta(x)}B_\mu \\ SU(3)_\mathrm{c} : A^A_\mu &\to A'^A_\mu = e^{-i\delta^A(x)t_A}A^A_\mu. \end{aligned} \tag{1.16}$$

Dirac sector

The Dirac part of the Lagrangian corresponds to the kinematic terms of the fermions, and their interaction with the gauge bosons as follows:

$$\begin{aligned} L_D = i\overline{L}_{i,L}\gamma^\mu D_\mu L^i_L + i\overline{L}_{i,R}\gamma^\mu D_\mu L^i_R + i\overline{Q}_{i,L}\gamma^\mu D_\mu Q^i_L + i\overline{u}_{i,R}\gamma^\mu D_\mu u^i_R + i\overline{d}_{i,R}\gamma^\mu D_\mu d^i_R, \text{ with} \\ D_\mu L^i_L = \left(\partial_\mu - igT_iW^i_\mu - ig'\frac{Y}{2}B_\mu\right)L^i_L \\ D_\mu L^i_R = \left(\partial_\mu - ig'\frac{Y}{2}B_\mu\right)L^i_R \\ D_\mu Q^i_L = \left(\partial_\mu - igT_iW^i_\mu - ig'\frac{Y}{2}B_\mu \ - ig"t_AA^A_\mu\right)Q^i_L \\ D_\mu u^i_R = \left(\partial_\mu - ig'\frac{Y}{2}B_\mu \ - ig"t_AA^A_\mu\right)u^i_R \\ D_\mu d^i_R = \left(\partial_\mu - ig'\frac{Y}{2}B_\mu \ - ig"t_AA^A_\mu\right)d^i_R. \end{aligned} \tag{1.17}$$

The symbol g' is the coupling parameter of the hypercharge force, and is related to g as it will be demonstrated in the next section. The multiplets L_L, Q_L are respectively the $SU(2)_\mathrm{L}$ doublets for both left leptons and quarks. The multiplets L_R, u_R, and d_R are respectively the $SU(2)_\mathrm{L}$ singlets for right leptons, up-, and down-quarks. The corresponding weak hypercharges are $Y = -1, \frac{1}{3}, -2, \frac{4}{3}$, and $-\frac{2}{3}$. The covariant derivatives ensure the invariance gauge under $SU(2)_\mathrm{L}$, $U(1)_\mathrm{Y}$, and $SU(3)_\mathrm{c}$ transformations such as:

$$\begin{aligned} SU(2)_\mathrm{L} : L'_L = e^{-i\alpha^i(x)T_i}L_L,\ L'_R = L_R \quad \text{(identical for the quarks)} \\ U(1)_\mathrm{Y} : L'_L = e^{-\frac{i}{2}\beta(x)}L_L,\ L'_R = e^{\frac{i}{2}\beta(x)}L_R \quad \text{(identical for the quarks)} \\ SU(3)_\mathrm{c} : Q'_L = e^{-i\delta^A(x)t_A}Q_L,\ u'_R = e^{i\delta^A(x)t_A}u_R,\ d'_R = e^{i\delta^A(x)t_A}d_R. \end{aligned} \tag{1.18}$$

Higgs sector

The Higgs part of the Lagrangian corresponds to the terms of the Higgs field, and its interaction with the gauge bosons as follows:

$$\begin{gathered}
L_{Higgs} = (D_\mu\phi)^\dagger(D^\mu\phi) - V(\phi), \text{ with} \\
D_\mu\phi = \left(\partial_\mu - igT_iW^i_\mu - ig'\frac{Y}{2}B_\mu\right)\phi \\
V(\phi) = -\mu^2\phi^\dagger\phi + \lambda(\phi^\dagger\phi)^2.
\end{gathered} \tag{1.19}$$

The weak hypercharge is $Y = 1$. After the spontaneous symmetry breaking in the EW sector, the scalar field can be written such as:

$$\phi = e^{\frac{i}{v}T_i\xi^i}\begin{pmatrix} 0 \\ (v+H)/\sqrt{2} \end{pmatrix}. \tag{1.20}$$

The unitary $SU(2)_\mathrm{L}$ transformation $U(\xi) = e^{-\frac{i}{v}T_i\xi^i}$ modifies the scalar field such as:

$$\phi \to \phi' = U(\xi)\phi = \frac{1}{\sqrt{2}}(v+H)\chi, \text{ with } \chi = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{1.21}$$

The other fields must undergo the same transformation in the whole Lagrangian such as:

$$\begin{gathered}
T_iW'^{i}_\mu = U(\xi)T_iW^i_\mu U(\xi)^{-1} - \frac{i}{g}\left(\partial_\mu U(\xi)\right)U^\dagger(\xi), \;\; L'^i_L = U(\xi)L^i_L, \;\; Q'^i_L = U(\xi)Q^i_L \\
A'^{A}_\mu = A^A_\mu, \;\; B'_\mu = B_\mu, \;\; L'^i_R = L^i_R, \;\; u'^i_R = u^i_R, \;\; d'^i_R = d^i_R.
\end{gathered} \tag{1.22}$$

The Higgs part of the Lagrangian is now the following:

$$\begin{aligned}
L_{Higgs} =& \frac{1}{2}\chi^\dagger\left(\partial_\mu + igT_iW'^i_\mu + i\frac{g'}{2}B_\mu\right)(v+H)\left(\partial^\mu - igT_jW'^{j\mu} - i\frac{g'}{2}B^\mu\right)(v+H)\chi \\
&+ \frac{\mu^2}{2}(v+H)^2\chi^\dagger\chi - \frac{\lambda}{4}(v+H)^4\left(\chi^\dagger\chi\right)^2 \\
=& \frac{1}{2}\partial_\mu H\partial^\mu H + \frac{1}{8}\left(g^2W'^i_\mu W'^\mu_i + g'^2B_\mu B^\mu - 2gg'B_\mu W'^{3\mu}\right)(v+H)^2 \\
&+ \frac{\mu^2v^2}{4} + \frac{1}{2}(2\mu^2)H^2 + \lambda vH^3 + \frac{\lambda}{4}H^4 \\
=& \frac{1}{2}\partial_\mu H\partial^\mu H + \frac{1}{8}\left(g^2W'^1_\mu W'^{1\mu} + g^2W'^2_\mu W'^{2\mu} + (gW'^3_\mu - g'B'_\mu)^2\right)(v+H)^2 \\
&+ \frac{\mu^2v^2}{4} + \frac{1}{2}(2\mu^2)H^2 + \lambda vH^3 + \frac{\lambda}{4}H^4.
\end{aligned} \tag{1.23}$$

We define the W^+ and W^- gauge fields using the following change of variables:

$$W^\pm_\mu = \frac{W'^{1}_\mu \mp iW'^{2}_\mu}{\sqrt{2}}. \tag{1.24}$$

The mass terms in the Lagrangian are quadratic to the gauge fields, such that the mass term of the W$^+$ and W$^-$ gauge bosons is $\frac{1}{4}g^2v^2W^+_\mu W^{-\mu}$. The W gauge bosons then acquire a mass $m_W = \frac{1}{2}gv$ through the BEH mechanism. The mass terms of the W'^{3}_μ and B'_μ gauge fields are mixed as follows:

$$\frac{v^2}{8}\begin{pmatrix} W'^{3}_\mu & B'_\mu \end{pmatrix}\begin{pmatrix} g^2 & -gg' \\ -gg' & g^2 \end{pmatrix}\begin{pmatrix} W'^{3\mu} \\ B'^\mu \end{pmatrix}. \tag{1.25}$$

Diagonalizing the matrix leads to:

$$\frac{v^2}{8}\left(Z_\mu \; A_\mu\right)\begin{pmatrix} g^2+g'^2 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} Z^\mu \\ A^\mu \end{pmatrix} = \frac{v^2}{8}(g^2+g'^2)Z_\mu Z^\mu + 0 \times A_\mu A^\mu, \tag{1.26}$$

where Z_μ and A_μ are defined thanks to the GSW model. The Z boson acquires a mass $m_Z = \frac{1}{2}v\sqrt{g^2+g'^2}$, while the photon is massless. The Weinberg angle is directly linked to the parameters g and g' such as:

$$\tan\theta_W = \frac{g'}{g},\;\; \sin\theta_W = \frac{g'}{\sqrt{g^2+g'^2}},\;\; \cos\theta_W = \frac{g}{\sqrt{g^2+g'^2}},\;\; m_Z = \frac{m_W}{\cos\theta_W}. \tag{1.27}$$

The mass term of the Higgs boson is $\frac{1}{2}(2\mu^2)H^2$ such that the Higgs boson acquires a mass $m_H = \sqrt{2\mu^2}$. The remaining terms of the Lagrangian correspond to the interactions between the Higgs boson and the gauge bosons, and among the bosons themselves.

Yukawa sector

The Yukawa part of the Lagrangian corresponds to the terms of interaction between the fermions and the Higgs field as follows:

$$L_{Yukawa} = \overline{L}_{i,L}Y_l^{ij}\phi L_{j,R} + \overline{Q}_{i,L}Y_u^{ij}\phi u_{j,R} + \overline{Q}_{i,L}Y_d^{ij}\phi d_{j,R} + h.c., \tag{1.28}$$

where Y^e, Y^u, and Y^d are complex 3×3 matrices describing the Yukawa couplings. After the unitary transformation and the redefinition of the matrices, the Yukawa part of the Lagrangian is the following:

$$L_{Yukawa} = \frac{v+H}{\sqrt{2}}\left(\overline{L'}_{i,L}g_l^{ij}\chi L_{j,R} + \overline{Q'}_{i,L}g_u^{ij}\chi V_{jk}u_R^k + \overline{Q'}_{i,L}g_d^{ij}\chi d_{j,R}\right) + h.c., \tag{1.29}$$

where g_l, g_u, and g_d are complex diagonal 3×3 matrices describing the Yukawa couplings per fermion g_f, and V is the CKM matrix. The mass terms of the leptons are $\frac{1}{\sqrt{2}}vg_f(\overline{L'}_L L_R)$ and $\frac{1}{\sqrt{2}}vg_f(\overline{Q'}_L Q_R)$, such that the fermions acquire a mass $m_f = \frac{1}{\sqrt{2}}vg_f$. The Dirac spinors can be redefined to remove the CKM matrix from the mass terms. However, the CKM matrix cannot be removed from the remaining terms of the Lagrangian which correspond to the interaction between the Higgs boson and the fermions. This fact leads to the existence of a complex phase, explicitly violating the CP-symmetry in the quark sector.

Experimental validation of the Standard Model

The SM is widely regarded as one of the most successful physical theories to date. First, it predicted the existence of numerous particles, later confirmed experimentally.

- The Z and W bosons predicted in 1967, hinted first at the Gargamelle experiment at CERN in 1971 [63], and confirmed at the Super Proton Synchroton (SPS) at CERN in 1983 [64].
- The charm quark predicted in 1964, and discovered following the J/ψ ($\overline{c}c$) meson detection at the Stanford Linear Accelerator Center (SLAC) and the Brookhaven National Laboratory (BNL) in 1974 [65].

- The tau lepton predicted in 1960 by Zichichi [66], and discovered by Perl in 1975 [67].
- The bottom quark, or 'b'-quark, predicted in 1973, and discovered at the E288 experiment at Fermilab in 1977 [68].
- The top quark predicted in 1973, and discovered at the Collider Detector (CDF) [69] and D0 [70] experiments at Fermilab in 1994.
- The Higgs boson predicted in 1964, and discovered at the CMS [71] and ATLAS [72] experiments at CERN in 2012.

Secondly, the various properties of the particles have been measured with an associated uncertainty of under 10^{-6} per cent [8], making the SM one of the most precise theory in physics. Finally, the complete Lagrangian of the SM is renormalizable, as 't Hooft and Veltman demonstrated in 1971 [73].

However, the Lagrangian of the SM involves 18 free parameters, all of them determined experimentally. These include 9 parameters g_f for the leptons masses, 2 parameters v and m_H for the bosons masses, 3 parameters g, g', and g" for the weak, hypercharge, and strong interactions respectively, and 4 parameters for the mixing angles and the complex phase in the CKM matrix. This unpredictability poses challenges in considering the SM as a robust theory, and motivates the exploration into possible extensions to push the boundaries of the current SM.

1.2 Beyond the Standard Model

We have already discussed several limitations of the SM, such as the unknown origin of the color confinement and the mass of the neutrinos, but other problems remain. How can we unify gravity with quantum mechanics? What causes the significant matter-antimatter asymmetry in the universe? And why do particles have such disparate masses? The latter point is particularly problematic within the Higgs sector, as the SM cannot predict the mass of the Higgs boson. This broader problem is known as the hierarchy problem.

1.2.1 The hierarchy problem

Origin of the hierarchy problem

The hierarchy problem arises when a theory needs to consider several energy scales simultaneously. If this happens, certain terms in the Lagrangian may become negligible, and have minimal impact. There are various levels to this problem, and we can distinguish two cases [74].

- Gravity is much weaker than the other fundamental forces (by a factor of 10^{-33} compared to the weak force [75]). Yet it becomes non-negligible at approximately 1 cm, so why? Suppose there exists a graviton. If the terms involving the graviton outweighs the ones of the Higgs boson in the Lagrangian, the latter becomes completely negligible. The characteristic energy scale shifts then to the Planck scale $\Lambda_{Pl} = G_N^{-1/2} \sim 10^{18}$ GeV, and the EW scale $\Lambda_{EW} \sim 10^2$ GeV is negligible [76]. The SM does not have this problem currently, but this could pose significant challenges if we aim to unify the SM with gravity. The hierarchy problem can be considered as a 'tree' level problem here, because it compares different forces across distinct energy scales.

- Let us assume that the EW scale is not the characteristic energy scale of the SM anymore, but has been replaced by a higher energy scale like the Planck scale. As shown in equation 1.3, electromagnetic, weak, and strong forces require renormalization through β-functions to take loop corrections into account. At this new energy scale, higher-order corrections may be not renormalizable, which would make the theory divergent. The hierarchy problem can be considered as a 'loop' level problem here, because the behavior of the same force varies depending on the energy scale.

The hierarchy problem in the Higgs sector

The Higgs boson hierarchy problem is a fundamental loop problem. Let consider an energy scale $\Lambda \gg \Lambda_{EW} \sim \sqrt{2\mu^2} = m_H$. Without taking any loop corrections into account, the Higgs boson mass is denoted as the bare mass. The major one-loop corrections within the SM for calculating the Higgs boson mass are summarized in Figure 1.3. The Higgs boson interacts either with itself, a virtual Z* boson, or top quarks through HHH, HZ^*H, and $Ht\bar{t}$ vertices. The different contributions to these Feynman diagrams result in:

$$\delta\mu^2 = \frac{1}{16\pi^2}\left(12\lambda + \frac{3}{4}(3g^2 + g'^2) - 2g_t^2\right)\Lambda^2. \tag{1.30}$$

The loop corrections cause the Higgs boson mass to diverge at least quadratically with energy and not logarithmically, which would make the theory renormalizable [77]. Applying a cutoff like Λ_{UV} at the UV scale could mitigate the divergence, but this approach does not predict the behavior at the Planck scale. To solve the Higgs boson hierarchy problem, the Higgs boson mass must yield unambiguously finite results for arbitrarily high energies like the Planck scale. The theory will be then characterized as a UV-complete theory, without any cutoff needed. Moreover, the Higgs boson mass must be consistent with the bare mass plus the different corrections at both the EW scale and the TeV scale, such as at the LHC, simultaneously. One solution is to empirically modify the expression of the β-function to match the predicted mass value, corrections included, with the experimental results at both EW and LHC levels. This method, known as fine-tuning, is very effective but lacks theoretical foundation. Furthermore, it is conceptually problematic or not 'natural', as 't Hooft would describe it in his naturalness theory [78]. Consequently, new BSM models have been proposed to fix the hierarchy problem in the Higgs sector.

Supersymmetry

The initial intuition to remove the quadratic divergence in the Higgs sector involves introducing new BSM fields which interact with the Higgs boson, in a way which cancels out the SM and BSM loop contributions. In other words, '$\delta\mu_{BSM} = -\delta\mu$'. It must be the consequence of a symmetry to ensure cancellation to all orders in perturbations. According to the spin-statistics theorem [79, 80], this cancellation can be achieved if fermions were turned into bosons in Figure 1.3 and vice versa. This symmetry between fermions and bosons, known as supersymmetry, has been the leading theory as a solution to the hierarchy problem for decades [81]. It assumes that for each fermion (boson) in the SM, there exists a corresponding boson (fermion) with opposite spin properties [82]. By considering the last contribution in Figure 1.3 only, the additional correction is:

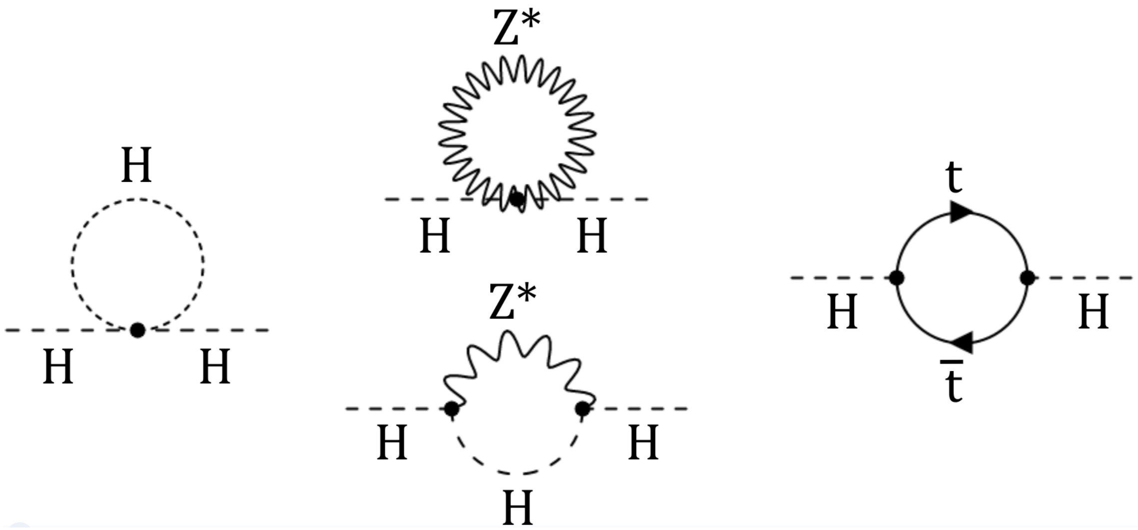


Figure 1.3: Major one-loop corrections to the Higgs boson mass. The Higgs boson interacts either with itself, a virtual Z* boson or top quarks.

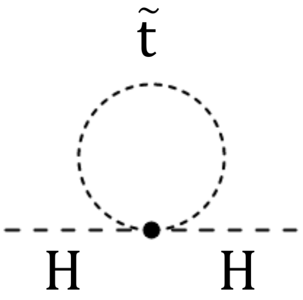


Figure 1.4: One-loop correction to the Higgs boson mass interacting with the top quark superpartner.

$$\delta\tilde{\mu}^2 = \frac{\tilde{g_t}^2}{8\pi^2}\left(\Lambda^2 - \tilde{m_t}^2 \log\left\{\frac{\Lambda^2}{\tilde{m_t}^2}\right\}\right), \qquad (1.31)$$

where $\tilde{g_t}$ and $\tilde{m_t}$ are the coupling parameter and the mass of the top quark partner respectively. By adding the two contributions, the overall divergence becomes logarithmic if $g_t = \tilde{g_t}$, and the theory is renormalizable.

In the simplest formulation of supersymmetry, known as the Minimal Supersymmetric Standard Model (MSSM), the 'superpartner' shares the same quantum numbers, besides spin, as its partner [83, 84]. A spontaneous broken symmetry also occurs in the supersymmetric part of the theory to give superpartners their mass [85, 86]. More intricate theories, known as Next-to Minimal Supersymmetric Standard Model (NMSSM), adds a singlet scalar supersymmetric field to the MSSM, but we will not go into detail. However, there is still no experimental evidence of supersymmetry. Current findings from the LHC push the exclusion of supersymmetric particles up to $\Lambda \sim 1$ TeV [87].

Composite Higgs models

Composite Higgs models assume that the Higgs doublet is a composite field, whose constituents are bound by a strong interaction similar to QCD. This new strong sector must exhibit a mechanism leading to a large separation of the energy scales to explain why the Higgs boson, but no other new bound states, have been found yet [88]. Most of these theories search for new particles at $\Lambda > 1$ TeV (since no new physics have been found below this limit) [89, 90]. At this energy scale, the Higgs boson in the composite Higgs models is no longer viewed as an elementary particle, and the different corrections in Figure 1.3 become invalid. Similar to hadrons, a PDF must describe the behaviors of particles within the Higgs boson to stabilize its mass value. Additionally, a new term L_{int} is introduced in the Lagrangian which corresponds to the interaction between the SM Lagrangian and the new strong sector.

While most of the predicted particles which are excitations or ingredients of the composite Higgs have a mass around a TeV or more [91, 92], the Higgs boson is much lighter. One explanation for this is to consider the composite Higgs boson as a Goldstone boson associated with a global symmetry of the new strong sector. This extended symmetry of the SM is spontaneously broken by confinement such that the Higgs boson mass is no longer related to the energy scale, but is rather set by an external source of the global symmetry breaking. Composite Higgs models are barely searched for at the LHC [93, 94], because the center-of-mass energy might not be high enough to reach the scale where the composite models live.

Little Higgs models

Similar to composite Higgs models, little Higgs models assume the existence of an extended symmetry which includes the EW symmetry [95]. At the TeV energy scale, the extended symmetry is explicitly broken while preserving the EW symmetry. The Higgs boson transforms into a pseudo-Goldstone Boson (pGB) which is massless at tree level, and only acquires mass through loop corrections. It is important to note that the one-loop corrections, like in Figure 1.3, are naturally suppressed at 1-loop level for pGBs [96], thereby preserving the Higgs boson mass up to 10 TeV. This mechanism is a limited UV-complete solution to the hierarchy problem. Some little Higgs models predict the existence of new particles with a mass around 10 TeV or more [97, 98], which have obviously not been observed by the LHC for now.

Extra dimension models

Extra dimension models, as the name implies, propose the addition of n spatial dimensions to our four common dimensions. The characteristic scale of the new $4 + n$-dimensional space is the Planck scale, which could be enormous compared to the weak scale if the size of the new dimensions is very small [76]. The hierarchy problem at tree level is then resolved: there is no hierarchy between the weak scale and the true gravitational scale, as we do not compare the same spacetime. The main idea is to determine how our four-dimensional space would fit within the extra dimensions. New dimensions could be compact or not. In the case of compact dimensions, they follow the Kaluza-Klein mechanism, and predict new particles [99, 100]. Interaction between the Higgs boson and these particles has to be added to the ones presented in Figure 1.3, resulting in an overall renormalizable theory.

We can categorize extra dimension models into three types, each predicting a different nature of the Higgs boson.

- Large extra dimension models consider that our space exists on a brane in a higher dimensional space. The SM lives only in our four-dimensional space, while gravity can propagate across the extra dimensions. In this scenario, the Higgs field itself is not modified [76, 101].
- Conversely, universal extra dimension models suppose that all fields propagate not only in the four-dimensional space, but also in the extra dimensions. The Higgs field must be then modified as a $4 + n$-dimensional scalar field [102].

- Warped extra dimension models are similar to large extra dimensions, where EW symmetry does not arise from a consequence of gauge interaction, but is rather based on the $4 + n$-dimensional metric [103, 104].

Most of these theories predict one or two extra dimensions, and anticipate particles with a mass which could be at the TeV scale. Such particles have not been observed at the LHC either [105].

When comparing various models addressing the hierarchy problem, it becomes pretty clear that most of them have been constructed to predict new physics around $\Lambda \sim 1 - 10$ TeV scale, which closely matches with the current capabilities of the LHC. Dozens of models exist to solve the hierarchy problem but the ones discussed here are among the most prominent, and predict the existence of new particles which could be identified as VLQs.

1.2.2 The Vector-Like Quark, a possible solution to the hierarchy problem

General characteristics of Vector-Like Quarks

A VLQ is a quark with left- and right-handed components transforming in a left-right symmetric way under the SM gauge group $SU(3)_\mathrm{c} \otimes SU(2)_\mathrm{L} \otimes U(1)_\mathrm{Y}$ [106]. This contrasts with the SM chiral fermions, where only left-components carry charged-currents, while both left- and right-components of VLQs carry charged-currents. They interact with SM quarks and the Higgs boson through Yukawa couplings [107]. Gauge-invariant interaction requires that the VLQs are represented as $SU(2)_\mathrm{L}$ singlets, doublets or triplets. Furthermore, bare mass terms of the VLQs $M_Q\overline{Q}Q$ are gauge invariant, and must be incorporated into the Lagrangian [108]. As their mass is not generated by couplings to scalar doublets, VLQs cannot be regarded as a fourth quark generation [109]. The VLQs masses can be significantly larger than the characteristic mass of the SM quarks at the EW scale, since they are not derived from a spontaneous symmetry breaking.

The mixing terms between both quarks and VLQs imply that the CKM matrix no longer fully describes the charged-current couplings. Additionally, the unitarity of the CKM matrix is no longer preserved, with deviations proportional to the ratio m/M_Q [106]. Another consequence of this extended mixing is that neutral-currents via a Z boson can directly modify the quark's flavor at tree level. This phenomenon is known as Flavor-Changing Neutral Currents (FCNC), and is related to the ratio m/M_Q as well. Moreover, the coupling between the Higgs field and SM quarks is modified, potentially leading to Higgs-mediated FCNC, and related to the ratio v/M_Q. These effects are similar to those described in the seesaw mechanism, which predicts the existence of three Majorana neutrinos [110]. Majorana fermions are their own antiparticle, and their mass terms are analogous to those of the VLQs in the Lagrangian [111]. In the current seesaw mechanism, the mass values of the new neutrinos are constructed in a way that the discrepancy between the two mass scales predicts FCNC in the neutrino sector at the TeV scale. Similar to the alteration of the CKM matrix, deviations from the Pontecorvo-Maki-Nakagawa-Sakata (PMNS) matrix, which characterizes the flavor changing in the neutrino sector [112, 113], are also expected.

The different types of VLQs are presented in Table 1.1 [107, 114]. The model includes two singlets, three doublets, and two triplets of the $SU(2)_\mathrm{L}$ group. We assume that the spontaneous symmetry breaking in the EW sector has already taken place, such that the resulting Lagrangian is as follows [115, 116]:

$$\begin{aligned} L = L_{SM} &+ i\overline{X}\not{D}X - M_X\overline{X}X + i\overline{T}\not{D}T - M_T\overline{T}T + i\overline{B}\not{D}B - M_B\overline{B}B + i\overline{Y}\not{D}Y - M_Y\overline{Y}Y \\ &+ \Big[H\Big(\overline{B}\chi\left(g_L^B P_L + g_R^B P_R\right) d + \overline{T}\chi\left(g_L^T P_L + g_R^T P_R\right) u\Big) \\ &+ \frac{g}{2\cos\theta_W}\Big(\overline{B}Z\left(\tilde{g}_L^B P_L + \tilde{g}_R^B P_R\right) d + \overline{T}Z\left(\tilde{g}_L^T P_L + \tilde{g}_R^T P_R\right) u\Big) \\ &+ \frac{\sqrt{2}g}{2}\Big(\overline{Y}W\left(g_L^Y P_L + g_R^Y P_R\right) d + \overline{B}W\left(g_L^B P_L + g_R^B P_R\right) u \\ &+ \overline{T}W\left(g_L^T P_L + g_R^T P_R\right) d + \overline{X}W\left(g_L^X P_L + g_R^X P_R\right) u\Big) + h.c.\Big]. \end{aligned} \tag{1.32}$$

Einstein indices are omitted for convenience. The Dirac slash notation $\not{D} = \gamma^\mu D_\mu$ is also used. The X, T, B, and Y are the VLQs $SU(2)_\mathrm{L}$ multiplets, and u and d the usual $SU(2)_\mathrm{L}$ doublets for SM up- and down-quarks. For a VLQ field Q, $Q = Q_R + Q_L$ as both left- and right-components of the VLQ are SM gauge-invariant. The first line covers the kinetic and mass contributions of the VLQs. These terms are analogous to the SM quarks contributions, with the mass terms represented in a block diagonal matrix for both SM quarks and VLQs. The rest of the Lagrangian corresponds to the interaction terms between the SM bosons (Higgs, W, and Z bosons), the SM quarks, and the VLQs. The interaction with the gluons and photon will not be considered here, as current models predict a coupling 10 times smaller between the VLQs and massless SM bosons than with massive SM bosons [117]. It is important to note that since the VLQs couple to the SM quarks, electric charge preservation requires that X and Y VLQs do not carry neutral-currents [115]. It is also interesting to note that the coupling between the Higgs boson, and the VLQs B and T may introduce a mass term M'_Q in addition to their bare mass M_Q [106]. The symbols $\tilde{g}^{VLQ}$ and g^{VLQ} represent the EW couplings between the VLQs and the SM particles. Their values differ when considering left or right SM quarks. These couplings can also be redefined to yield a Lagrangian similar to the SM Lagrangian, where all coupling information is included in an extended version of the CKM matrix V'. The general form of the new matrix is a rectangular non-unitary matrix, hence the presence of FCNC in VLQ extended models.

All in all, the inclusion of VLQs adds $6n$ parameters to the model, where n is the number of VLQs under consideration [106]. Parameters include the VLQ masses, the mixing angles, and the complex phases for the extended CKM matrix.

Theories predicting the Vector-Like Quarks

The models discussed in section 1.2.1 are possible solutions to the hierarchy problem by adding new particles to the SM, like the VLQs. The VLQs T and B exhibit direct interaction with the Higgs boson in particular, making them components of various UV-complete models aimed at addressing the hierarchy problem.

- In MSSM models, an extra doublet of VLQs $(T\ B)$ can interact and decay into the superpartners of the SM particles, and particularly the supersymmetric partner of the Higgs doublet. This extension would double the Lagrangian mentioned in equation 1.32, including the supersymmetric contributions [118, 119].

- The VLQs are a common feature in composite Higgs models. To generate the SM particles masses, the composite Higgs boson requires additional gauge interaction with VLQs. The VLQs serve as 'intermediaries' between the composite Higgs model and SM particles [120, 121].

	SM quarks	Singlets		Doublets			Triplets	
	$\begin{pmatrix} u \\ d \end{pmatrix} \begin{pmatrix} c \\ s \end{pmatrix} \begin{pmatrix} t \\ b \end{pmatrix}$	$\begin{pmatrix} T \end{pmatrix}$	$\begin{pmatrix} B \end{pmatrix}$	$\begin{pmatrix} X \\ T \end{pmatrix}$	$\begin{pmatrix} T \\ B \end{pmatrix}$	$\begin{pmatrix} B \\ Y \end{pmatrix}$	$\begin{pmatrix} X \\ T \\ B \end{pmatrix}$	$\begin{pmatrix} T \\ B \\ Y \end{pmatrix}$
$SU(2)_{\mathrm{L}}$	$Q_L = \mathbf{2}$ $u_R/d_R = \mathbf{1}$	$\mathbf{1}$		$\mathbf{2}$			$\mathbf{3}$	
Y	$Q_L = 1/3$ $u_R = 4/3$ $d_R = -2/3$	4/3	-2/3	7/3	1/3	-5/3	4/3	-2/3
Q_{em}	$u = 2/3$ $d = -1/3$	$X = 5/3$ $T = 2/3$ $B = -1/3$ $Y = -4/3$						
Number of parameters (Masses and mixing angles + Complex phases)	9 1	$4n$ $2n$						

Table 1.1: Allowed representations for SM quarks and VLQs with their corresponding quantum numbers and the number of parameters in the quark sector. The number n is the number of VLQs under consideration. The bold numbers are $SU(2)_{\mathrm{L}}$ multiplets [107].

- In little Higgs models, the previous gauge interaction with VLQs provokes an explicit symmetry breaking which transforms the Higgs boson into a pGB [122–124].
- In extra dimension models, the Kaluza-Klein modes produce new massive particles, potentially identified as VLQs. These particles manifest as excitations of SM particles [103].

The VLQs are not only predicted in BSM models to explain the hierarchy problem. Their properties can also answer many questions arising from the limitations of the SM.

- The VLQs offer a solution to the CKM unitarity problem. Measurements of meson and neutron decay processes have revealed a deviation of the unitarity of the CKM matrix at the level of three STandard Deviations (STD), or 3σ, from the SM predictions [125–127].
- The VLQs provide a simple explanation of the spontaneous CP-violation origin for every BSM theory mentioned before. In most of the theories, CP-symmetry violation is only introduced through *ad hoc* coefficients in the Yukawa sector. To predict a spontaneous CP-violation in a model, the Lagrangian must be CP-invariant, but the CP-invariance must be broken by the vacuum [108]. Any realistic model for spontaneous CP-violation must then generate both a CP-violating vacuum phase and a complex CKM matrix. By adding the VLQs to the model, these conditions are directly fulfilled as the coupling between the VLQs, the Higgs boson and the SM quarks introduces a FCNC interaction [106].
- The VLQs can play a key role for the strong CP problem. 't Hooft demonstrated in 1976 [128] that the strong sector of the SM leads to CP-violation similar to the EW sector. The Lagrangian includes a CP-violating term with the complex phase θ_{QCD}, known as the 19th parameter introduced in the SM, and originating from the QCD vacuum. However, the CP-violation in the strong sector has not yet been observed. A first solution was proposed by Peccei and Quinn [129, 130] leading to the prediction of new particles called axions, which protect the CP-violation in the strong sector, and explain the small value

of θ_{QCD}. But the VLQs can play the same role as axions in other models, where an additional CP-symmetry in the strong sector is spontaneously broken, and characterized by a naturally small θ_{QCD} phase [131, 132]

- It is possible to build models where all the CP-violations (SM weak sector, SM strong sector, and spontaneous CP-violation) have the same origin [133]. In this scenario, the previous models including VLQs merge to explain the different CP-violations simultaneously.

The VLQs are also predicted in the Grand Unified Theories (GUT) which aim to unify the electromagnetic, strong, and weak forces at high energies [134, 135]. Nevertheless, these models remain unconfirmed until the hypothetical particles they predict (above the TeV scale) can be experimentally identified. This would be a long-term objective for the future of the LHC.

Production mechanisms of Vector-Like Quarks

At the LHC, there are two ways of producing the VLQs: the pair production and the single production.

The VLQs are produced in pairs either via QCD or EW interactions (respectively up and down diagrams of Figure 1.5). Additional yet negligible processes can be found in Ref. [136]. The VLQ Q can be X, T, B or Y. In those diagrams, the two VLQs are produced at Leading Order (LO) in perturbative QCD (pQCD) in the final state, and could have the same electric charge or not. If they do, the production must be mediated by the t-channel exchange of a Higgs, W or Z boson, as illustrated in the two last subprocesses in Figure 1.5. Pair production is mostly dominated by QCD contributions [116], which is why we will only consider these latter for the rest of the chapter. A more complete description of the influence of EW contributions on the VLQs pair-production can be found in Ref. [137]. The cross section in the QCD-initiated VLQ production does not depend on the production mode but only on the mass of the VLQ. Being independent of the production mode has many advantages: the VLQs mass bounds are absolute, different production modes can be simultaneously considered, and the combination of several searches is simplified. Figure 1.7 presents the cross section at LO for both pair (dashed line) and single (solid line) production of a VLQ. The pair production dominates over single production for masses $M_Q < 650$ GeV and cross sections $\sigma > 50$ fb. For this reason, double-produced VLQ processes have been searched for in the first place. But pair production decreases faster than single production as a function of the VLQs mass. With the increasing center-of-mass energy of colliders through years, single production becomes a much more appealing process to detect the existence of the VLQs.

The VLQs are single-produced via EW interaction as presented in Figure 1.6. Additional yet even smaller cross section processes can be found in Ref. [136]. In the EW production, the VLQ Q is always produced associated with a heavy SM quark (top or bottom quark) coming from a gluon splitting, plus a light SM quark of the first or second generation [138]. The EW contributions depend on the nature of the VLQ, its mass, and the coupling strength between the VLQ, the SM quark, and the boson (Higgs, W or Z boson): the stronger the coupling, the higher the cross section. These information must always be presented in addition to the bounds on the VLQ single production cross sections. Because of their electric charge, the X (Y) can only be produced in association with a top (bottom) quark via a XtW (YbW) vertex. The T (B) can be produced in association with a bottom (top) quark via a TbW (BtW) vertex, or in association with a top (bottom) quark via a TtZ (BbZ) vertex. The cross sections at LO of these different modes of production are presented in Figure 1.7. They include the $SU(2)_{\mathrm{L}}$

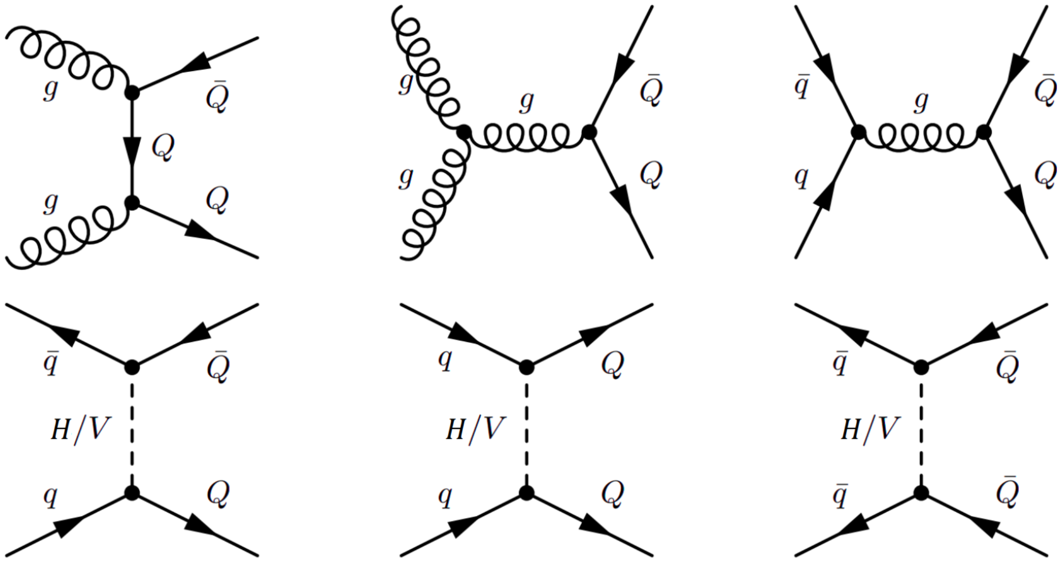


Figure 1.5: LO Feynman diagrams at tree level for pair production of a VLQ Q. They show the VLQ pair-production diagrams via strong (first line) and EW interactions (second line). The V corresponds to a W or Z boson [116].

singlets and doublets for the VLQs single production modes, in association with a jet j produced systematically in EW production. To compute the cross section, it is assumed that the mixing is the largest for each $SU(2)_\mathrm{L}$ multiplet. The dotted lines correspond to the experimental exclusions of the masses of VLQs in 2013 (more recent bounds are presented below). The production of $Y\bar{b}j$, $T\bar{b}j$, and $B\bar{b}j$ have the highest cross sections, especially for T in the $SU(2)_\mathrm{L}$ singlet representation (red line), and $(B\ Y)$ in the $SU(2)_\mathrm{L}$ doublet representation (yellow and purple lines). All these production modes have a higher cross section than the pair production for masses 650 GeV $< M_Q <$ 1500 GeV, hence their importance to identify other VLQs signatures at higher energies. In comparison, $T\bar{t}j$ and $X\bar{t}j$ production modes are negligible, and their cross section is always lower than that of the pair production for $M_Q <$ 2000 GeV.

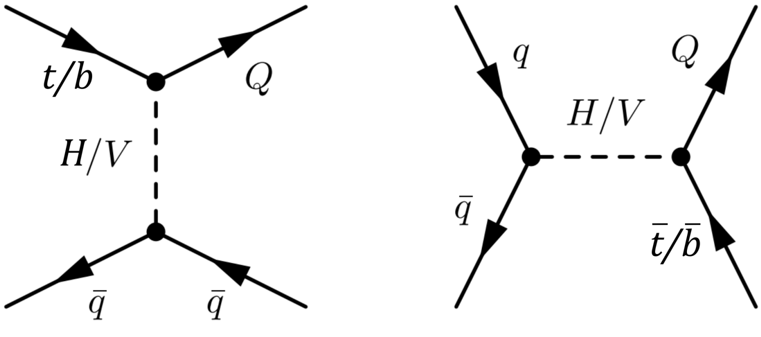


Figure 1.6: LO Feynman diagrams at tree level for single production of a VLQ Q. They show the VLQ single production diagrams via EW interaction in association with a SM light or heavy quark t/b. The V corresponds to a W or Z boson [116].

Decay modes of Vector-Like Quarks

Because the same couplings appear in the production and decay modes, the VLQs decay into one boson (Higgs, W or Z boson) and a SM quark. Other models propose VLQ decays into

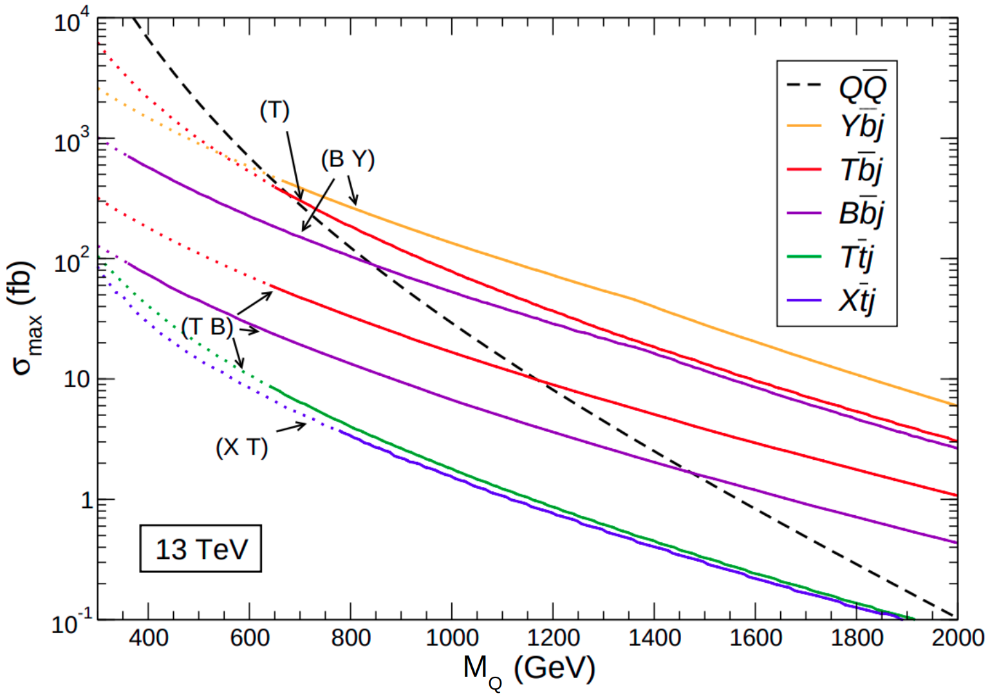


Figure 1.7: VLQs $SU(2)_{\mathrm{L}}$ multiplets production cross sections for double (dashed) and single (solid) production at the LHC at a center-of-mass energy of $\sqrt{s} = 13$ TeV. The dotted part of the lines indicate the range of masses already excluded by direct searches in 2013 [138].

non-SM particles, as we will touch on later, but we will not go into detail now. The decay rates of VLQs into SM particles are predicted as follows at LO [106]:

$$\begin{aligned}
\Gamma(Q_i \to W^+ d_j) =& \frac{g^2 M_{Q_i}^3}{64\pi m_W^2} \lambda[r_W, r_j] \left[\left(1 - r_j^2\right)^2 + r_W^2 \left(1 + r_j^2\right) - 2r_W^4 \right] |V'_{ij}|^2 \\
\Gamma(Q_i \to W^- u_j) =& \frac{g^2 M_{Q_i}^3}{64\pi m_W^2} \lambda[r_W, r_j] \left[\left(1 - r_j^2\right)^2 + r_W^2 \left(1 + r_j^2\right) - 2r_W^4 \right] |V'_{ji}|^2 \\
\Gamma(Q_i \to Z q_j) =& \frac{g^2 M_{Q_i}^3}{128\pi m_W^2} \lambda[r_Z, r_j] \left[\left(1 - r_j^2\right)^2 + r_Z^2 \left(1 + r_j^2\right) - 2r_Z^4 \right] |F^q_{ij}|^2 \\
\Gamma(Q_i \to H q_j) =& \frac{g^2 M_{Q_i}^3}{128\pi m_W^2} \lambda[r_H, r_j] \left[\left(1 + r_j^2\right)^2 - r_H^2 \left(1 + r_j^2\right) + 4r_j^4 \right] |F^q_{ji}|^2, \text{ with} \\
r_a = \frac{m_a}{M_Q}, \quad F^q =& V'V'^{\dagger} \text{ and } \lambda[a, b] = \sqrt{1 - 2\left(a^2 + b^2\right) + \left(a^2 - b^2\right)^2}.
\end{aligned} \tag{1.33}$$

The symbols r_a and $\lambda[a, b]$ are negligible for SM quarks of the first and second generations, with respect to the ones for SM quarks of the third generation. Therefore, we consider that the VLQs only decay into one Higgs, W or Z boson, and a top or bottom quark [139]. The possible decay channels are as follows:

$$\begin{gathered}
X \to W^+ t, \quad Y \to W^- b \\
T \to W^+ b, \quad T \to Ht, \quad T \to Zt \\
B \to W^- t, \quad B \to Hb, \quad B \to Zb.
\end{gathered} \tag{1.34}$$

The Branching Ratios (BRs) for the decay modes of T and B for all $SU(2)_{\mathrm{L}}$ multiplets are presented in Figure 1.8. The mass range of the VLQs goes from the lower-bound mass limits presented in Figure 1.7 (black crosses) to $M_Q = 2$ TeV (red dots), and the purple lines indicate the intermediary masses. Once again, we assume the largest mixing for every production mode. We do not consider X and Y as they only have a single decay mode. Only the neutral-current decays of T and B are presented in Figure 1.8, since the charge-current decay can be deduced from the following equations:

$$\begin{cases} \text{BR}(T \to W^+b) = 1 - \text{BR}(T \to Ht) - \text{BR}(T \to Zt). \\ \text{BR}(B \to W^-t) = 1 - \text{BR}(B \to Hb) - \text{BR}(B \to Zb) \end{cases} \tag{1.35}$$

The BRs are quite different between the $SU(2)_\text{L}$ singlets and doublets for both T and B. They range from 0 to a maximum of 50% for the neutral-current decays, and from 0 to 100% for the charged-current decay. The BRs can be modified by 15% at maximum for a mass variation of 1 TeV. For $SU(2)_\text{L}$ singlets, the BRs are set by convention in international collaborations as follows [106]:

$$\begin{cases} \text{BR}(T \to W^+b) = 0.5, \quad \text{BR}(T \to Ht) = 0.25, \quad \text{BR}(T \to Zt) = 0.25 \\ \text{BR}(B \to W^-t) = 0.5, \quad \text{BR}(B \to Hb) = 0.25, \quad \text{BR}(B \to Zb) = 0.25. \end{cases} \tag{1.36}$$

These relations are obtained with the equations 1.33, when r_a and $\lambda[a,b]$ are considered equal for every decay rate such that:

$$\begin{aligned} \Gamma(Q_i \to W^+d_j) &\simeq \frac{g^2 M_{Q_i}^3}{64\pi m_W^2}|V'_{ij}|^2 \\ \Gamma(Q_i \to W^-u_j) &\simeq \frac{g^2 M_{Q_i}^3}{64\pi m_W^2}|V'_{ji}|^2 \\ \Gamma(Q_i \to Zq_j) &\simeq \frac{g^2 M_{Q_i}^3}{128\pi m_W^2}|F^q_{ij}|^2 \\ \Gamma(Q_i \to Hq_j) &\simeq \frac{g^2 M_{Q_i}^3}{128\pi m_W^2}|F^q_{ji}|^2. \end{aligned} \tag{1.37}$$

This leads to:

$$\frac{\Gamma(Q_i \to W^+d_j)}{2|V'_{ij}|^2} \simeq \frac{\Gamma(Q_i \to W^-u_j)}{2|V'_{ji}|^2} \simeq \frac{\Gamma(Q_i \to Zq_j)}{|F^q_{ij}|^2} \simeq \frac{\Gamma(Q_i \to Hq_j)}{|F^q_{ji}|^2}. \tag{1.38}$$

If $|V'_{ji}| \simeq |F^q_{ji}|$ and $|V'_{ij}| \simeq |F^q_{ij}|$ is assumed, as is actually the case for third-generation SM quarks [106], we obtain the BRs presented in equations 1.36.

All the results presented so far assume the Narrow-Width Approximation (NWA), wherein the VLQs mass is treated as a Dirac delta function rather than a Breit-Wigner (BW) distribution. Moreover, only LO contributions are considered. However, higher-order corrections would change the behavior of these particles.

Impact of the Vector-Like Quarks width

There are two ways to achieve a large VLQ width Γ_Q: either the coupling strength of the single-produced VLQs with SM particles is increased, or the number of decay channels including non-SM particles is expanded.

- In the first case, the coupling strength between a single-produced VLQ Q_i and a SM quark q_j is implied in both the production and decay modes of Q_i through the term $|V'_{ij}|$ mentioned in equations 1.33. The modification of this term will directly alter the properties of Q_i, and particularly its width. For example, the addition of more than one

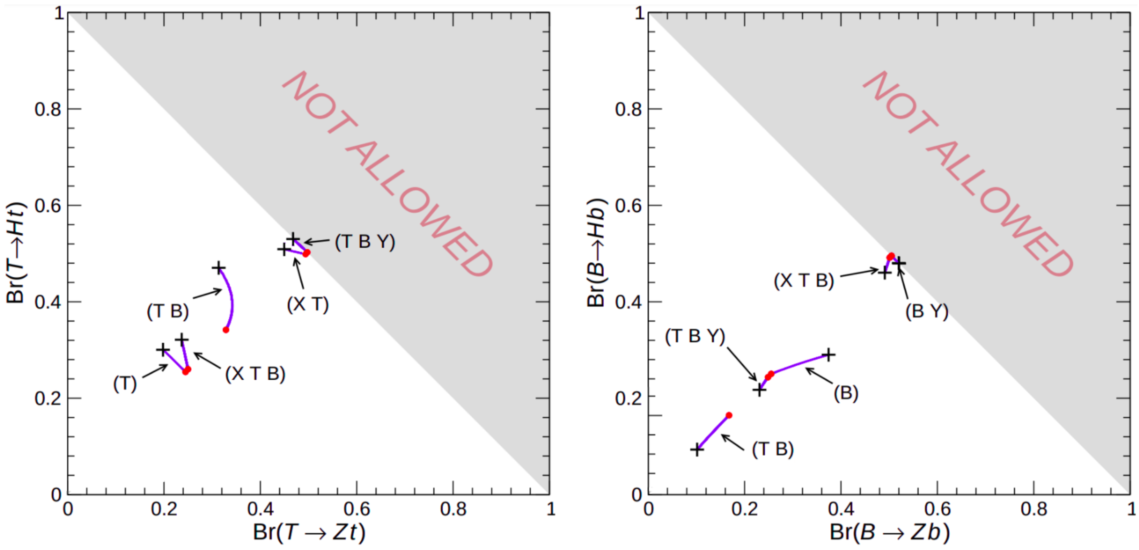


Figure 1.8: Allowed BRs for the decays of the VLQs T (left) and B (right) for $SU(2)_\mathrm{L}$ multiplets. The masses run from lower limits presented in Figure 1.7 (black crosses) to $M_Q = 2$ TeV (red dots) [138], with intermediary mass values in purple.

VLQ to the SM introduces additional interaction terms in the Lagrangian, and so new couplings in the matrix V' [140, 141]. These interactions not only affect the coupling of SM bosons (Higgs, W, and Z boson) to SM quarks, but also the mixing to first- and second-generation SM quarks which is not negligible anymore.

- In the second case, the BRs are completely altered as new final states emerge, defining distinct VLQs signatures. Since Γ_Q is the sum of the widths of all the different decay modes, the width automatically increases by introducing more decay channels. We can think about various decay modes beyond the SM: supersymmetric particles [118], new neutral or charged scalars [142, 143]...

Three schemes incorporating a non-zero finite VLQ width are explored in the literature.

- The propagator distribution includes a new term in the denominator of the width, resulting in a BW distribution [144]:

$$\mathrm{BW}(p^2) = \frac{1}{\left(p^2 - M_Q^2\right)^2} \to \frac{1}{\left(p^2 - M_Q^2\right)^2 + \Gamma_Q^2 M_Q^2}. \tag{1.39}$$

  p is the norm of the VLQ momentum vector. The introduction of the finite width prevents the divergence at $p^2 = M_Q^2$.

- In the complex mass scheme, the VLQ mass is modified such that [145]:

$$M_Q^2 \to \tilde{M}_Q^2 = M_Q^2 - iM_Q\Gamma_Q. \tag{1.40}$$

- The running width scheme uses a more intricate expression of the self-energy Σ_Q of the VLQ. The self-energy includes all corrections resulting from interaction between the VLQ and its environment, such that the propagator G is defined through the Dyson equation:

$$G = G_0 + G_0\Sigma_Q G, \tag{1.41}$$

  where G_0 is the bare propagator, without any correction. In the running width scheme, the self-energy can be expressed as [146]:

$$\Sigma_Q(p^2) = \frac{p^2}{M_Q^2}\Gamma_Q M_Q. \tag{1.42}$$

Let us consider the single production of a VLQ T in association with a bottom quark and a light quark. It could be mediated by the s-channel or t-channel, as depicted on the left in Figure 1.9. We assume that T has a mass $M_T = 1000$ GeV, and decays into a Higgs boson accompanied with a top quark. The most obvious distribution to study the effects of the VLQ width is the invariant mass of the T decay products M_{inv}. In the NWA, the distribution corresponds to a Dirac delta function located at the mass M_T. With a finite width, the distribution is modified, as presented on the right in Figure 1.9 for different width-over-mass ratios Γ_T/M_T and width schemes. The normalized distributions, and the ratio between each scheme and the BW scheme are shown. The shape of the invariant mass highly widens with an increasing Γ_T/M_T ratio, and the resonance is prominent for $\Gamma_T/M_T \leq 10\%$, with a finite size of $\Delta M_T \sim 100$ GeV. For $\Gamma_T/M_T \geq 30\%$, the peak at M_T is no longer prominent, and the invariant mass reaches a plateau at low energies. The resonance peak is asymmetric, and does interfere with SM background. The different width schemes produce distributions relatively similar at low energies, although the running width scheme tends to shift the peak to lower energies. The distributions are however quite different at high energies with more than a 50% difference for $M_{\text{inv}} > 1500$ GeV. The different schemes can then be treated as systematic uncertainties. The effects discussed here are similar for all types of VLQs and decay channels [147].

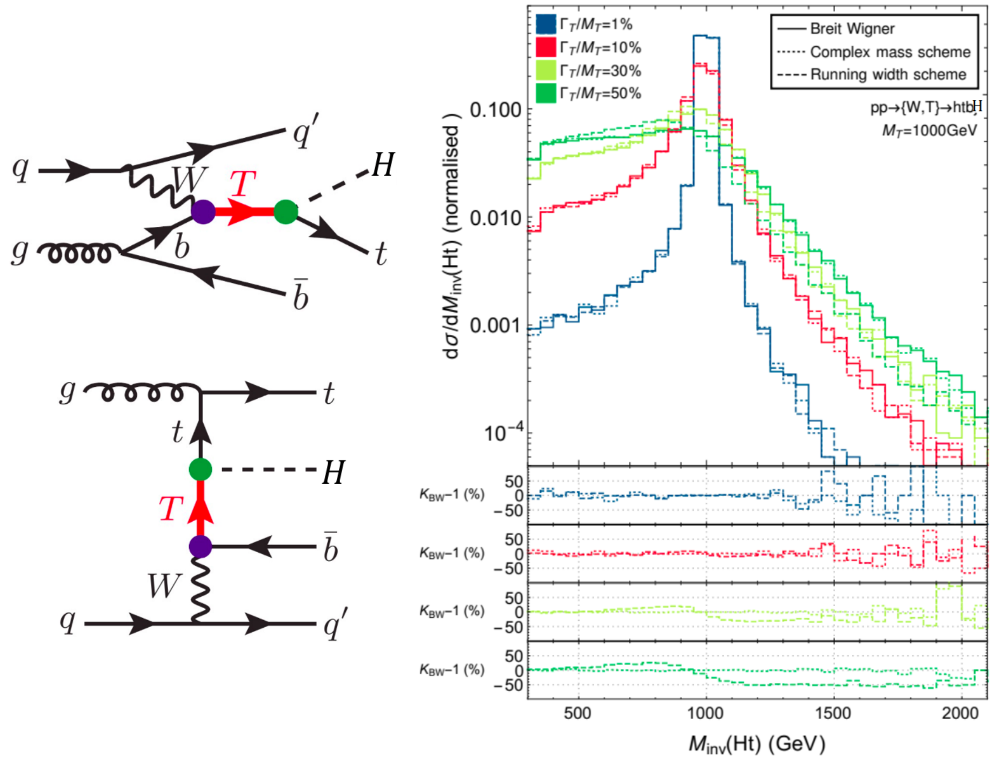


Figure 1.9: Invariant mass distribution for a single-produced VLQ T decaying into a Higgs boson accompanied with a top quark. The left panel shows the diagrams corresponding to the production modes of this process. The right upper panel shows the normalized distributions for different schemes and for several width-over-mass ratios. The right lower panel shows the ratio between each scheme and the BW scheme $K_{\text{BW}} - 1 = (\text{scheme} - \text{BW})/\text{BW}$ (%) [147].

Impact of higher-order corrections in QCD

Higher-order corrections are expected to substantially impact the predictions with colored VLQs [116, 148]. The Ref. [149] examines two schemes for the single production of the VLQ T at LO, and at Next-to-Leading-Order (NLO) QCD. In the Four-Flavor-Number Scheme (4FNS), a bottom quark is involved in the VLQ production, and an additional bottom quark is present in the final state, hence the process name Tjb. In the Five-Flavor-Number Scheme (5FNS), only the bottom quark involved in the VLQ production is considered with no extra bottom quark in the final state, hence the process name Tj. The variation of the effective cross section $\hat{\sigma}_T$ with the mass M_T, for the two schemes at LO and NLO QCD, is presented in Figure 1.10. The effective cross section $\hat{\sigma}_T$ is presented with its scale and parton density uncertainties (more details in Ref. [149]). In the 4FNS scheme, the cross section central value increases by 10% relative to LO for $M_T < 1100$ GeV, but decreases by 10% above this mass. In the 5FNS scheme, the cross section central value always decreases by 10% at NLO. The NLO corrections induce a significant reduction of the uncertainties by 50% compared to the LO results. The 4FNS and 5FNS schemes are also more compatible at NLO than at LO, especially for low masses, and the difference can be treated as a systematic uncertainty. A combination of both the finite width and higher-order corrections can be found in Ref. [147], but we will not go into detail as no experimental analysis uses this method currently.

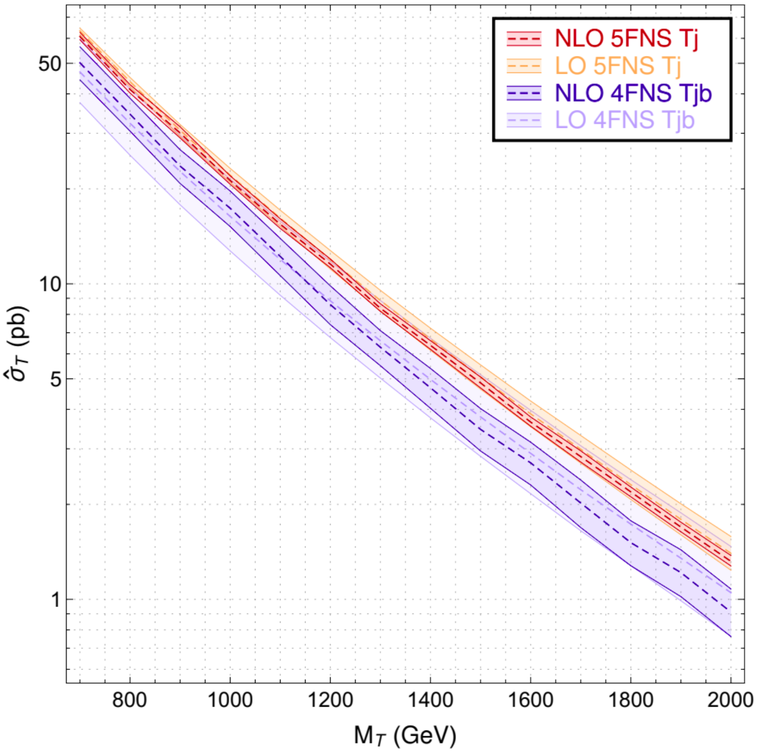


Figure 1.10: Cross section as a function of the single-produced VLQ T mass at LO in the 4FNS (light blue) and 5FNS (orange) schemes, and at NLO QCD in the 4FNS (dark blue) and 5FNS (red) schemes. Scale and parton density function uncertainties are included in the error bands [149].

Current results at the LHC

We present the various VLQ searches conducted at the LHC during the Run 2 period, and highlight the most updated findings. The ATLAS and CMS experiments collected data from the LHC at a center-of-mass energy of $\sqrt{s} = 13$ TeV, respectively between 2015 and 2018 and between 2016 and 2018. Searches focus on $SU(2)_L$ singlets for the VLQs T and B. The decay channel $T \to W^+b$ ($B \to W^-t$) is analogous to the decay channel $Y \to W^-b$ ($X \to W^+t$), allowing results for T and B in these specific decay channels to be extended to Y and X. The results assume the production of VLQs in the NWA at LO, with a coupling to the third-generation SM quarks only, and the BRs presented in equations 1.36 unless otherwise specified.

They are presented with a 95% Confidence Level (CL). It is important to note that the CDF and D0 experiments at the Tevatron, along with the ATLAS and CMS experiments during the Run 1 period (2010-2013), were not as relevant to perform searches, since data were collected at a respective center-of-mass energy of $\sqrt{s} = 1.96$ TeV and $\sqrt{s} = 7 - 8$ TeV. More information on VLQs searches during that period can be found in Ref. [106].

The pair production searches are production mode-independent, and do not need to specify the coupling between the VLQs and SM particles. The results are presented in Table 1.2. They are categorized based on the following studied final states.

- The all-hadronic event selection only requires jets.
- The semileptonic event selection requires exactly one electron or muon, Missing Transverse Energy (MET) coming from the neutrinos (see section 2.2.1), and jets.
- The Opposite Sign (OS) dilepton event selection requires exactly two oppositely-charged electrons or muons, MET, and/or jets.
- The Same Sign (SS) dilepton event selection requires exactly two same-charged electrons or muons, MET, and/or jets.
- The multilepton event selection requires three or more electrons or muons, MET, and/or jets.
- The diphoton event selection requires exactly two photons, electrons, muons, MET, and/or jets.

The single production searches are production mode-dependent, and need to specify the coupling $|V'_{Qq}|$ between the VLQs and SM particles, or equivalently the VLQ width-over-mass ratio Γ_T/M_T, and the production mode. The results are presented in Table 1.3, and are categorized like double production results. The first analysis, with a VLQ T decaying into a Z or Higgs boson, accompanied with a top quark in the all-hadronic final state, will be the analysis reference for this thesis. The results with the same VLQ T decaying into a Higgs boson accompanied with a top quark, but in the SS dilepton final state, will be presented in Chapters 4 and 5.

Experimental results suggest the potential exclusion of VLQs below the TeV scale. Results from both the CMS and ATLAS experiments show solid agreement across the different types of VLQs. The full Run 2 period analyses have low statistical errors, and establish precise constraints on the existence of VLQs. The major discrepancies in the results come from the studied production modes of VLQs. First, the searches using double-produced VLQs have better exclusion limits than single-produced VLQs. Second, the impact of the BR is not negligible, as indicated by the double-produced VLQ B in the all-hadronic final state [150]. Finally, the impact of coupling strength or width single-produced VLQs modifies the bounds by a factor of 2 at maximum, especially for $\Gamma_Q/M_Q = 25 - 30\%$. Beyond this threshold, interference with the SM becomes significant, and the VLQ Monte-Carlo (MC) generation does not include this effect.

Many improvements are necessary for a more accurate description of the VLQs in the future. First, simultaneous consideration of the width and higher-order corrections is essential. Most analyses to date have relied on simulated signal events at LO, despite the evident impact of higher-order corrections on the VLQs production modes and their associated uncertainties. Second, the BRs and the coupling to first- and second-generation SM quarks are crucial for better characterizing the VLQ signatures and widths, as mentioned before. These factors might be intricate as new decay channels will affect the BRs presented in equations 1.36 [151]. Besides, the results presented here focus on the VLQs $SU(2)_{\mathrm{L}}$ singlets only, while the BRs vary drastically across different $SU(2)_{\mathrm{L}}$ multiplets, as shown in Figure 1.8. Finally, the pair-production of the VLQs via EW interaction, as illustrated in Figure 1.5, must be consistently taken into account into the VLQ production process even with smaller cross sections. In future experiments following the LHC such as the Future Circular Collider (FCC) or the Compact LInear Collider (CLIC), the search for VLQs is expected to extend the bounds up to $M_Q \sim 3$ TeV for both single- [152] and double-produced VLQs [153, 154]. This will open prospective avenues of research for potentially uncovering the presence of the VLQs.

Experiment (L_{int} in fb^{-1})	Decay channels	Specific considerations	Final state	95% CL bounds (M_Q in GeV)	Refs.
CMS (35.9)	$T \to Wb/Ht/Zt$ $B \to Wt/Hb/Zb$	-	All-hadronic	$M_T > 960, \;\; M_B > 890$	[155]
ATLAS (36.1)	$T \to Wb/Ht/Zt$ $B \to Wt/Hb/Zb$	-	All-hadronic	$M_T > 1010, \;\; M_B > 1010$	[156]
CMS (137.6)	$B \to Hb/Zb$	BR($B \to Zb$) = 100% BR($B \to Hb$) = 100% BR($B \to Hb/Zb$) = 50%	All-hadronic	$M_B > 1390$ $M_B > 1570$ $M_B > 1450$	[150]
ATLAS (139.0)	$T \to Wb/Ht/Zt$ $B \to Wt/Hb/Zb$	-	Semileptonic	$M_T > 1260, \;\; M_B > 1330$	[157]
CMS (35.9)	$T \to Wb/Ht/Zt$ $B \to Wt/Hb/Zb$	NNLO QCD	OS dilepton	$M_T > 1095, \;\; M_B > 955$	[158]
ATLAS (36.1)	$T \to Wb/Ht/Zt$ $B \to Wt/Hb/Zb$	NNLO QCD	OS dilepton	$M_T > 1030, \;\; M_B > 1010$	[159]
ATLAS (36.1)	$T \to Wb/Ht/Zt$ $B \to Wt/Hb/Zb$	-	SS dilepton	$M_T > 980, \;\; M_B > 1000$	[160]
CMS (137.6)	$T \to Wb/Ht/Zt$ $B \to Wt/Hb/Zb$	-	multilepton	$M_T > 1480, \;\; M_B > 1470$	[161]
ATLAS (139.0)	$T \to Wb/Ht/Zt$ $B \to Wt/Hb/Zb$	-	multilepton	$M_T > 1270, \;\; M_B > 1200$	[162] [163]

Table 1.2: Exclusion limits at 95% CL for the $SU(2)_{\mathrm{L}}$ singlets VLQs T and B in pair production in multiple final states. They are presented at a center-of-mass energy of $\sqrt{s} = 13$ TeV.

Experiment (L_{int} in fb^{-1})	Decay channels	Specific considerations	Final state	95% CL bounds (M_Q in GeV)	Refs.
CMS (137.6)	$T(bW/tZ) \to Ht/Zt$	-	All-hadronic	$M_T > 650,\ \Gamma_T/M_T = 5\%$ $M_T > 600,\ \Gamma_T/M_T = 10\%$	[164] [165]
ATLAS (139.0)	$T(bW/tZ) \to Ht$	-	All-hadronic	$M_T > 1200,\ \Gamma_T/M_T = 5\%$ $M_T > 1500,\ \Gamma_T/M_T = 10\%$ $M_T > 1600,\ \Gamma_T/M_T = 25\%$ $M_T > 1700,\ \Gamma_T/M_T = 50\%$	[166]
CMS (35.9)	$B(tW/bZ) \to Hb$	-	All-hadronic	$M_B > 800,\ \Gamma_B/M_B = 20\%$ $M_B > 1150,\ \Gamma_B/M_B = 30\%$	[167]
ATLAS (139.0)	$B(tW/bZ) \to Hb$	-	All-hadronic	$M_B > 1600,\ \Gamma_B/M_B = 5\%$ $M_B > 2200,\ \Gamma_B/M_B = 10\%$	[168]
CMS (137.6)	$T(bW) \to Ht$	-	Diphoton	$M_T > 720,\ \Gamma_T/M_T = 1\%$ $M_T > 770,\ \Gamma_T/M_T = 2\%$ $M_T > 860,\ \Gamma_T/M_T = 3\%$ $M_T > 940,\ \Gamma_T/M_T = 4\%$ $M_T > 1000,\ \Gamma_T/M_T = 5\%$	[169]
ATLAS (36.1)	$T(bW) \to Wb$	NLO QCD	Semileptonic	$\lvert V'_{Tb}\rvert < 0.18,\ M_T = 800$ $\lvert V'_{Tb}\rvert < 0.35,\ M_T = 1200$	[170]
ATLAS (139.0)	$T(bW/tZ) \to Ht/Zt$	-	Semileptonic	$M_T > 1100,\ \Gamma_T/M_T = 5\%$ $M_T > 1700,\ \Gamma_T/M_T = 10\%$ $M_T > 1900,\ \Gamma_T/M_T = 20\%$ $M_T > 2100,\ \Gamma_T/M_T = 50\%$	[171]
CMS (35.9)	$B(tW/bZ) \to Wt$	-	Semileptonic	$M_B > 900,\ \Gamma_B/M_B = 10\%$ $M_B > 1300,\ \Gamma_B/M_B = 20\%$ $M_B > 1450,\ \Gamma_B/M_B = 30\%$	[172]
CMS (35.9)	$T(bW) \to Zt$	-	Semileptonic/ All-hadronic	$M_T > 975,\ \Gamma_T/M_T = 5\%$ $M_T > 1050,\ \Gamma_T/M_T = 10\%$ $M_T > 1250,\ \Gamma_T/M_T = 20\%$ $M_T > 1350,\ \Gamma_T/M_T = 30\%$	[173]
ATLAS (36.1)	$T(bW) \to Zt$	-	Semileptonic/ All-hadronic	$M_T < 1400,\ \lvert V'_{Tb}\rvert < 0.7$	[174]
ATLAS (139.0)	$T(bW/tZ) \to Zt$	-	OS dilepton/ multilepton	$M_T > 1600,\ \Gamma_T/M_T = 10\%$ $M_T > 1850,\ \Gamma_T/M_T = 30\%$ $M_T > 1950,\ \Gamma_T/M_T = 50\%$	[175]

Table 1.3: Exclusion limits at 95% CL for the $SU(2)_{\mathrm{L}}$ singlets VLQs T and B in single production in multiple final states. They are presented at a center-of-mass energy of $\sqrt{s} = 13$ TeV.

Chapter 2

The CMS experiment at the Large Hadron Collider

During World War II, experimental research in nuclear physics focused on the development of the atomic bomb. But the end of the conflict completely changed the objectives of the researchers, especially in Europe. Louis de Broglie was the first to propose the establishment of a European scientific laboratory in a neutral zone, in Switzerland. This initiative led to the creation of the European Organization for Nuclear Research, commonly known as the 'Conseil Européen pour la Recherche Nucléaire' (CERN) in French, by 12 European governments in 1952. Although the acronym has remained unchanged, CERN's initial focus on studying atomic nuclei quickly shifted to high-energy physics, and the interaction between subatomic particles. Since its inauguration, CERN has played a major role in discovering and characterizing all the particles discussed in Chapter 1 [176]. In 1994, CERN approved the construction of a new accelerator, designed to achieve unprecedented energies for proton-proton and heavy ions collisions: the Large Hadron Collider (LHC).

The LHC is the largest and most powerful accelerator in the world since the first collisions of the Run 1 period in 2010. It consists of a 27-kilometer ring, where hadrons are accelerated to high energies, and subsequently collided at four locations. Each of the collision points is associated with a specific experiment: ATLAS and CMS which are two general-purpose experiments, A Large Ion Collider Experiment (ALICE) which studies heavy ions and quark-gluon plasma properties, and Large Hadron Collider beauty (LHCb) dedicated to B-meson physics and CP-violation. Following the discovery of the Higgs boson in 2012 [71, 72], the CMS experiment in particular has expanded its physics objectives to include both probing the SM of particle physics through precision measurements, and investigating BSM physics. The results presented in the following chapters will focus on the CMS experiment.

This chapter will be divided into two sections. The first section will provide an overview of CERN and the LHC. Some of the characteristics described here are indeed not exclusive to CMS, but are general considerations which apply to all experiments at the LHC. The second section will focus on the CMS experiment. The design of every sub-detector is crucial to identify the particles produced as a result of the collisions. Additionally, the data processing will be explained to demonstrate how we can transform simple electronic signals in the detector into usable data, for a comprehensive particle physics analysis. Eventually, we will discuss the perspectives of the LHC and the CMS experiment for the upcoming data-taking period, known as the High-Luminosity (HL) LHC, and starting in 2029. The presentation of this chapter is inspired by the work which can be found in Ref. [177–179].

2.1 CERN and the Large Hadron Collider

2.1.1 Key steps in CERN history

The first accelerator ever constructed at CERN was the Synchro-Cyclotron (SC) in 1957, with a circumference of 628 m. It accelerated proton beams up to 600 MeV. It was soon replaced by the Proton Synchrotron (PS) in 1959, which accelerated protons up to 28 GeV [178]. The Gargamelle experiment, which hinted at the existence of W and Z bosons in 1971, was conducted at the PS. That same year, the Intersecting Storage Rings (ISR) were constructed using the PS as an injector. It became the world's first hadron collider, accelerating protons up to 62 GeV.

In 1971, a new site was constructed at CERN to house the Super Proton Synchrotron (SPS). Serving as the first underground collider at CERN, it crossed the border between France and Switzerland, with a circumference of 6.9 km. Protons were accelerated up to 400 GeV, enabling the confirmation of the existence of W and Z bosons in 1983 through the UA1 and UA2 experiments. The advancement in particle detection was highly enhanced by the introduction of new detectors named multi-wire proportional chambers, designed by Charpak in 1968 [180]. These chambers replaced bubble chambers, facilitating data processing through computers, and increasing the counting rate by a factor of 1,000.

In 1989, the Large Electron-Positron (LEP) collider was inaugurated, with a record circumference of 26.7 km. Electrons and positrons were accelerated up to 209 GeV, making it the most powerful lepton collider to date. LEP's contributions include the confirmation that only three generations of leptons couple to the Z boson [181], and the precise measurement of the Z boson's peak line shape [182]. In 2000, LEP was dismantled, and replaced by the current LHC collider [183] (see section 2.1.2).

In addition to the collaborations directly associated with the colliders over time, CERN initiated smaller experiments which took advantage of particle collisions to investigate more specific subjects. In 1995, the PS210 experiment confirmed the creation of antihydrogen atoms, utilizing the Low Energy Antiproton Ring (LEAR) with the PS as an injector [184]. In 1999, the evidence for direct CP-violation was reported in neutral kaons in the NA48 experiment, using the SPS as an injector [185]. It it finally noteworthy that CERN is the birthplace of the World Wide Web in 1989. Tim Berners-Lee and Rover Cailliau developed a hypertext information system allowing CERN physicists to communicate globally. This marked the birth of the first website, followed by the release of this system under a free license.

2.1.2 The Large Hadron Collider

The primary goal of the LHC was to confirm the existence of the Higgs boson. Following its discovery by both the CMS and ATLAS experiments in 2012 [71, 72], a broader 'Higgs program' began at the LHC, covering a wide range of measurements aiming at elucidating the properties of this particle. With a center-of-mass energy on the order of several TeV, the LHC also serves as an ideal instrument for investigating BSM phenomena, and identifying signatures of new particles. Finally, the LHC plays a crucial role in advancing the precision measurements previously conducted at the LEP in QCD, electroweak, and flavor physics.

General description of the collider

The LHC lies in a tunnel with a circumference of 26.7 km and a diameter of 3.7 m, situated at depths ranging from 45 to 175 m, and with a slope of 1.4% [186]. It uses the same infrastructure as the LEP, but it primarily uses protons or heavy ions instead of leptons. The center-of-mass energy is $\sqrt{s} = 7$ and 8 TeV, 13 TeV, and 13.6 TeV for proton-proton collisions during the Run 1 (2010-2012), Run 2 (2015-2018), and Run 3 (2022-2025) periods respectively, making the LHC the hadron collider with the highest center-of-mass energy to date. Inside the accelerator, proton beams traverse two tubes maintained at an ultrahigh vacuum of 10^{-13} atm to prevent unwanted collisions. They are guided by a powerful magnetic field of 8.3 T, generated by 9,593 copper-clad niobium-titanium (TiNb) superconducting magnets [178]. These magnets include 1,232 dipole magnets which bend the beam, and 392 quadrupole magnets which maintain particles focus in narrow beams. Additionally, other quadrupoles compress the beams just before the interaction point, while magnets of higher-orders correct minor deviations in the magnetic field. The superconducting state is achieved at a temperature of 1.9 K, while the tunnel is globally cooled at 4.5 K. Approximately 96 tons of superfluid helium-4 ^{4}He are employed to maintain these conditions.

The LHC machine is part of a large complex including the various accelerators mentioned earlier [187]. At each step, the energy of the proton beam is progressively increased before it is injected into the main accelerator. The complete chain is illustrated in Figure 2.1, and follows this procedure.

- Step 1: hydrogen ions H^- are generated by introducing hydrogen gas into an ion source. These ions are then injected into the LINear Accelerator 4 (LINAC4), where they are accelerated up to 160 MeV.
- Step 2: electrons are stripped from hydrogen ions as they pass through a thin carbon foil. This process leaves only nuclei containing a single proton. The resultant proton beam is then injected into the Proton Synchrotron Booster (PSB). It is composed of four superimposed rings, where protons are bunched together, and accelerated up to 2 GeV before being merged into a single proton beam.
- Step 3: the proton beam is directed to the PS, where its energy is increased to 26 GeV. Proton bunches are divided into smaller bunches, and brought closer together.
- Step 4: the proton beam is directed to the SPS, which pushes its energy to 450 GeV. This is the final stage before injection into the LHC.

Within the LHC, protons are divided into two counter-rotating beams, and further accelerated to their maximum energy level. This process takes approximately 20 minutes. It is achieved using metallic chambers, known as Radio Frequency (RF) cavities, which contain an electromagnetic field. There are a total of 8 RF cavities, located at 4 different points along the accelerator, and each contributing to an energy increase of 0.5 MeV per revolution such that the final center-of-mass energy is $\sqrt{s} = 13.6$ TeV. Under nominal operating conditions, the beams can circulate for several hours before they have to be re-filled. The procedure for lead-lead collisions follows a similar pattern, where all electrons are stripped from lead nuclei, and lead ions Pb^{82+} are accelerated through the Low Energy Ion Ring (LEIR). The next steps remain identical to those for protons.

Beam characteristics in proton-proton collisions

Protons are arranged into bunches within beams, with each bunch containing 115 billions of protons under nominal operating conditions. These bunches are prepared along the injection

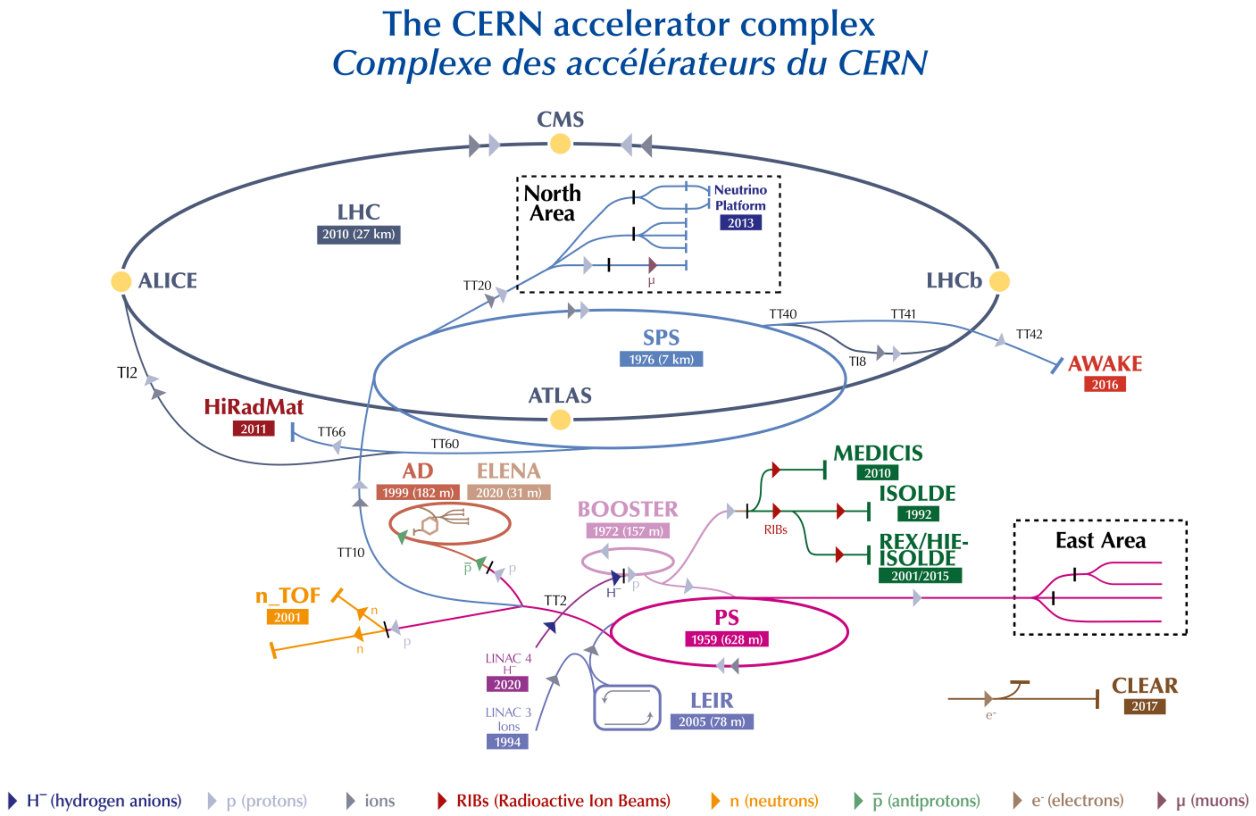


Figure 2.1: Scheme of the CERN accelerator complex and the LHC injection chain [188].

chain, and are separated by a time interval of $\Delta\tau = 25$ ns. Consequently, collisions between two beams occur at discrete intervals, with a frequency of approximately 40 MHz. The gaps between the bunches are used for synchronization, calibration data acquisition, and provide resets to front-end electronics [189]. Overall, the LHC accommodates 2,808 bunches during each fill [190]. The protons beams are collimated to minimize the spatial extension of the bunches, and maximize the number of collisions detected at the interaction point. The latter defines the instantaneous luminosity L [8]:

$$L = \frac{N_p^2 n_b f_{rev}}{4\pi\sigma_x\sigma_y} F, \tag{2.1}$$

where N_p is the number of particles per bunch, n_b the number of bunches per beam, f_{rev} the revolution frequency, σ_x and σ_y the transverse widths of the beam along the respective x and y-axes, and F the geometric factor for the reduction of the bunch size, due to the crossing angle at the interaction point. The values of these parameters are summarized in Table 2.1.

Center-of-mass energy	$\sqrt{s} = 13.6$ TeV
Number of protons per bunch	$N_p = 1.15 \times 10^{11}$
Number of bunches per beam	$n_b = 2.808 \times 10^3$
Horizontal extension of bunches	$\sigma_x = 16.7$ m
Vertical extension of bunches	$\sigma_y = 16.7$ m
Time interval between bunches	$\Delta\tau = 25$ ns
Revolution frequency of bunches	$f_{rev} = 11.25$ kHz
Crossing angle at the interaction point	285 μrad
Geometric reduction factor	$F \sim 10^{-4}$
Number of proton-proton collisions per beam	$\sim$50
Instantaneous luminosity	$L = 10^{34}$ cm^{-2}.s^{-1}

Table 2.1: Nominal values of LHC characteristics for proton-proton collisions in the Run 3 period.

The number of events produced per second as a result of a collision for a given process is determined by the product of the cross-section σ of this process and L, such that:

$$\frac{\partial N_{event}}{\partial t} = \sigma L. \tag{2.2}$$

The values of the cross-section for proton-proton collisions can be found for different center-of-mass energies at the bottom right corner of Figure 2.2. The LHC has been designed to operate at a nominal value of $L = 10^{34}$ cm^{-2}.s^{-1}. However, L gradually decreases along the fill, due to beam losses during collisions, and requiring periodic re-filling. The total amount of data delivered to the experiments is represented by the integrated luminosity L_{int} through the equation $L_{int} = \int_0^t L \, \mathrm{d}t$, as illustrated in Figure 2.3 for CMS.

Luminosity is a critical parameter of the LHC machine, and must be maximized to enhance the potential for new discoveries. However, higher luminosity levels lead to multiple proton-proton collisions, where interesting high-energy, or 'hard' collisions are accompanied by additional unwanted, or 'soft' interactions. Hard collisions occur when one parton (see section 1.1.2) from each proton exchanges an important fraction of momentum, resulting in particle production with a large transverse momentum p_T. On the other hand, soft collisions can be either elastic, diffractive, or inelastic, and produce numerous particles with a small p_T along the beam pipe, typically described by non-perturbative QCD [191]. These collisions, known as PileUp (PU), are illustrated in Figure 2.2 for CMS. The PU pollutes the final state of interest, complicating the reconstruction process. Special techniques are employed during the reconstruction process to mitigate PU, as it will be discussed in section 3.2.2.

The LHC data-taking: Run 1, Run 2, Run 3 periods and perspectives

The operation of the LHC began in 2010, with the first physics data colliding at $\sqrt{s} = 7$ TeV, as illustrated in Figure 2.4. This marked the beginning of the first data-taking period, known as the Run 1 period, with a maximum center-of-mass energy $\sqrt{s} = 8$ TeV. During this period, a total luminosity $L_{int} = 29.4$ fb^{-1} of data were collected at CMS, with a PU ranging from 10 to 20. Following a Long Shutdown (LS) from 2013 to 2015, the LHC entered its second operational phase, known as the Run 2 period, reaching $\sqrt{s} = 13$ TeV. Until 2018, a total luminosity $L_{int} = 159.3$ fb^{-1} of data were collected, with an average PU number of 35. Another LS (LS2) took place from 2018 and 2022, to maintain and upgrade the accelerator complex. Operations resumed on 22 April 2022, initiating the Run 3 data-taking period with $\sqrt{s} = 13.6$ TeV. Since

2022, a total luminosity $L_{int} = 177.7\ \mathrm{fb}^{-1}$ of data were collected, surpassing the Run 2 total luminosity, with a PU number ranging from 45 to 60. The LHC will continue to operate in the current configuration until the end of 2025, with an expected luminosity of $L_{int} = 300\ \mathrm{fb}^{-1}$ for the full Run 3 period.

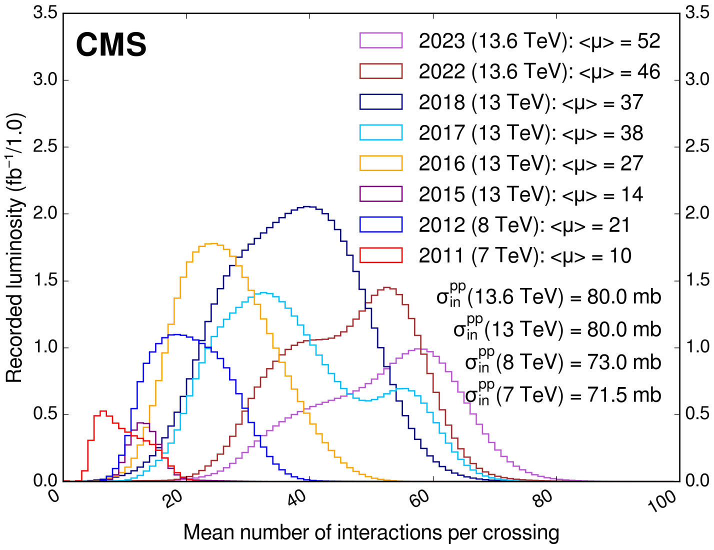


Figure 2.2: Number of proton-proton collisions per beam or PU during the Run 1, Run 2 and Run 3 periods. The mean PU per year and the total cross-section for proton-proton collisions per center-of-mass energy are shown in the legend [192].

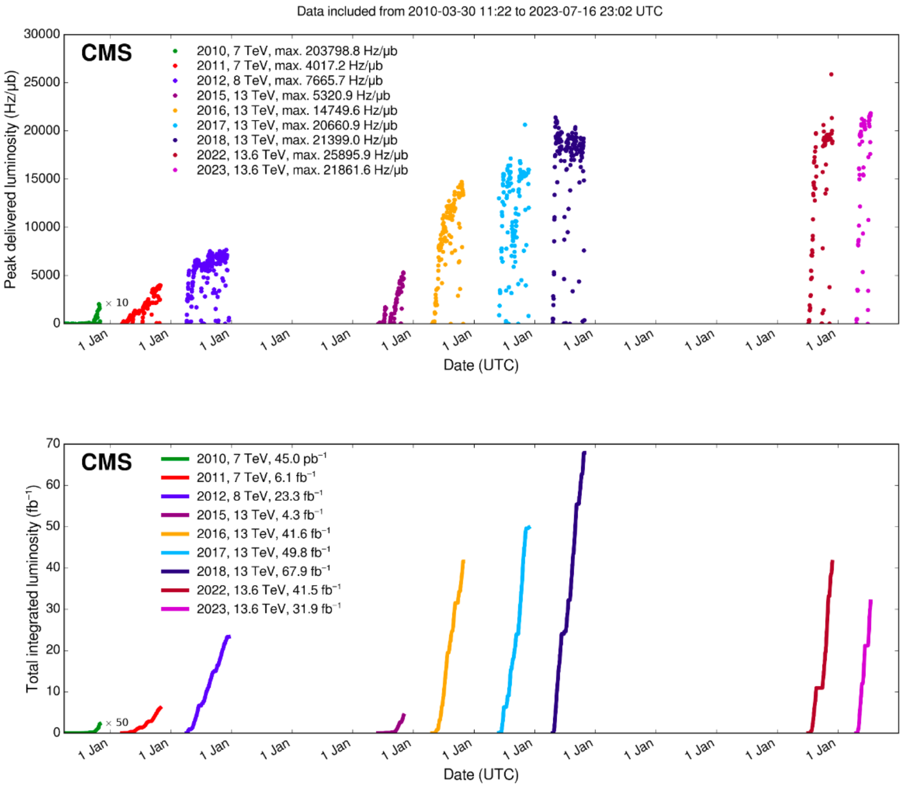


Figure 2.3: Peak L (top) and L_{int} (bottom) delivered to CMS versus day during the Run 1, Run 2 and Run 3 periods [192].

Following the Run 3 period, a 3-year LS (LS3) will lead the LHC to undergo a major upgrade, and start a new era: the HL-LHC [193]. The primary objective is to increase the instantaneous luminosity by a factor of 5, relative to the original design value in 2010. The targeted integrated luminosity is $L_{int} = 3{,}000\ \text{fb}^{-1}$ over the 10-year operation period, with an associated center-of-mass energy $\sqrt{s} = 14$ TeV. While this upgrade promises new possibilities for discoveries and precision measurements, it also presents unprecedented technical challenges for the experiments. One of the biggest concerns stems from the significant increase in the number of PU events, projected to rise from an average number of 50 during the Run 3 period to approximately 140-200 interactions per beam during the HL-LHC phase. To address this issue, the LHC experiments have already undergone various detector upgrades during LS2 (see section 2.2.1), with further enhancements planned for LS3. Furthermore, the development of new detectors is underway to mitigate the effects of PU [179]. This will be discussed more in detail in section 2.2.4 for the CMS experiment.

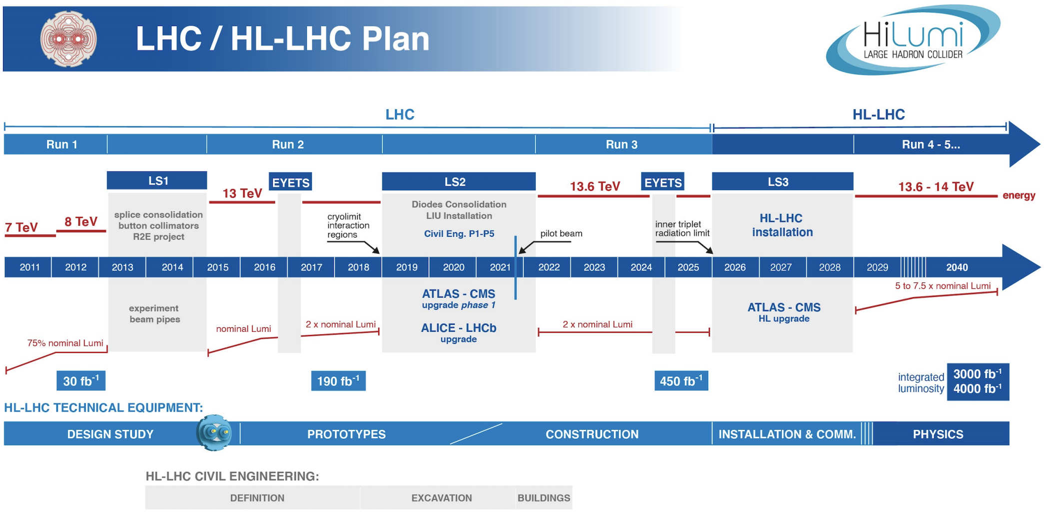


Figure 2.4: The timeline of the current LHC and the following HL-LHC [194].

2.2 The CMS experiment

The CMS experiment is a general-purpose detector located in an underground cavern in Cessy, France [195], and stands as one of the largest scientific collaborations worldwide, involving approximately 6,000 members from 57 countries. The physics scope of CMS is to probe the SM of particle physics, and search for BSM physics through proton-proton and lead-lead collisions.

2.2.1 Description of the CMS detector

An overview of the CMS detector is presented in Figure 2.5. The CMS detector is relatively compact, measuring 21.0 meters in length, 15.0 meters in width and 15.0 meters in height, with a weight of approximately 14,000 tons. It comprises various sub-detectors arranged in cylindrical layers around the beam axis, each serving a specific purpose. Starting from the interaction point, produced particles from proton-proton collisions enter the tracker first. A strong magnetic field of 3.8 T bends the trajectories of charged particles, enabling the measurement of their electric charges and momenta in the tracker. Electrons and photons are captured by the Electromagnetic CALorimeter (ECAL), where their energies and directions are determined from ElectroMagnetic (EM) showers. Both charged and neutral hadrons are detected in the Hadron CALorimeter (HCAL), where their energies and directions are determined from hadronic showers. Hadrons can generate showers within the ECAL, which are absorbed in the HCAL afterwards. Muons, which can pass through the calorimeters with little to no interaction, are measured in the tracker and muon chambers, representing the final layer of the CMS experiment.

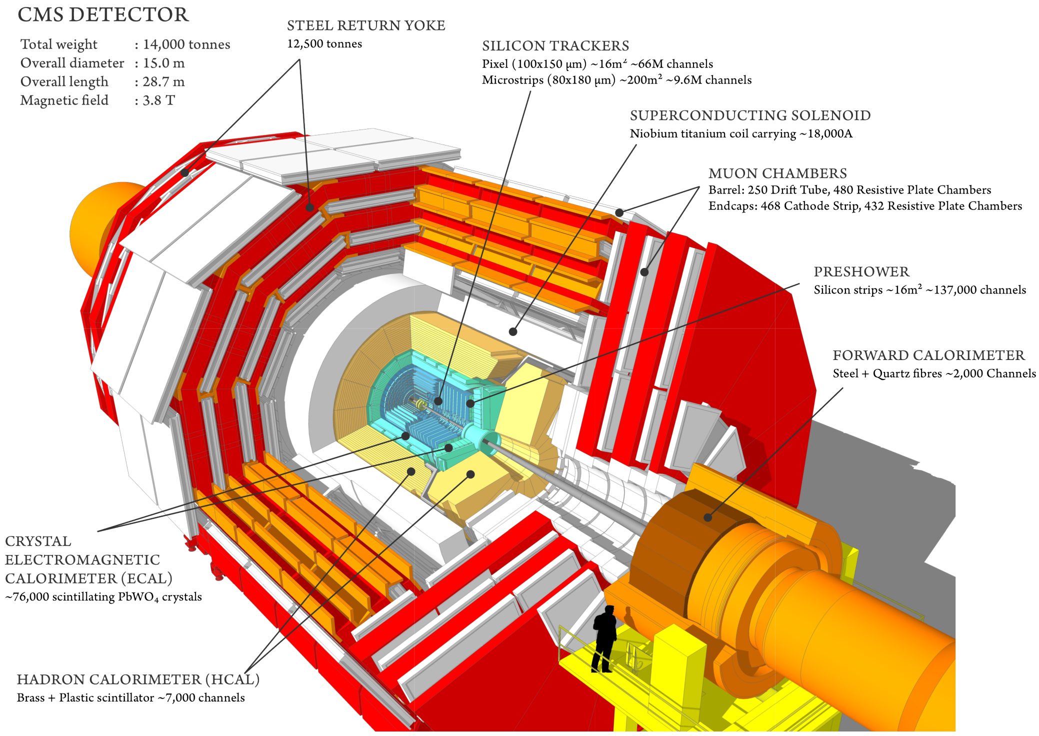


Figure 2.5: The CMS detector overview [196].

Coordinate system

The coordinate system used by the CMS experiment is shown in Figure 2.6. Its origin is centered at the collision point, with the x-axis pointing toward the center of the LHC, the y-axis pointing vertically upward, and the z-axis aligned along the beam direction. The momentum vector is denoted by $\vec{p}$, and the components in the three directions are denoted by p_x, p_y, and p_z respectively. Considering the cylindrical shape of the detector, polar coordinates are frequently used. The polar angle θ is measured from the z-axis, while the azimuthal angle φ is measured from the x-axis in the x-y plane, also known as the transverse plane. In this plane, the transverse momentum vector is denoted by $\vec{p_T}$, and its norm is defined as:

$$p_T = \sqrt{p_x^2 + p_y^2}. \tag{2.3}$$

Given that the collisions take place along the beam axis, the total transverse momentum vector is expected to be null. However, an imbalance of energy in the transverse plane may arise due to the presence of neutrinos in the final state or detector miscalibration, and should be minimized. This phenomenon is referred to as the MET (see section 2.2.2).

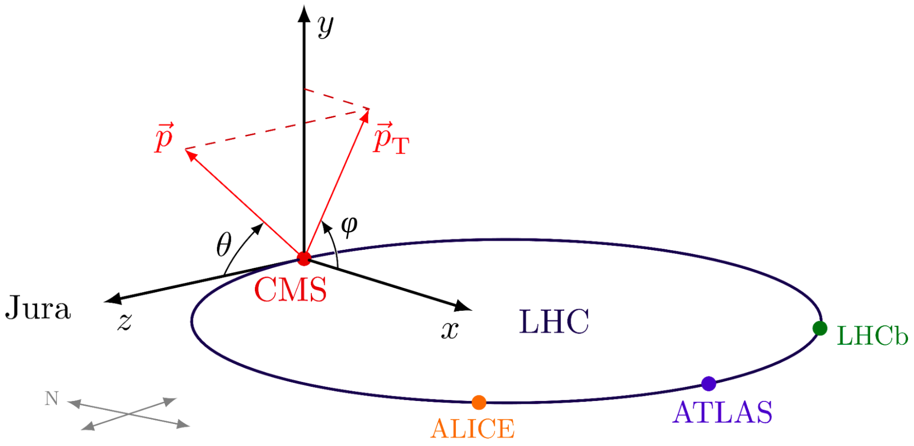


Figure 2.6: The CMS coordinate system [197].

The pseudorapidity η is defined from θ as:

$$\eta = \frac{1}{2}\log\left(\frac{p+p_z}{p-p_z}\right) = -\log\left[\tan\left(\frac{\theta}{2}\right)\right]. \tag{2.4}$$

The values of η corresponding to different θ angles are shown in Figure 2.7. Pseudorapidity is a critical measure in particle physics. When partons collide within the detector, they indeed carry a fraction of the initial proton's momentum, resulting in varying longitudinal boosts in their center-of-mass frames. However, the difference in pseudorapidity $\Delta\eta$ between two particles is Lorentz invariant along longitudinal boosts. Therefore, if the emission angle of each particle is defined by (η, φ), the spatial separation between two particles $\Delta R(1,2)$ is invariant with respect to boosts along the beam axis, and is defined as:

$$\Delta R(1,2) = \sqrt{\Delta\eta^2 + \Delta\varphi^2} = \sqrt{|\eta_2 - \eta_1|^2 + |\varphi_2 - \varphi_1|^2}, \tag{2.5}$$

where η_i and φ_i are respectively the pseudorapidity and the azimuthal angle of the particle i.

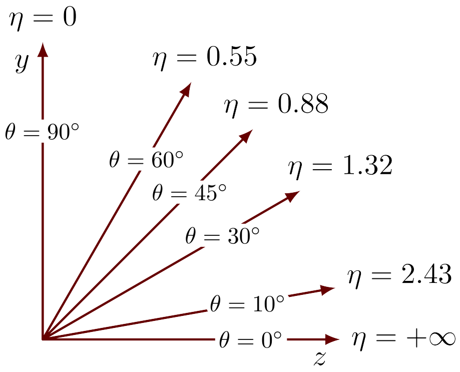


Figure 2.7: Illustration of the relation between the η and θ values [197].

The tracker

The inner tracker system is a cylindrically-shaped detector positioned the closest to the interaction point, as illustrated in Figure 2.8. The main function of the tracker is to reconstruct the trajectory of charged particles, known as 'tracks', which are curved due to the presence of the magnetic field created by the CMS magnet. Measuring the particle tracks provide precise information on the location of particle collision, known as the vertex. Typically, we distinguish the Primary Vertex (PV), resulting from a hard collision, from the Secondary Vertex (SV), resulting from a soft collision.

The tracker has an outer radius of 1.20 m and a length of 5.6 m, covering the pseudorapidity range up to $|\eta| = 2.5$. It is divided into two parts: the pixel tracker and the silicon microstrip detector.

- The pixel tracker constitutes the innermost part of the tracking system, situated in a particularly harsh radiation environment [198]. It is divided into 1,856 segmented sensor modules, each consisting of 160×416 pixels, and depicted as green lines in Figure 2.8. The total silicon area is 1.9 m^2 with a standard pixel size of 100×150 $\mu\mathrm{m}^2$.

- Surrounding the pixel tracker is the silicon microstrip detector, comprising 24,244 sensor modules and a total of 9.6 millions strips [199]. Sensors are designed in 15 different geometries for an active area of 198 m^2, and depicted as blue and red lines in Figure 2.8. Modules made of silicon were chosen thanks to their semiconducting properties and low production costs.

The tracker is segmented into four parts: the Tracker Inner Barrel (TIB) and Tracker Outer Barrel (TOB), featuring modules arranged in concentric layers, and the Tracker Inner Disk (TID) and Tracker End Caps (TEC), where modules are arranged in disks. The transition between cylindrical layers and disk layers is progressive from $|\eta| = 0.8$ to $|\eta| = 1.6$, as illustrated in Figure 2.8.

The fine granularity of the detector allows for the differentiation of closely spaced particle tracks. For high-p_T tracks exceeding 100 GeV, the p_T resolution is approximately 1-2% in the barrel and slightly worse in the endcap, due to the shorter lever arm of these tracks in the transverse plane of the tracker [200]. However, the tracker represents a significant amount of material in front of the calorimeters. At $\eta \approx 1.5$, there is an approximate 85% probability for a photon to convert, or for an electron to emit a bremsstrahlung photon upon interacting with this material. Similarly, a hadron has a 20% probability of experiencing a nuclear interaction before reaching the ECAL. These complications need to be accounted for in the final object reconstruction process (see section 2.2.2). Lastly, the tracker, and especially the pixel tracker, is the CMS sub-detector which suffers the most from radiation damage, due to its proximity to the interaction point. It must be replaced every 10 years at most, as was done in 2016 for the pixel tracker.

The electromagnetic calorimeter

The ECAL plays a crucial role in the reconstruction of photons, electrons, and in measuring jet energies and MET [201]. It is a homogeneous calorimeter made of lead tungstate ($PbWO_4$) scintillating crystals, covering the pseudorapidity range up to $|\eta| = 3.0$, as illustrated in Figure 2.9. The ECAL is divided into two parts.

- The barrel section (EB) covers the pseudorapidity region $0 < |\eta| < 1.479$, and has an inner radius of 129 cm. The 61,200 crystals are arranged in 36 identical supermodules, and depicted in green in Figure 2.9.

- The two endcaps (EE) cover the pseudorapidity region $1.479 < |\eta| < 3.0$, and are positioned at a distance of 314 cm from the interaction point. Each endcap is split into two

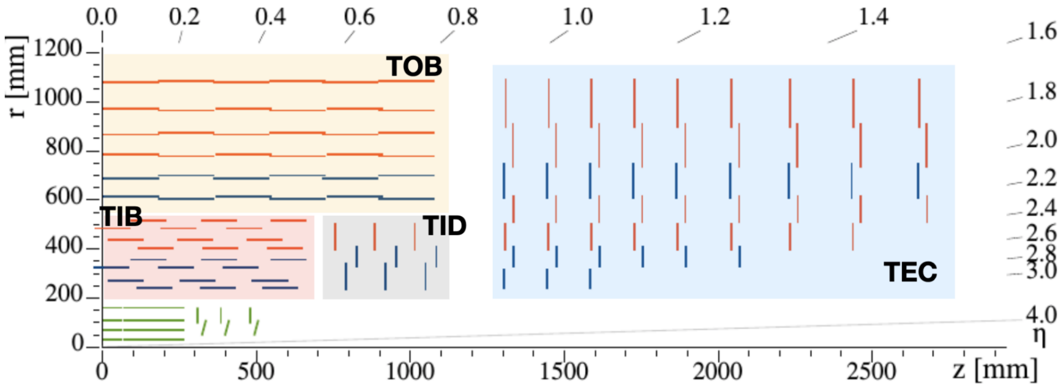


Figure 2.8: Schematic cross-section view of the CMS tracker. The pixels modules are represented in green and the silicon strips are represented in red and blue lines. The TIB, TOB, TID and TEC are represented in red, yellow, gray and blue respectively [178].

halves, or Dees, containing 3,662 crystals. These crystals are organized into 5×5 units, known as supercrystals, and depicted in blue in Figure 2.9. In addition, a preshower detector with a much finer granularity is installed in front of each endcap disk. The preshower is a sampling calorimeter made of lead followed by silicon sensors, which improves the spatial resolution for the pseudorapidity region $1.65 < |\eta| < 2.6$, and depicted in red in Figure 2.9.

When electrons and protons have a momentum exceeding 1 GeV, multiple decays give rise to a shower of new particles with decreasing energies. The main parameters describing this shower are the radiation length X_0 and the Molière radius R_M, representing the longitudinal extension of the shower and the radius of the cylinder containing 90% of the shower energy deposition on average respectively [179]. It is imperative to minimize X_0 and R_M to ensure that the majority of the EM shower's energy is contained within the detector. For $PbWO_4$, $X_0 = 0.89$ cm and $R_M = 2.2$ cm, facilitating the construction of a compact detector with precise energy measurements and spatial resolution. Moreover, $PbWO_4$ emits 80% of the scintillation light within 25 ns, and exhibits resilience to radiations, ensuring efficient signal detection. All these reasons lead to the choice of $PbWO_4$ for the ECAL design.

The most important characteristic of the ECAL is the resolution σ_E, measured as [202]:

$$\frac{\sigma_E}{E} = \frac{0.028}{\sqrt{E}} + \frac{0.12}{E} + 0.003, \tag{2.6}$$

where E is the energy of the EM shower in GeV. The first term accounts for stochastic effects, including statistical fluctuations in the energy measurement. The second term comes from the electronic noise, while the last term relates to signal losses and calibration errors. Unfortunately, the resolution decreases over time, as the ECAL crystals lose transparency due to continuous irradiation during collisions. To sustain the ECAL's excellent resolution performance, a dedicated laser monitoring system has been developed to mitigate these losses [203].

The hadron calorimeter

The HCAL measures the energy of hadronic showers, and is the only sub-detector capable of detecting those coming from neutral hadrons [205]. The HCAL is a heterogeneous calorimeter made from different materials, and covering the pseudorapidity range up to $|\eta| = 5.0$, as illustrated in Figure 2.10. The HCAL is divided into four parts.

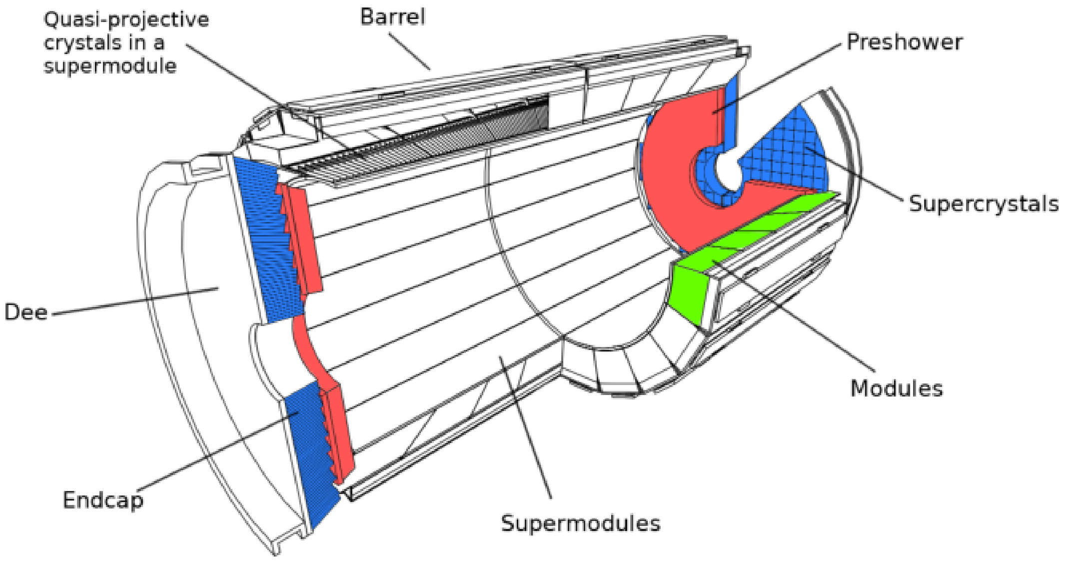


Figure 2.9: Schematic cross-section view of the CMS ECAL. The supermodules, supercrystals and preshower detectors are represented in green, blue and red respectively [204].

- The HCAL Barrel (HB) and HCAL Endcap (HE) cover the respective pseudorapidity regions of $0 < |\eta| < 1.4$ and $1.4 < |\eta| < 3.0$. These detectors are sampling calorimeters consisting of brass layers serving as the absorber, with a length of 16.5 cm, and plastic scintillator as the active material. The modules are arranged into units known as 'towers', each with a fine granularity.
- The HCAL Outer (HO) covers the pseudorapidity region $0 < |\eta| < 1.4$. Positioned outside the magnet solenoid, it acts as a 'tail-catcher' for hadronic showers. The HO calorimeter uses the same active material as before, but the brass is replaced by the 'return yoke' steel as an absorber. The return yoke, which contains the magnetic field outside the solenoid coil, comprises three layers of steel interleaved with four muon detector planes.
- The HCAL Forward (HF) covers the pseudorapidity region $2.9 < |\eta| < 5.2$. It is a Cherenkov calorimeter based on a steel absorber and quartz fibers, able to handle the high particle flux next to the beam pipe.

The combined (ECAL + HCAL) calorimeter energy resolution σ_E is measured as [206]:

$$\frac{\sigma_E}{E} = \frac{0.847}{\sqrt{E}} + 0.074, \tag{2.7}$$

where E is the energy of the EM shower in GeV. The first term accounts for stochastic effects, while the second term relates to signal losses and calibration errors.

The superconductive magnet

The CMS magnet bends the trajectories of the charged particles, allowing us to deduce their electric charge and momentum by measuring the radius of curvature [207]. This magnet is a NbTi superconducting solenoid magnet with a length of 12.5 m, an outer radius of 3.15 m, and a weight of approximately 220 tons. It produces an axial and uniform magnetic field of 3.8 T, through ultrahigh vacuum conditions and a temperature of 4.1 K. To minimize the material obstructing the calorimeters while accommodating the tracker, the magnet is strategically placed between the HCAL and the muon chambers.

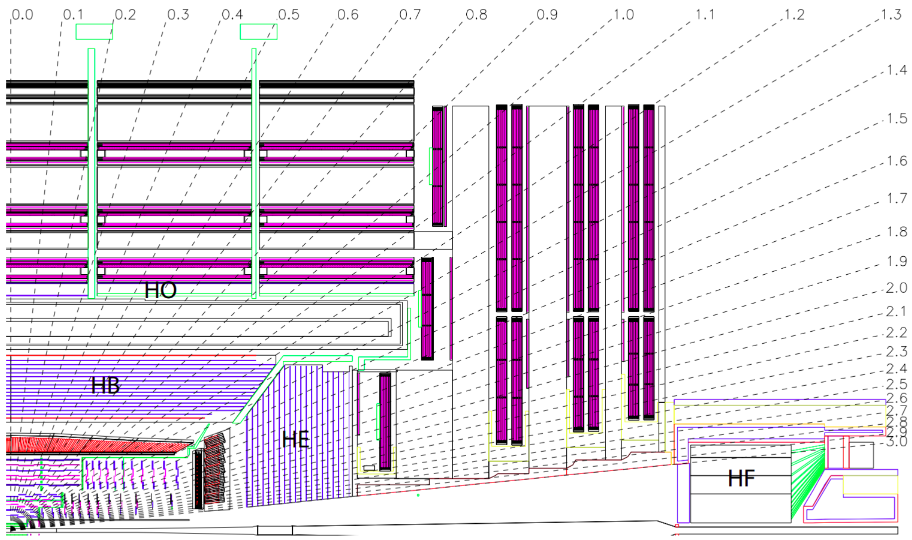


Figure 2.10: Schematic cross-section view of the CMS HCAL [195].

The magnetic field is confined by the return yoke outside the solenoid, which accounts for the majority of the weight of CMS (approximately 10,000 tons). The return yoke serves to filter out residual particles, allowing only neutrinos and muons to be detected in the muon chambers. The magnetic field strength in the muons chambers is 2 T.

The muon chambers

The muon chambers constitute the outermost layers of the CMS detector, since muons have the highest penetration power, excepting neutrinos, and traverse all preceding sub-detectors [208]. The primary goal of the muon chambers is to detect muons, and to measure their momenta. They are made mostly of four kinds of muon stations, as illustrated in Figure 2.11, totaling 1,400 muon chambers. All muon chambers are gaseous detectors. The selection of these gases is based on criteria such as the radiation resilience, the rejection of neutron-induced background, the capacity to operate in high magnetic fields, and the response time.

- The Drift Tubes (DTs) cover the pseudorapidity region $0 < |\eta| < 1.3$. They consist of four layers of DTs filled with a gaseous mixture of argon (Ar) and carbon dioxide (CO_2) [209], providing a response time of 400 ns. They are depicted in pale yellow in Figure 2.11.

- The Cathode Strip Chambers (CSCs) cover the pseudorapidity region $0.9 < |\eta| < 2.4$. They consist of five disks of CSCs filled mostly with Ar and CO_2, for a response time of 30 ns, and are depicted in green in Figure 2.11. These chambers occupy a total area of 5,000 m^2 for a gas volume of 50 m^3.

- The Gas Electron Multiplier (GEM) covers the pseudorapidity region $1.6 < |\eta| < 2.2$, and is depicted in red in Figure 2.11. It was installed next to the CSCs to complement muon detection in this pseudorapidity range.

- The Resistive Plate Chambers (RPCs) cover the pseudorapidity region $0 < |\eta| < 2.1$. They are made of two parallel bakelite plates separated by a few millimeters, and filled with a predominant freon ($C_2H_2F_4$) gas mixture, providing a response time of 25 ns. They are depicted in blue in Figure 2.11.

The data from the muon chambers must finally be integrated with the tracker data to improve the resolution of muon momentum. The muon momentum resolution varies as a function of the pseudorapidity, but the overall resolution ranges from 5-8% at 10 GeV to 20-40% at 1 TeV when relying only on the muon system. However, these figures improve notably from 1-1.5% at 10 GeV to 6-17% at 1 TeV when both the tracker and outer muon systems are combined [209].

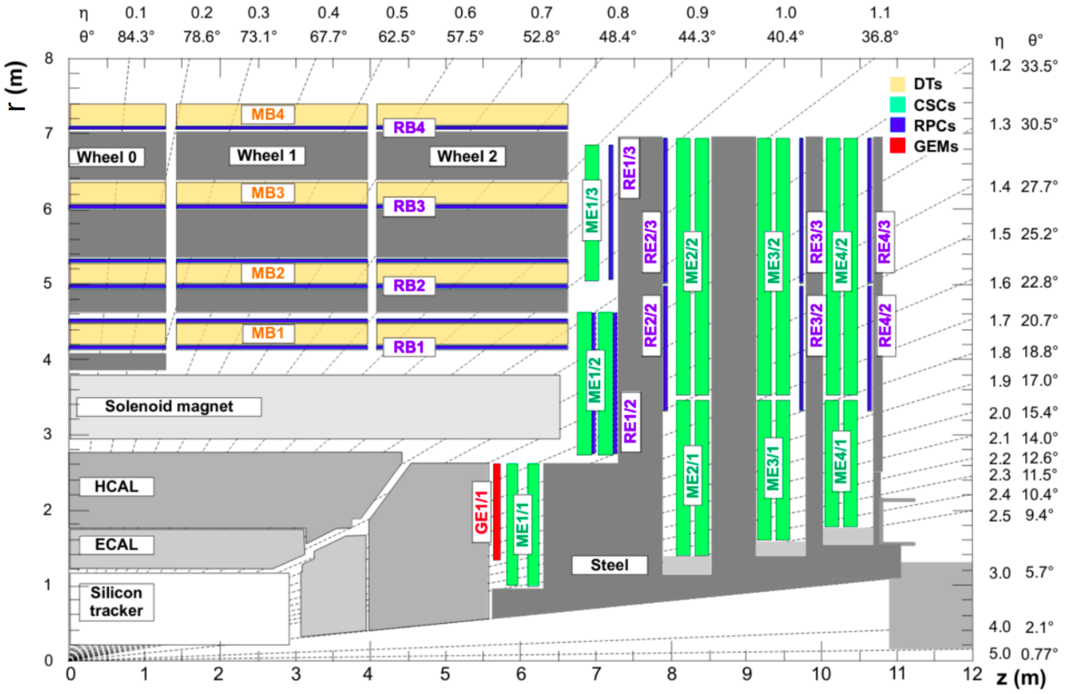


Figure 2.11: Schematic cross-section view of the CMS muon chambers. The DTs, CSC, GEM and RPCs are represented in pale yellow, green, red and blue respectively [210].

2.2.2 Object and event reconstruction in the CMS experiment

Particles resulting from proton-proton collisions leave distinct traces in the CMS sub-detectors as shown in Figure 2.12. With a collision frequency of 40 MHz, employing a dedicated trigger system in CMS is imperative to record the particles of interest. This aspect will be widely explored in Chapter 3 for both the Run 2 and Run 3 periods. To efficiently reconstruct these particles, and thus the overall event, the CMS collaboration has developed a dedicated algorithm called Particle Flow (PF) [211]. This algorithm gathers information from the sub-detectors, combining them to infer the nature of the particles in the event, and reconstruct them to build higher-level objects and quantities.

Particle tracks and clusters

Signals in the tracker above a certain threshold are initially grouped into what is known as 'hits'. These hits provide an estimate of the position and associated uncertainty by summing the charge collected in neighboring sensors, corrected for potential contributions from electronic noise and Lorentz drift effect induced by the magnetic field [178]. Hits are then used for the reconstruction of the tracks themselves, using the Combinatorial Track Finder (CTF) algorithm [212]. The CTF algorithm scans for hits in each tracker layer, and tests the compatibility of the track through a χ^2 test. The track is refined, and hit positions are calculated with high precision, using an additional filter which eliminates the false tracks, corresponding to wrong combinations or poorly reconstructed tracks. The entire procedure is repeated twelve times.

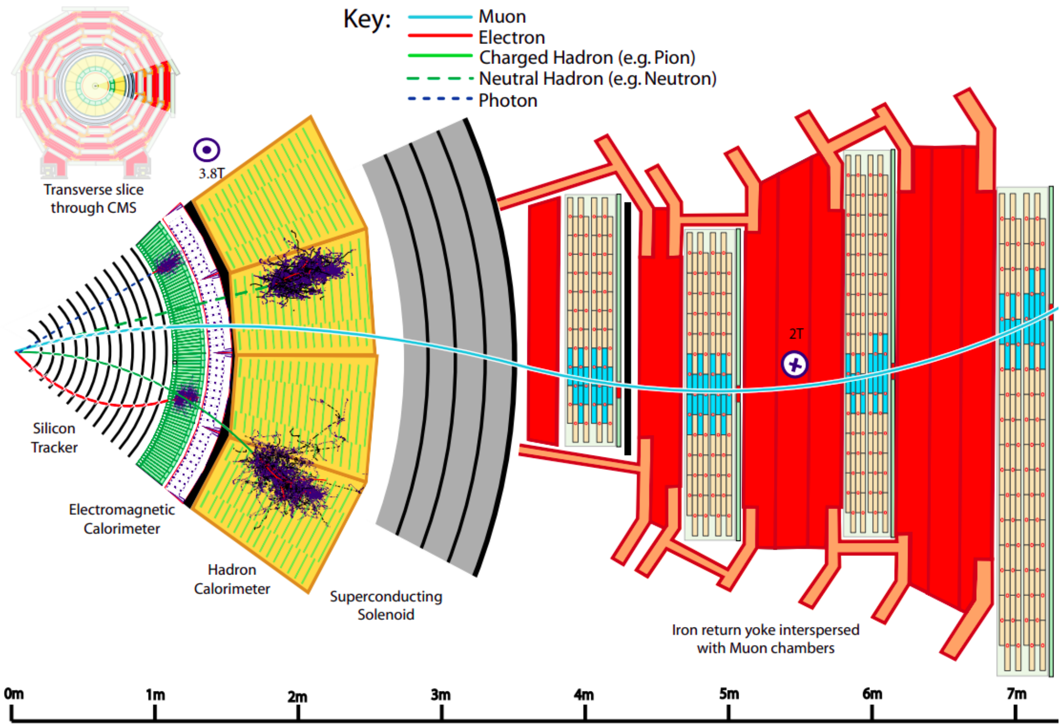


Figure 2.12: Schematic cross-section view of the CMS detector. The muon, electron, hadron and photon signatures are represented as light blue, red, green and dark blue lines respectively, with solid lines for charged particles and dashed lines for neutral particles [211].

In the first iteration, high-p_T tracks originating from the PV are reconstructed. Since the next iterations no longer consider previously used hits, the reduction in combinatorial complexity of hit association allows for relaxed selection criteria, and the reconstruction of less apparent tracks.

In evaluating the reconstruction performance, we define various Figures Of Merit (FOM) estimated from MC simulations (see section 2.2.3). A reconstructed track is associated with a simulated particle if at least 75% of its hits originate from the simulated particle. If a reconstructed track cannot be associated with any simulated particle, it is considered to result from the combination of uncorrelated hits, and is labeled as 'fake'. The number of fake tracks varies significantly depending on the measurement precision and the layout of the tracker modules.

- The track reconstruction efficiency is the fraction of matched reconstructed tracks among all simulated tracks. It exceeds 90% for particles with 2 GeV $< p_T <$ 40 GeV and $|\eta| < 1.0$, and remains relatively stable with various PU levels [213].
- The track fake rate is the fraction of non-matched reconstructed tracks among all reconstructed tracks. It remains below 2% for particles with 0.2 GeV $< p_T <$ 40 GeV, $|\eta| < 1.5$, and PU < 35, but it can rise up to 20% for PU $= 70$, where the number of fake tracks is inevitably higher [213].

To mitigate the impact of fake tracks and maximize reconstruction efficiency, only tracks which pass tight criteria, labeled as 'HighPurity' tracks, are used in most of the CMS analyses [200]. It is worth noting that the track reconstruction process is also applied in the muon chambers to reconstruct muon tracks, in addition of the tracker.

Besides tracks which are relevant only for charged particles, the presence of energy aggregates in calorimeters, known as 'clusters', allows for the reconstruction of both charged and neutral particles. Clusters are defined using local energy maxima, and the energies of neighboring sensors are summed if they exceed a certain threshold [178].

The Particle Flow algorithm

The starting point of the PF framework are the tracks and the clusters. These elements are associated with each other by a geometrical connection, known as a 'link', in the η-φ plane. Based on these distinct signatures in the CMS experiment, the PF framework generates a list of identified particles which can serve for reconstructing higher-level objects such as jets, and calculating the MET. The particles are presented in the sequence of the reconstruction process, with tracks and clusters used first being subsequently discarded.

- The muons are the first particles to be reconstructed. They are identified by the presence of tracks in both the tracker and muon chambers. Combining the information of all subdetectors, the muon reconstruction efficiency exceeds 95% for $p_T < 1.5$ TeV and any value of η, and 90% for $p_T > 1.5$ TeV in the barrel region, for both data and MC [214].

- The electrons are identified by the presence of clusters in the ECAL which are associated with a track. The reconstruction of electron tracks uses a specific tracking algorithm, known as the Gaussian Sum Filter (GSF) algorithm, which accounts for trajectory alterations due to bremsstrahlung photon emissions. The electron reconstruction efficiency exceeds 95% for any value of E_T and η, for both data and MC [215]. The electron energy resolution is within 0.1% in the barrel region and 0.3% in the endcap region for 10 GeV $< E_T <$ 50 GeV, for both data and MC.

- The isolated photons are identified by the presence of clusters in the ECAL which are not associated with any track. Additionally, photons may convert to an electron-positron pair before reaching the ECAL, effectively producing a signature in the tracker as well. Cluster candidates must satisfy $E_T > 10$ GeV, and be isolated from other ECAL clusters, with the main background coming from neutral mesons decaying into two photons. The photon reconstruction efficiency and energy resolution are similar to that of electrons, for both data and MC [215].

- The neutral and charged hadrons are identified by the presence of clusters in both the ECAL and HCAL, which are associated with zero or at least one track respectively. Jets may include hundreds of such particles, and are reconstructed using a dedicated algorithm, known as the anti-k$_T$ (AK) algorithm [216]. This algorithm aggregates particles based on their transverse momentum and the distance d_{ij} defined as:

  $$d_{ij} = \min\left(p_{T_i}^{-2}, p_{T_j}^{-2}\right) \frac{\Delta R(i,j)^2}{\mathcal{R}^2}, \tag{2.8}$$

  with p_{T_k} the transverse momentum of the particle k, $\Delta R(i,j)$ the spatial separation defined in equation 2.5, and $\mathcal{R}$ the radius of the cone inside which particles are associated. The value of the radius may vary depending on the CMS analysis, but $\mathcal{R} = 0.4$ is a common value within the CMS collaboration, defining 'AK4' jets (where AK stands for the AK algorithm). Owing to the pseudorapidity acceptance of the tracker, jets from charged hadrons are reconstructed up to $|\eta| = 2.4$, and jets from neutral hadrons up to $|\eta| = 5.0$. The jet energy resolution ranges from 25% at maximum for $p_T = 20$ GeV to 5% for $p_T = 1$ TeV [217].

- After the matching of all measured tracks and clusters to the reconstructed particles, a portion of the initial parton energy may not be detected in CMS. This MET could arise from 'invisible' particles, like neutrinos, as well as from instrumental imperfections. Therefore, the detector must cover a range as wide as possible of the pseudorapidity to ensure the reconstruction of all particles, and hence facilitate the measurement of MET. The MET is usually defined using the transverse momentum of all reconstructed particles as:

$$\vec{\not{p}_T} = -\sum_i \vec{p_{T_i}}, \tag{2.9}$$

with p_{T_i} the transverse momentum of the particle i. The corrections applied to the MET and all reconstructed objects will be discussed in Chapter 4.

An example of jet reconstruction from its constituents is shown in Figure 2.13. The illustrated jet has a p_T of 65 GeV, and is made of five particles: two charged hadrons π^+ and π^-, two photons coming from a π^0 decay, and one neutral hadron K_L^0. The track T_1 is linked to the ECAL cluster E_1 and to the HCAL clusters H_1 and H_2, while the track T_2 is linked only to the HCAL clusters H_2 and H_1. These elements form two PF blocks: the first one (T_1, E_1, H_1) corresponds to the charged hadron π^-, and the second one (T_2, H_2) to the charged hadron π^+. The remaining ECAL clusters E_2, E_3 and E_4 are not linked to any tracks or clusters, and each forms its own PF block, corresponding to the two photons coming from the pion π^0 and the neutral hadron K_L^0 respectively.

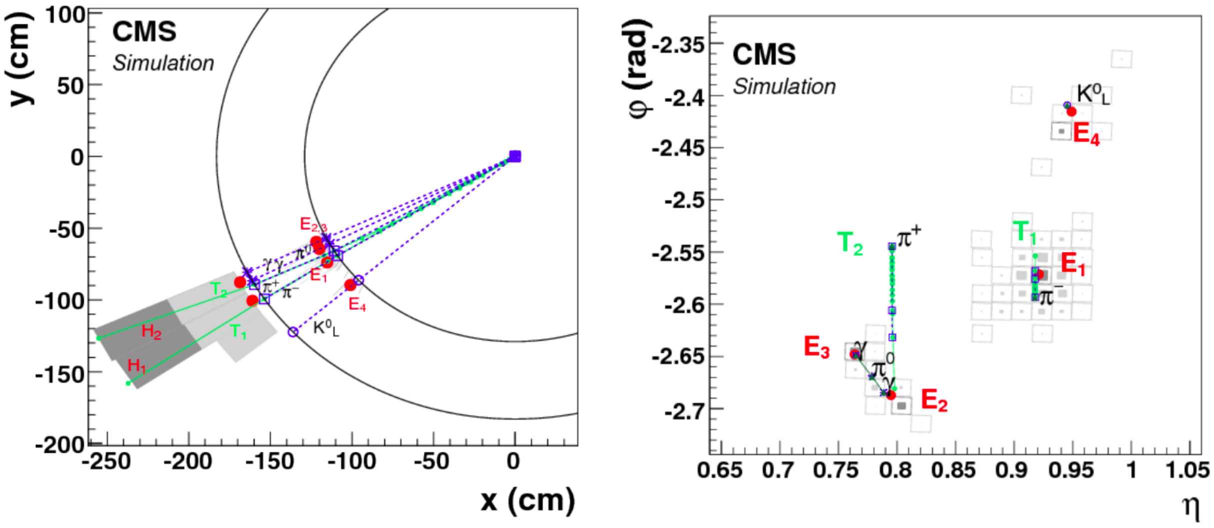


Figure 2.13: Event display of a jet made of five particles in the transverse plane (left) and the η-φ plane on the ECAL surface (right). On the left plot, the ECAL and HCAL detectors are represented as circles centered around the interaction point. The tracks and clusters are represented as green lines and red deposits respectively [211].

2.2.3 Monte-Carlo simulations in the CMS experiment

Numerical simulations conducted using MC methods [218] play a crucial role in high-energy physics to directly compare theoretical predictions with experimental results, by accurately representing the CMS detector. The CMS collaboration ensures the integrity of its simulations by providing MC samples produced centrally, created by adhering to the following procedure [189].

- Step 1: events for a given process are generated using MC programs known as matrix element generators, like Madgraph [219] or Powheg [220, 221], at LO in pQCD or including higher-order corrections. They provide the cross-section of the studied process and the four-momentum vectors of the partons involved in the hard collision, and are formatted in Les HouchEs (LHE) format [222].

- Step 2: outputs from the matrix elements generators flow into parton-shower event generators like Pythia [223]. They calculate the four-momenta of spectator partons involved in soft collisions, and proceed to the fragmentation and hadronization for both hard and soft partons. They can also generate the Initial State Radiation (ISR) and Final State Radiation (FSR) of photons and gluons. The decay of unstable particles is performed until only stable decay products remain in the final state. This GENeration (GEN) step is fully interfaced with the CMS SoftWare (CMSSW) [224]. A ROOT file [225] is created, listing the four-momentum vectors of all initial, intermediary, and final particles of the event.
- Step 3: the GEN file flows into the GEANT4 program [226] to simulate the interaction of the particles with the material of the CMS detector. Furnishing the geometry and features of the detector, GEANT4 can simulate a wide range of particle-matter interaction processes. This SIMulation (SIM) step creates a ROOT file containing all the simulated energy deposits in the detector, known as simhits.
- Step 4: the SIM file flows into CMSSW to digitize the energy deposits in all sub-detectors, and simulate the electronic response calibrated from beam test data. This RAW step provides preliminary raw data.
- Step 5: the electronic noise is simulated through a dispersion function, and the signal response to these fluctuations is digitized. This is the DIGItized (DIGI) step.
- Step 6: the DIGI file flows into CMSSW to simulate the trigger response and the object reconstruction process, as discussed in section 2.2.2. This RECOnstruction (RECO) step creates ROOT files which can be directly used for CMS analyses.

It is important to note that this procedure may be repeated multiples times, following several data acquisition eras. Regarding potential imperfections which may arise within the CMS detector, simulated events need reprocessing to accommodate any calibration discrepancies. The first MC campaign is typically referred to as the End Of Year (EOY) campaign, succeeded by new campaigns which profit from the latest calibrations of the CMS detector for the data era under investigation.

2.2.4 HL-LHC perspectives of the CMS detector

The main objective of the HL-LHC upgrade is to significantly increase the volume of acquired data to facilitate new physics searches, Higgs boson coupling measurements, and other precision tests of SM predictions. The aim is to reach an instantaneous luminosity L = 5-8 $\times$ 10^{34} cm^{-2}.s^{-1} and an integrated luminosity L_{int} = 3,000 fb^{-1} over a 10-year period, representing a 15 times larger luminosity compared to the already acquired luminosity. The HL-LHC will introduce unprecedented challenges for the sub-detectors due to high radiation levels, especially along the beam axis as illustrated in Figure 2.14. The impact of the PU will be even more dominant, with an expected number of 200 collisions per beam by the end of this data-taking period. Consequently, the CMS sub-detectors must undergo major upgrades, known as the phase-2 upgrades, to maintain the excellent performance of the CMS detector in terms of efficiency, resolution, and background rejection within a high-luminosity environment. The installation of these upgraded systems started during LS2 (phase-1 upgrades), and will be completed during LS3. With the goal of predicting the impact of the harsh operating conditions of the HL-LHC on the sub-detectors, and to outline the required upgrades, CMS has undertaken extensive simulations efforts.

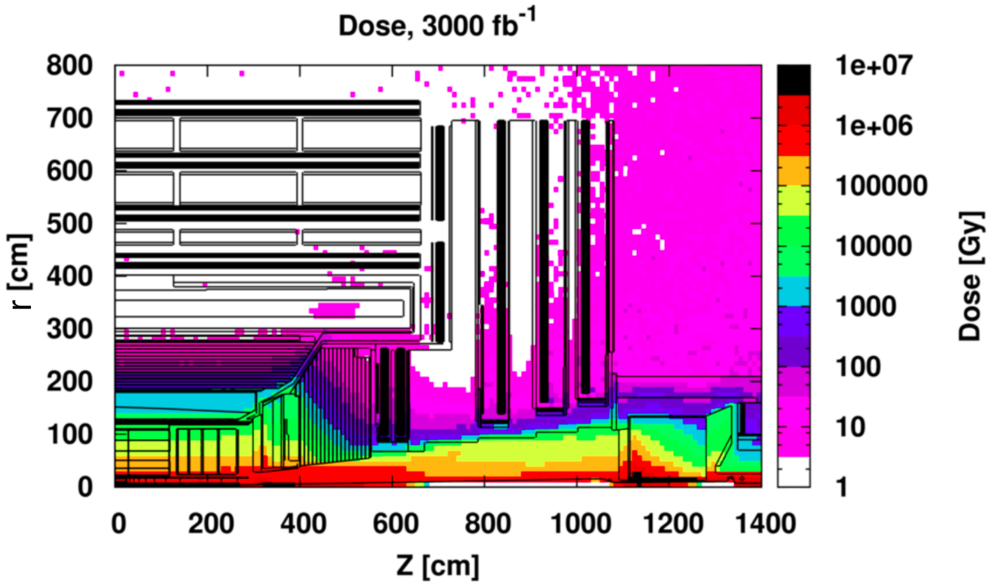


Figure 2.14: Absorbed dose of radiation in the CMS cavern with an integrated luminosity $L_{int} = 3{,}000\ \mathrm{fb}^{-1}$ [227].

The phase-2 tracker

Given its proximity to the interaction point, the tracker is anticipated to endure significant radiation damage, necessitating a complete replacement as part of the phase-2 upgrades [228]. The upgraded tracker will be composed of two detectors which will extend into the forward region, and will be able to reconstruct tracks up to approximately $|\eta| = 4.0$.

- The Inner Tracker (IT) covers the pseudorapidity region $0 < |\eta| < 4.0$. It comprises silicon pixel modules, depicted as green and orange lines in Figure 2.15.

- The Outer Tracker (OT) covers the pseudorapidity region $0 < |\eta| < 2.4$. It is made from new silicon modules denoted as p_T-modules, depicted as blue and red lines in Figure 2.15, and which consist of two single-sided sensors. The sensors are designed to produce correlated signals, ensuring that only the hit pairs, or 'stubs', compatible with particles above a p_T threshold of 2 GeV are retained, as illustrated in Figure 2.16. This information will be directly integrated into the first stage of the trigger system, or Level 1 (L1) (see section 3.1.1), to significantly reduce the data volume resulting from collision events. The modules aligned parallel to the z-axis or to the transverse plane consist of two strip sensors (2S modules), or a combination of one strip and one pixel sensors (PS modules) respectively.

The upgraded tracker must provide a very high resolution to accurately associate the produced tracks with their respective vertices, a critical requirement for managing the increased PU levels. To maintain the track reconstruction performance under these new conditions, the granularity of the IT will be increased by a factor of 6, relative to the current detector.

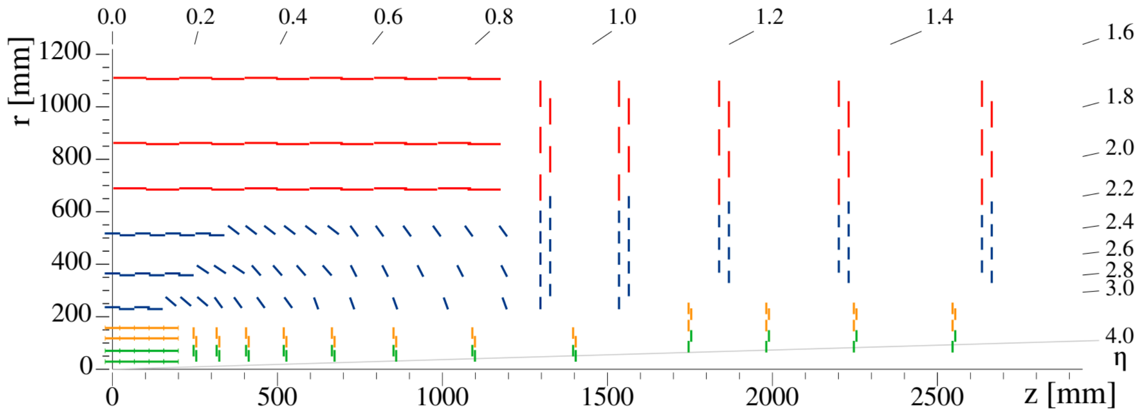


Figure 2.15: Schematic cross-section view of the CMS upgrade tracker. The silicon modules are represented in green and orange and the silicon pixel modules are represented in red and blue [228].

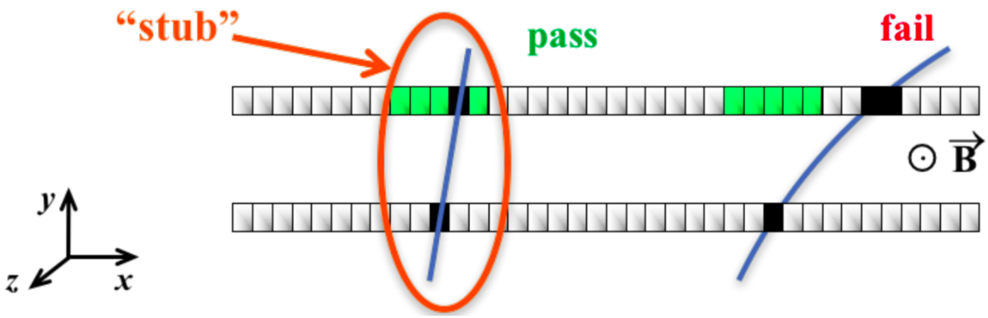


Figure 2.16: Illustration of the p_T-module concept. The channels depicted as green squares represent the selection window to define an accepted stub. Low-p_T particles will be rejected [228].

The Minimum ionizing particles Timing Detector

A new detector, known as the Minimum ionizing particles Timing Detector (MTD), will be added between the tracker and the calorimeters to cover a pseudorapidity region $0 < |\eta| < 3.0$. The MTD's primary function is to contribute to the measurement of the precise timing of the collision vertices, by associating tracks from a vertex with both the hits and their respective timing within the detector. This process enables the identification and exclusion of external tracks, which align approximately with the vertex but arrive at incorrect times, thereby disregarding their contribution to that specific collision. As its name suggests, the MTD is designed to achieve precise time of arrival measurements and even for Minimum Ionizing Particles (MIPs), which reconstruction is challenging due to their minimal energy loss in the material. Initially, the target timing resolution for the start of the HL-LHC is set at 30-40 ps, gradually deteriorating to 50-60 ps by the end of operations due to the radiation-induced damage [179]. The MTD is divided into two parts, as illustrated in Figure 2.17.

- The Barrel Timing Layer (BTL) covers the pseudorapidity region $0 < |\eta| < 1.48$. It is a cylindrical detector made of scintillating crystals bars in LYSO(Ce), with a length of 5.7 cm in the transverse plane and a width of 3 mm along the z-axis. It is depicted in light purple in Figure 2.17. Each crystal is coupled with two Silicon PhotoMultipliers (SiPMs), performing the readout from both ends of the bar. This configuration was chosen to enhance resolution by mitigating light travel delays within the crystal.

- The Endcap Timing Layer (ETL) covers the pseudorapidity region $1.48 < |\eta| < 3.0$. It is a two-disk system made of silicon devices called LowGain Avalanche Detectors (LGADs), selected for their superior radiation resilience relative to SiPMs. It is depicted in red in Figure 2.17.

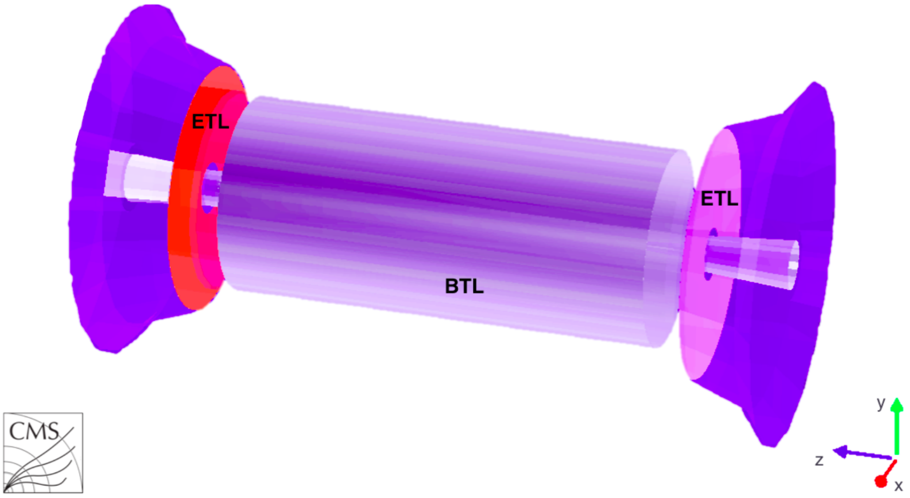


Figure 2.17: A schematic layout of the MIP Timing Detector design in GEANT4. The BTL and ETL are represented in light purple and red respectively [229].

The phase-2 calorimeters

The upgrade plan entails the replacement of both the front-end electronics for the EB and the off-detector electronics for the EB and HB [230]. This revision will enhance timing precision by reducing the current resolution of approximately 120 ps to below 50 ps, and mitigate noise issues stemming from the photodetectors.

Extensive radiation-induced degradation has been observed in both the EE and HO during ongoing operations [231], necessitating the construction of a new detector, known as the High-Granularity CALorimeter (HGCAL), and illustrated in Figure 2.18. It will cover the pseudorapidity region $1.5 < |\eta| < 3.0$, and promises excellent transverse and longitudinal segmentation while measuring the timing. It is divided into two parts.

- The electromagnetic part (CE-E) of the HGCAL consists of 26 layers of a copper-tungsten-lead alloy (Cu/CuW/Pb) serving as absorbers, interleaved with silicon sensors as the active material. They are depicted as blue lines in Figure 2.18.
- The hadronic part (CE-H) of the HGCAL has 21 layers of steel absorbers, interleaved with silicon sensors in high-radiation regions, and depicted as green lines in Figure 2.18. They are complemented by scintillating tiles in the outer regions as the active material, depicted as purple lines in Figure 2.18.

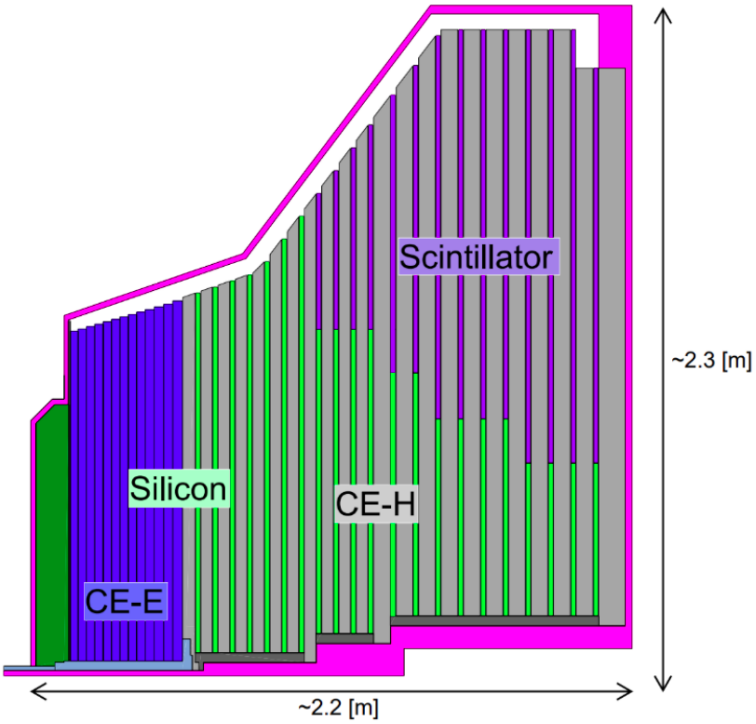


Figure 2.18: Schematic illustration of the HGCAL. The CE-E is represented in blue and the CE-H is represented in green and purple [232].

Timing within object reconstruction

Following the tracker and calorimeter upgrades, the phase-2 upgrade of CMS will integrate timing measurements directly into the reconstruction process. This represents a significant improvement to discriminate the PV from the vertices coming from the PU. Indeed, collisions occur within a time interval characterized by a Root Mean Square (RMS) extent of 180-200 ps for a 25 ns bunch crossing and a PU below 200, as illustrated in Figure 2.19. By incorporating timing as the fourth dimension into vertex reconstruction, the misreconstruction efficiency decreases from 15% in spatial reconstruction to 1% in space-time reconstruction.

The overview presented here highlights only some of the key ingredients of the phase-2 upgrade. The muon chambers will undergo a moderate upgrade, but the overall structure of the sub-detectors will remain unaltered [233]. Numerous other upgrades including those of the detector electronics, trigger system, data acquisition, and luminosity measurements are currently under construction.

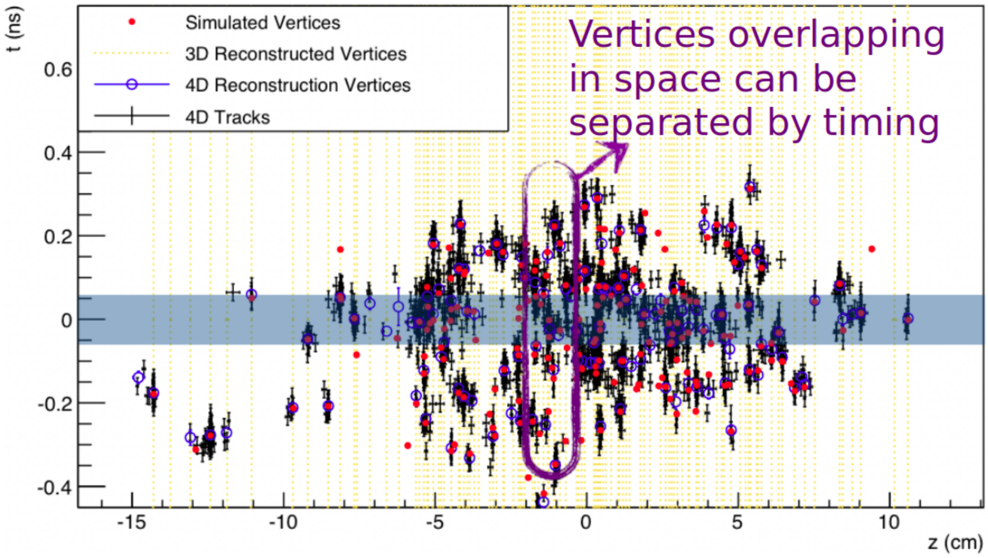


Figure 2.19: Simulated and reconstructed vertices within a 25 ns bunch crossing under a PU of 200, assuming a MTD detector with a timing resolution of approximately 30 ps. The inclusion of the timing measurement into the reconstruction process improves significantly the vertex reconstruction performance [229].

Chapter 3

Trigger performances for the Run 2 and Run 3 periods

Under nominal operating conditions, proton bunch collisions occur every 25 ns within the CMS detector, resulting in a collision frequency of 40 MHz. However, the current computing capabilities can only process a limited number of collisions, corresponding to a frequency of approximately 1 kHz. Consequently, the events must undergo a preselection before being used for physics analyses: this constitutes the primary goal of the trigger system. This preselection requires the reconstruction of all particles to be within a few microseconds time frame. Given this extremely brief interval, achieving a full event reconstruction is impractical, necessitating a focus on simplified particle signatures within the detector. To accommodate the targeted final states of the analyses within the CMS collaboration, hundreds of selection criteria, known as 'trigger paths', are optimized to efficiently capture the events of interest. The selection performed at the trigger level, or 'online' selection, is crucial in all CMS analyses, as any miscalibration of the trigger paths could introduce biases in the subsequent analysis-level selection, or 'offline' selection.

This chapter will be divided into two sections. The first section will focus on the properties of the trigger system in the CMS experiment. The trigger system is divided into two parts, the Level 1 (L1) trigger and the High Level Trigger (HLT), both contributing to reduce the acquisition frequency to approximately 1 kHz. The second part will concentrate on the HLT performance within a specific analysis, targeting the search for a VLQ T' decaying into a Z or Higgs boson, accompanied with a top quark in the all-hadronic final state [165] (see Table 1.3). After a brief description of the targeted final state and the event selection, the different algorithms used to reconstruct the particles at both the online and offline levels will be presented. Following a presentation of the algorithm performances during the Run 2 period, I have conducted several studies to improve these performances for the Run 3 period. The comparison with early data of the Run 3 period will close the chapter.

3.1 The trigger system in the CMS experiment

3.1.1 The trigger structure

The CMS trigger system consists of two levels [234]: the L1 trigger and the HLT.

The Level 1 trigger

The L1 is a hardware trigger, which uses information from the CMS sub-detectors to determine the selection or rejection of events. It uses information from the ECAL, the HCAL, and the muon chambers. Each event is buffered during 3.8 μs. During that time, information from the sub-detectors electronics reach the CMS processors, where it is processed for approximately 1 μs, before giving back the decision to the electronics to proceed the event at the HLT or not. The L1 algorithms, known as the Trigger Primitives (TPs), are based on the track segments, energy deposits, and hits (see section 2.2.2). For example, an event passes the L1 selection if the energy deposits exceed a certain threshold. It exists 506 algorithms at the L1 for the Run 3 period [178], each of them targeting a different final state. Events which successfully pass the L1 trigger are forwarded to the Data AcQuisition system (DAQ), where information from the 16 million readout channels in the CMS sub-detectors is combined to form a single event [179]. At this stage, the major limitation comes from the readout channels, and the rate of events passing the L1 criteria is decreased from 40 MHz to 100 kHz.

The L1 trigger relies not only on sub-detectors information, but also on additional data from sensors positioned ahead of and behind the CMS detector [195]. The Beam Pick-up Timing Experiment (BPTE) and the Beam Scintillator Counter (BSC), located 175 meters in front and 10.5 meters behind the interaction point respectively, detect the bunch crossing and the collisions between the beam and residual gas particles, despite the ultrahigh vacuum conditions. This additional information is known as the 'technical' trigger, which selects events known as the 'minimum bias' events, and correspond mostly to elastic and diffractive collisions. It usually happens at the beginning of each data-taking period, where the number of bunches per beam is low and the time interval between two collisions greater than 25 ns. Otherwise, under nominal operating conditions, the selected events are known as the 'zero bias' events.

The High Level Trigger

The HLT is a software-based trigger, which uses information from all the CMS sub-detectors. Each event passing the L1 trigger is buffered during 350 ms, reconstructed thanks to the CMS processors, and processed with more complex algorithms than those used at the L1 trigger. These algorithms are selected to target specific analysis objects (see section 3.1.2), and collectively forming a path. All the HLT paths constitute the HLT menu. In principle, each path represents a simplified version of the offline selection for a given analysis. However, the event rate at the HLT is constrained to approximately 1 kHz, leading to many analyses with similar event selections benefiting from the same HLT path. It exists 712 HLT paths for the Run 3 period.

Data monitoring at the trigger level

To guarantee that no issues arise in any of the CMS sub-detectors during the LHC collisions, the data quality is monitored continuously by shifters, which control every sub-detector online. This holds particularly true at the trigger level, where events are selected before being permanently stored in data centers, and distributed worldwide for all CMS analyses. The data monitoring process is similar at both the L1 and HLT stages, where the event rate must not exceed 100 kHz and 1 kHz respectively. If the event rate exceeds these thresholds, and no issues are reported by the other CMS sub-detector shifters, it indicates that the L1 or HLT paths are not sufficiently selective, and saturate the CMS processors at the HLT level. One solution is to implement a prescale, which involves selecting only a fraction of the events which pass the problematic HLT path, thereby reducing the total event rate. For example, if only one out of ten events is selected, the prescale for that HLT path would be set to 10. Ideally, high-p_T paths should maintain a prescale of 1 throughout the data-taking period. If this is not the case, a prescale is applied on the fly to keep the bandwidth at a reasonable level, and the corresponding HLT path must be redesigned promptly. It is important to note that the prescale is necessary for paths targeting low-p_T objects (such as those in B-physics analyses), which select a large number of events.

Another aspect of trigger data monitoring is the Dead Time (DT), which refers to the percentage of time during which the sub-detector electronics do not process collision events. Under nominal operating conditions, the DT is approximately 4% for the Run 2 period and 6% for the Run 3 period. However, a significant increase in DT could indicate a malfunction in a sub-detector, leading to a reduction in the event rate.

After passing both the L1 and the HLT selection stages, the events are cleaned of any information from noisy channels in the sub-detectors. They are then buffered one last time for data quality evaluation, known as Data Quality Monitoring (DQM). Each data-taking sequence, known as a 'run', is controlled per sub-detector to detect any anomaly in the energy density, timing... If no issues are detected, the run is certified as valid for physics analyses, and the events are proceeded to the final reconstruction for the offline selection. The synchronization of both the trigger and DQM monitoring is crucial to ensure the highest quality of data collection.

3.1.2 Event selection at the High Level Trigger

Every HLT path consists of a series of algorithms, each designed to target a specific object (electron, muon, jet...). Due to time constraints, the object reconstruction cannot be as efficient as in the offline selection, necessitating a comparison between multiple algorithms to determine the most effective at the trigger level.

General description of a High Level Trigger path

The Figure 3.1 presents the path 'HLT_AK8PFJet140_v24' as a general illustration of an HLT path structure. The selection requires at least one AK8 jet, which corresponds to a jet defined in a cone of a radius parameter $\mathcal{R} = 0.8$ with the AK algorithm, reconstructed using the PF algorithm (see section 2.2.2), and with a transverse momentum $p_T > 140$ GeV. Each HLT path is composed of five sequences.

- The first sequence ('HLTBeginSequence' sequence) sets up the L1 and HLT environments.
- The second sequence ('hltL1SingleJt90' filter) lists all the events which pass the different L1 trigger. This includes only the L1 paths which are relevant for the HLT path, based on the targeted objects.
- The third sequence ('hltPreAK8PFJet140' filter) applies the HLT prescale.
- The fourth sequence (from 'HLTAK8CaloJetsSequence' sequence to 'hltSinglePFJet140AK8' filter) reconstructs the targeted objects, and filters the events. The jets are first reconstructed as AK8 jets using only calorimeter information, known as Calo jets, and must have $p_T > 110$ GeV. Then, they are reconstructed as AK8 jets using the PF algorithm, matched to the Calo jets, and must have $p_T > 140$ GeV.

- The fifth sequence ('HLTEndSequence' sequence) exits the HLT environment, returning a boolean indicating whether the event passes the HLT path.

The design of an effective HLT path is thus focused on optimizing the reconstruction algorithms and filters within the fourth block (see section 3.2.2).

HLT_AK8PFJet140_v24
  HLTBeginSequence
  hltL1sSingleJet90
  hltPreAK8PFJet140
  HLTAK8CaloJetsSequence
  hltSingleCaloJet110AK8
  HLTAK8PFJetsSequence
  hltPFJetsCorrectedMatchedToCaloJets110AK8
  hltSinglePFJet140AK8
  HLTEndSequence

Figure 3.1: General structure of the HLT path HLT_AK8PFJet140_v24 obtained with the ConfDB database [235].

Evaluation of High Level Trigger performance

The evaluation of trigger performance at the HLT is based on three FOM.

- The efficiency is the ratio of the number of events passing the offline selection, the reference HLT path, and the target HLT path, to the number of events passing the offline selection and the reference HLT path as follows:

$$\text{Efficiency} = \frac{\text{Selection}_{\text{offline}} + \text{HLT}_{\text{reference}} + \text{HLT}_{\text{target}}}{\text{Selection}_{\text{offline}} + \text{HLT}_{\text{reference}}}. \tag{3.1}$$

  The reference HLT path is orthogonal to the target HLT path. For example, a final state which identifies 5 jets will trigger an HLT path which requires at least 5 jets. A good reference HLT path would then require at least one muon. This ensures there is no bias in calculating the efficiency since data samples trigger HLT paths before applying the offline selection, which is not the case for simulated samples. Since the HLT selection is a simplified version of the offline selection, it is hoped that efficiency is close to 100%. However, discrepancies between algorithms used at both the trigger and analysis levels may reduce the efficiency. If the offline algorithms are well defined, any reduction in efficiency could indicate a potential issue at the trigger level. The efficiency is specific to each analysis, and must be the first variable calculated to ensure the proper functioning of the HLT path. The efficiency is usually measured in bins of a distribution for each targeted object. In the ideal case, the efficiency is 100% when the value of the targeted object exceeds the HLT path threshold, and 0% otherwise. Since the reconstruction is different at both the trigger and analysis levels, this step function can nevertheless transform into a turn-on curve, where the efficiency reaches a 100% plateau for values beyond the HLT path threshold. The selection criteria used in the offline selection must reach this plateau to be considered acceptable.

- The pure rate is the difference of the total HLT menu rate with and without the contribution of the HLT path, where the total HLT menu rate is defined as the ratio of the number of events passing the HLT menu to the total processing time. Since the total event rate is constrained to 1 kHz, it is essential to verify that the HLT path under study does not saturate the processors while collecting events. Otherwise, a prescale must be applied, reducing the sensitivity of the HLT path. But a single event can be selected by two different HLT paths (with one path being a 'shadow' path in this case), making the pure rate the preferred measurement over the total HLT menu rate since it is specific to each HLT path. However, it is important to note that the pure rate of an HLT path can be significantly high, as it may benefit from multiple analyses simultaneously.
- The pure timing is the difference of the timing with and without the contribution of the HLT path, where the timing is defined as the duration required to process each event in the HLT menu. The timing can range from 0 to 2,000 ms per event, but the average timing must not exceed 500 ms per event. Again, the pure timing is preferred since it is specific for each HLT path. The pure timing must not exceed 10 ms per event.

For an HLT path to be considered valid in CMS analyses, the efficiency, the pure rate, and the pure timing should all meet the required standards simultaneously, ensuring both physical and technical feasibility.

3.2 Trigger performances for the Vector-Like Quark T' $\rightarrow$ t+H/Z analysis in the all-hadronic final state

The search for a VLQ T' decaying into a Z or Higgs boson, accompanied with a top quark in the all-hadronic final state has been published for the full Run 2 period [165], and is ongoing for the Run 3 period. A dedicated HLT path has been developed to target the final state of this specific decay. We will explain the design of the HLT path for the Run 3 period, focusing on the study of the choice of the jet reconstruction and b-tagging algorithms.

3.2.1 Presentation of the final state

A typical Feynman diagram of the studied T' decay is shown in Figure 3.2. The VLQ T' corresponds to the VLQ T described in section 1.2.2. It is single-produced via EW interaction, as illustrated on the left in Figure 1.6. The interaction between the radiation of a W or Z boson by a quark q produces an antiquark $\overline{q'}$, and a gluon splitting produces a pair of b-quark-antiquark $b\bar{b}$, leading to the production of the T'. As mentioned in section 1.2.2, the T' has three decay modes: T' $\rightarrow Wb$ (assumed BR = 50%), T' $\rightarrow Ht$ (assumed BR = 25%), and T' $\rightarrow Zt$ (assumed BR = 25%). We will focus on the second and third decay modes here. The top quark decays into one W boson and a b-quark in almost 100% of cases, and the W boson subsequently decays into a pair of quark-antiquark $q''\overline{q'''}$ in 67.4% of cases [8]. The Higgs or Z boson decays into a pair of b-quark-antiquark $b\bar{b}$ in 53% and 15.1% of cases respectively. The event selection requires 5 jets, including 3 jets originating from a b-quark, and known as b-jets. The targeted nominal mass of the T' ranges from 600 GeV to 1200 GeV. In that case, the T' does not produce jets with sufficient energy to be merged into larger jets, hence the search in the all-hadronic 'resolved' final state.

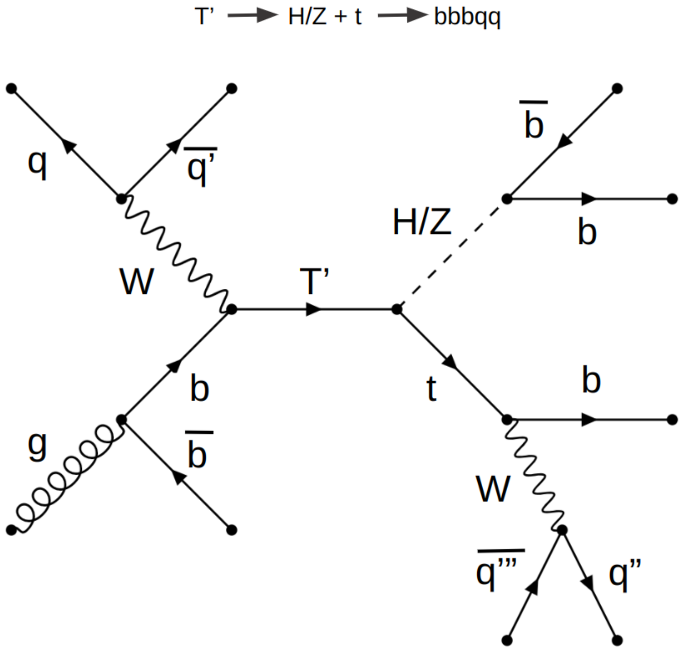


Figure 3.2: Example Feynman diagram for the single production of the VLQ T' and its decay into the all-hadronic resolved final state via a Z or Higgs boson accompanied with a top quark.

3.2.2 Jet algorithms at the High Level Trigger and analysis levels

With an event selection which only requires jets, the choice of the jet algorithms at both the trigger and analysis levels can impact profoundly the analysis performance, particularly for distinguishing the b-jets from the other jets.

Jet reconstruction algorithms at the trigger level: Calorimeter jets vs Particle Flow jets

As mentioned in section 3.1.2, we define two main types of jets at the trigger level: the Calo jets and PF jets. The PF jets have a better reconstruction efficiency compared to the Calo jets, as they use tracking information in addition to calorimeter information. The timing of the algorithm can be reduced while preserving the efficiency by analyzing the hits in areas around the specified track, known as Regions Of Interest (ROIs). Depending on whether information from all the sub-detectors regions or from the ROIs only is used, the jets are classified as 'global' or ROI jets respectively.

Pileup mitigation algorithms in jet reconstruction at the analysis level: CHS vs PUPPI

The objects are reconstructed at the trigger level thanks to the PF algorithm. At the analysis level, the PU is mitigated by applying either the Charge Hadron Subtraction (CHS) algorithm or the PileUp Per Particle Identification (PUPPI) algorithm [236], on top of the PF algorithm.

Both algorithms use vertex information reconstructed from charged particle tracks, and the AK algorithm.

The CHS algorithm is applied after the reconstruction of the PF particles and before the PF jet clustering process [237]. Using tracking information, this algorithm removes charged particles associated with vertices which are not defined as the PV or the SVs, corresponding to the PU vertex candidates. The charged particles which are not associated with any PU vertex, and the neutral particles, are kept.

The PUPPI algorithm is applied within the reconstruction of PF particles [238]. Using local particle distribution and tracking information, this algorithm calculates a weight in a range from 0 to 1 for each particle, whether it is a charged or a neutral particle.

- For a charged particle, a weight of 1 is assigned if the particle is associated only with vertices in a distance along the z-axis with respect to the PV d_z smaller than 0.3 cm, as illustrated in Figure 3.3. Otherwise, a weight of 0 is assigned, and the charged candidate is rejected. Charged particles are considered in the whole pseudorapidity range of the CMS detector.
- For a neutral particle i, a more complex weight α_i is defined as follows:

  $$\alpha_i = \log \sum_{j \neq i, \Delta R(i,j) < 0.4} \left(\frac{p_{T,j}}{\Delta R(i,j)} \right)^2 , \tag{3.2}$$

  where j are all the particles in the cone of the particle i within a radius $\mathcal{R} < 0.4$, and $\Delta R(i,j)$ defined in equation 2.5. Neutral particles are here considered in the pseudorapidity range $|\eta| > 2.5$, where tracking information is not available.

The weights are used to rescale the four-momentum of every particle and redefine the transverse momentum of the jet, thereby reducing the impact of PU contamination.

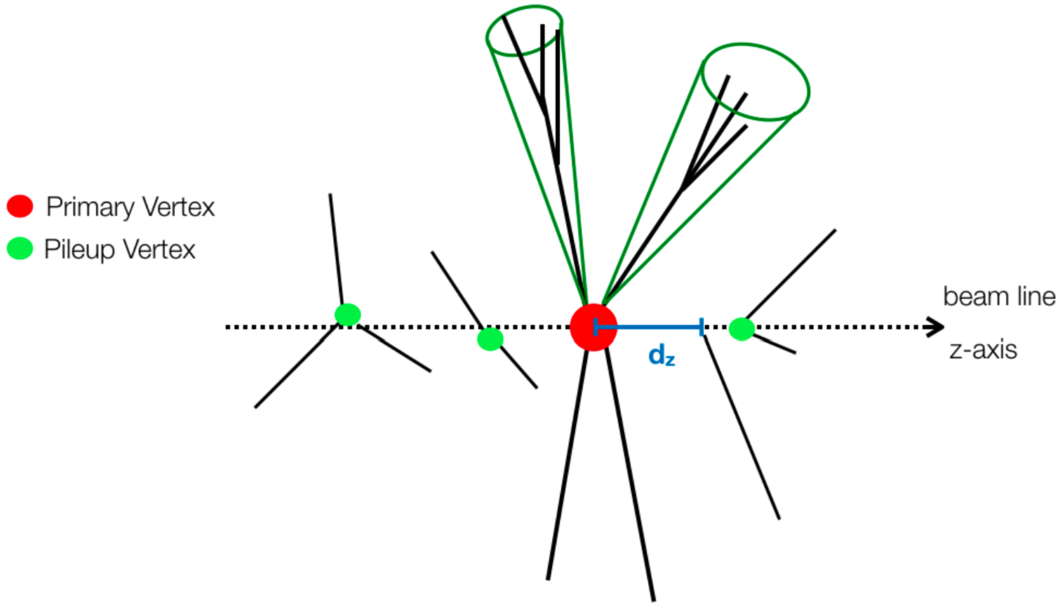


Figure 3.3: Schematic representation of the PV and the PU vertices for the PUPPI algorithm [239].

To compare both the CHS and PUPPI algorithms, we will use an additional algorithm known as the PU jet IDentification (ID) algorithm [240]. The PU jet ID algorithm aims at rejecting jets mostly composed of particles originating from PU interaction. To define the loose, medium, and tight Working Points (WPs), the PU jet ID criteria are designed to select 95%, 90%, and 80% of the jet collection respectively. A comparison of the performances of the CHS jets alone, the CHS jets in association with the PU jet ID algorithm, and the PUPPI jets alone is presented in Figure 3.4, with Run 2 $Z + jets$ simulated samples. The efficiency and purity are presented as a function of the PU, and computed in the pseudorapidity range $0 < |\eta| < 2.5$. The efficiency is defined as the fraction of jets with $p_T > 30$ GeV at the particle level, matched with a AK4 jet with $p_T > 20$ GeV at the reconstruction level. The purity is defined as the fraction of jets with $p_T > 30$ GeV at the reconstruction level, matched with a AK4 jet coming from the PV at the particle level with $p_T > 20$ GeV. For CHS jets, the efficiency exceeds 95% in the whole CMS detector, regardless of PU. The tight PU jet ID applied to the CHS jets reduces this efficiency to 80% for PU exceeding 50, a scenario observed during both the Run 2 and Run 3 periods [236], but at the same time improves the purity to 95% at maximum. For PUPPI jets, the efficiency is better than that of the CHS jets with the tight PU jet ID applied, regardless of PU. Furthermore, the purity of PUPPI jets is better than that of CHS jets with the tight PU jet ID applied, close to 100%, regardless of PU. This is expected as the PUPPI algorithm accounts for neutral particles, unlike the CHS algorithm. With this enhanced performance, the PUPPI algorithm has become the recommended algorithm for PU mitigation for the Run 3 period, whereas the CHS algorithm was the recommended choice during the Run 2 period. Finally, it is important to remember that the CHS and PUPPI algorithms are applied only in the offline selection due to time constraints at the trigger level. With overall better performances of these two algorithms compared to those of PF jets alone, the offline selection is tighter than the HLT selection, resulting in an expected efficiency of the HLT path of 100%.

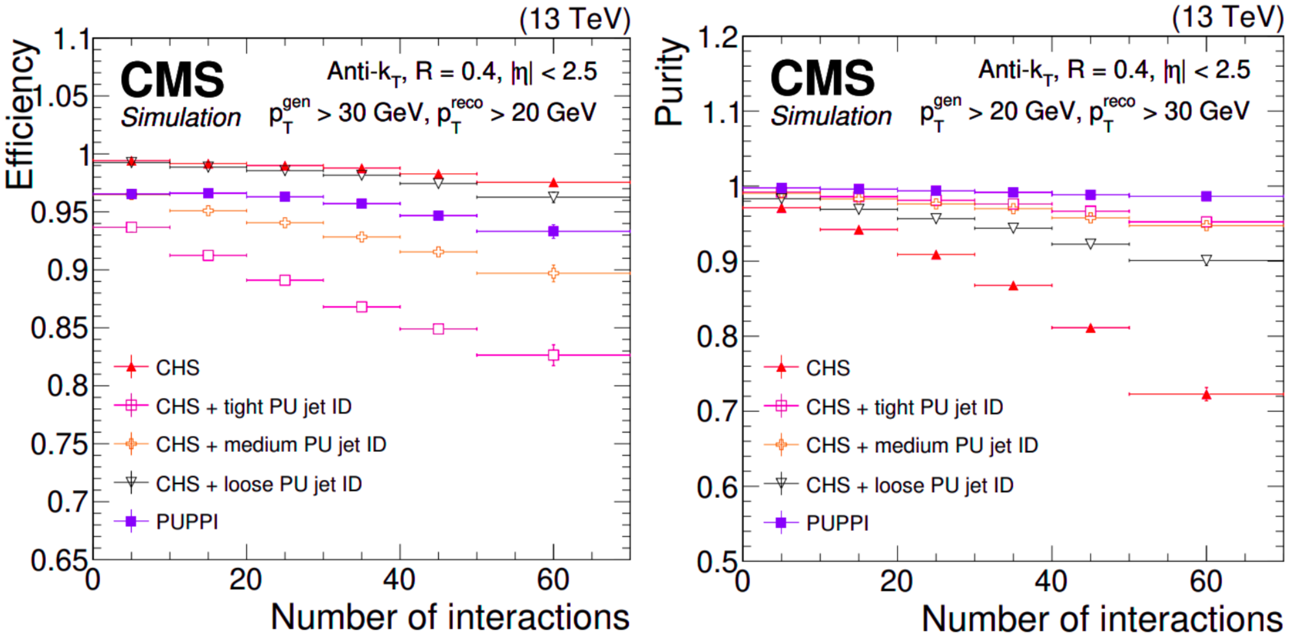


Figure 3.4: Efficiency (left) and purity (right) as a function of the number of interactions for CHS jets (red), CHS jets with a PU jet ID applied (pink, yellow and black) and PUPPI jets (blue), computed using Run 2 $Z + jets$ simulation. The efficiency is shown for AK4 jets with $0 < |\eta| < 2.5$. Statistical uncertainties in simulation are included in the error bars [236].

B-tagging algorithms: DeepCSV vs DeepJet

The b-jets leave distinct signatures in the CMS detector, as illustrated in Figure 3.5. The hadronization process of b-quarks results in the production of B-hadrons. Due to their large mass (a few GeV) and long lifetime (flight distance of approximately a millimeter), these hadrons generate around five displaced tracks in the detector each [241]. Furthermore, the B-hadrons carry 70% of the jet's momentum, and the b-quark hadronization involves leptonic decays in 20% of cases. Given the large production rate of top quarks, which decay into a W boson and a b-quark in almost 100% of cases, and the direct production of b-quarks in QCD processes at the LHC, the identification of b-jets via dedicated algorithms is crucial for analyses involving multiple jets in the event selection, such as the one presented in section 3.2.1. The Deep Combined Secondary Vertex (DeepCSV) and DeepJet are both Deep Neural Network (DNN) algorithms, which aims at identifying the b-jets against the jets originating from a charm quark, known as 'c'-jets, and from the light quarks (up, down, and strange quarks) and the gluons, known as the 'udsg'-jets. They are applied at both the trigger and analysis levels.

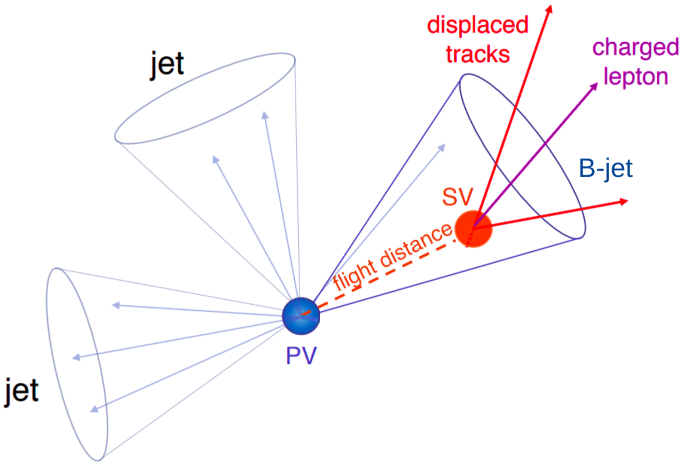


Figure 3.5: Schematic representation of a b-jet signature in the CMS detector [241].

The DeepCSV discriminator is a fully connected DNN consisting of 5 layers with 100 nodes each, taking as input information from the selected tracks and SVs [242]. For each jet candidate, at least two tracks are required, with an angular distance from the jet axis of $\Delta R < 0.3$. The DeepCSV model is trained using QCD and $t\bar{t}$ simulated events, with a splitting between the training, the validation, and the testing samples. Approximately 20 input variables are used per event, using the track variables, the SVs variables, and the correlated variables between them [241]. The training process discriminates the jet flavor f among the b-jets, c-jets, and udsg-jets, such that the DNN output is the probability of a jet belonging to a certain category $P(f)$.

The DeepJet discriminator is using both a Convolution Neural Network (CNN) and DNN [243]. Approximately 650 input variables are used per event, using the global variables, the charged PF candidates variables, the neutral PF candidates variables, and the SVs variables associated with the jets. To surpass the performance of the DeepCSV algorithm, no criterion is applied on the jet candidates. The DeepJet model is trained using QCD and $t\bar{t}$ simulated events, with a splitting between the training, the validation, and the testing samples. Each set of variables is passed separately through a CNN, and then combined in a fully connected DNN consisting of 7 layers with 100 nodes each. The training process discriminates the jet flavor among the b-jets, c-jets, and udsg-jets, such that the DNN output is the probability $P(f)$.

A comparison of both the DeepCSV and DeepJet algorithms performances is presented in

Figure 3.6, with Run 2 QCD simulated samples. The misidentification (mis-id) rate is evaluated as a function of the b-jet efficiency, also known as the b-tagging discriminative score, and computed in the two transverse momentum ranges 150 GeV $< p_T <$ 300 GeV and 300 GeV $< p_T <$ 600 GeV. The mis-id is calculated by comparing the assignment of the jet candidate as a b-jet versus a c-jet, or as a b-jet versus a udsg-jet. In all cases, the DeepJet discriminator reaches significant higher scores compared to those of the DeepCSV discriminator. For a mis-id probability of 1%, the b-jet discrimination against c-jets and udsg-jets increases by 15% at low p_T, and by 20% at high p_T. This difference between the two p_T regimes is expected, since high-p_T jets are likely to have more constituents which are all used during the DeepJet training. The b-tagging scores are also expected to be lower for the b-jet discrimination against the c-jets, since the c-quark hadronization process is similar to that of the b-quark (the c-quark hadronization involves leptonic decays in 10% of cases for example [241]). With this enhanced performance, the DeepJet algorithm has become the recommended algorithm for b-tagging identification for the Run 3 period, whereas the DeepCSV algorithm was the recommended choice during the Run 2 period. Finally, it is important to remember that the b-tagging discriminative scores must be estimated at both the trigger and analysis levels separately, as the jet reconstruction algorithms show better performance at the analysis level. Thus, the discriminative scores cannot be directly compared at both the trigger and analysis levels.

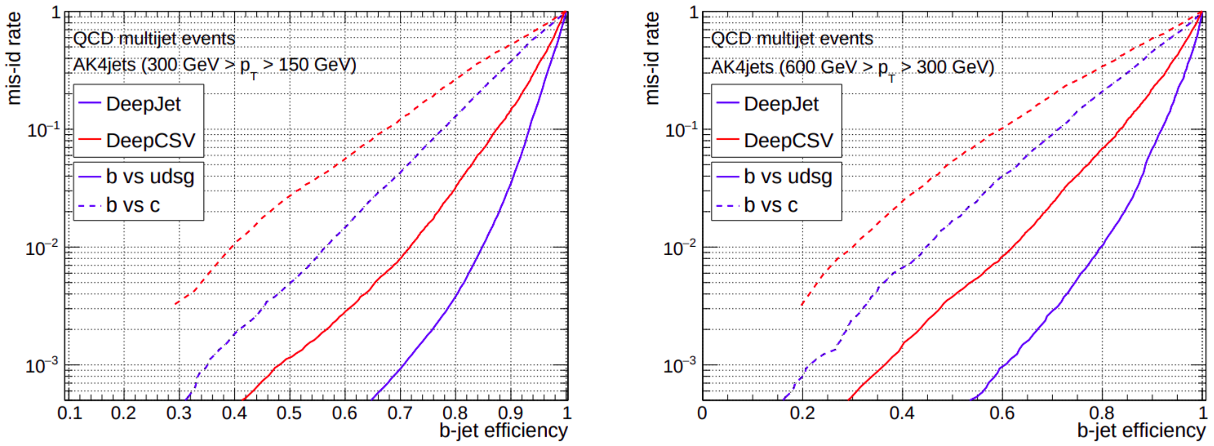


Figure 3.6: Misidentification rate as a function of the b-tagging efficiency for both the DeepCSV (red) and DeepJet (blue) algorithm, computed using Run 2 QCD multijet simulation. The performance is shown for b-jets versus c-jets (dashed lines) and versus udsg-jets (solid lines). [243].

3.2.3 Performances at the High Level Trigger during the Run 2 period

The details of the search for a VLQ T' decaying into the all-hadronic resolved final state are discussed in section 4.1.1. During the Run 2 period, the HLT path 'HLT_PFHT380_SixJet32_DoublePFBTagDeepCSV' was used for this analysis. It requires at least 6 PF jets, with a transverse momentum $p_T > 32$ GeV for each jet candidate, and a scalar sum of the transverse momentum of all jets H_T defined as:

$$H_T = \sum_{jets} p_T > 380 \text{ GeV}. \tag{3.3}$$

At least two of these jets must be identified as b-jets, using the DeepCSV algorithm. Unfortunately, this HLT path was not designed specifically for the targeted event selection, which only requires 5 jets. The offline selection initially required 5 jets with $p_T > 170$, 130, 80, 40, and 40 GeV respectively, while the HLT path requires 6 jets. To ensure that the offline selection was more restrictive than the HLT selection, a sixth jet with $p_T > 40$ GeV was required at the analysis level. The efficiencies for the p_T of the sixth leading jet and the H_T are presented in Figure 3.7, with 2016 $t\bar{t}$ simulated and data samples. The HLT reference path is the 'HLT_IsoMu24', which requires at least one isolated muon with $p_T > 24$ GeV. The efficiency is above 95% for any value of both variables, and the total efficiency is 98% in the T' → Ht decay channel and 95% in the T' → Zt decay channel [165], which is considered acceptable. Nevertheless, the requirement for a sixth jet at the analysis level could introduce a potential bias in the event selection: a new HLT path must be designed and optimized for the Run 3 period. The rest of the chapter will present the studies I have conducted for this period unless otherwise specified.

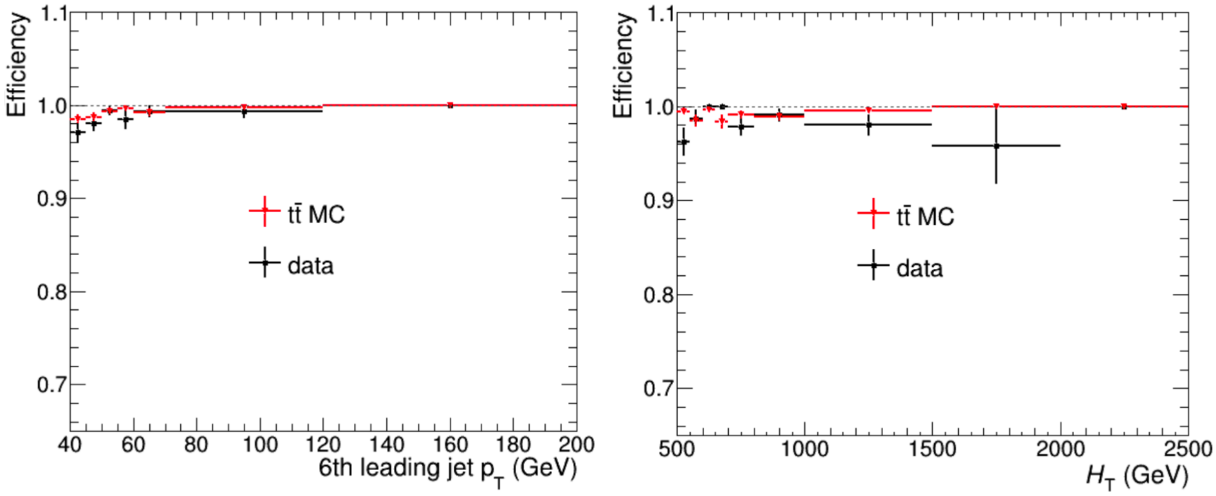


Figure 3.7: Trigger efficiencies of the HLT path as a function of the p_T of the sixth leading jet (left) and the H_T (right) HLT_PFHT380_SixJet32_DoublePFBTagDeepCSV, comparing 2016 $t\bar{t}$ simulation (black) and data (red) [164].

3.2.4 Performances at the High Level Trigger during the Run 3 period

The new Run 3 HLT path requires at least 5 jets in a pseudorapidity range $0 < |\eta| < 4.5$, with a transverse momentum $p_T > 100$, 100, 60, 30, and 30 GeV respectively, $H_T > 400$ GeV and with at least two of these jets identified as b-jets. Events must first pass L1 paths which require at least 4 jets with lower thresholds on p_T, $|\eta|$, and H_T. The criteria which defined the offline selection for the Run 2 analysis are preserved for the Run 3 period. Moreover, I have conducted two major studies to guide the choice of the jet reconstruction method and the b-tagging algorithm. These studies were performed in simulation with a signal sample with a nominal mass of the T' $m_{T'} = 650$ GeV.

Jet reconstruction strategy

Three jet reconstruction methods are compared at the trigger level. A jet which comes from the PV only is labeled as a ROI jet.

- The jets must be reconstructed both as ROI Calo Jets and global PF jets, which was the configuration used for the Run 2 HLT path.

- The jets must be reconstructed both as ROI Calo Jets and ROI PF jets.
- The jets must be reconstructed as ROI PF jets.

A comparison of the HLT efficiencies and pure rates between the three options is presented in Table 3.1. The HLT efficiency was calculated using the same offline selection and HLT reference path as in the Run 2 analysis. The first method achieves a 94.7% efficiency, which is comparable to the Run 2 performance (95%). One part of the decrease in efficiency from 100% to 94.7% is attributed to the use of the PUPPI algorithm at the analysis level only as the PU mitigation algorithm for the Run 3 period, as detailed in section 3.2.2. The impact of this algorithm results in a 2% decrease in efficiency, which is considered acceptable. The remaining decrease in efficiency from 100% to 94.7% may arise from a mismatch between the reconstruction algorithms at both the trigger and analysis levels, which is also considered acceptable. The first method achieves a 1.5% better efficiency with similar pure rates within uncertainties, leading to its selection by the 'Beyond SM particles decaying 2 Higgs, top and Gauge bosons' (B2G) trigger group of the CMS collaboration. The global PF jets reconstruct the jet candidates with all information of the sub-detectors, explaining the higher efficiency of the first method. In contrast, the Calo jets reconstruct the jet candidates with calorimeter information only, explaining similar efficiencies and pure rates between the second and third methods.

Jet reconstruction at the trigger level	Efficiency	Pure rate (Hz)
ROI Calo + global PF jets	94.7%	13.7 ±0.9
ROI Calo + ROI PF jets	93.2%	13.1 ±0.8
No Calo + ROI PF jets	93.2%	13.2 ±0.9

Table 3.1: Efficiencies and pure rates of the studied HLT path for different jet reconstruction methods, computed using 2022 simulation with $m_{T'} = 650$ GeV.

B-tagging algorithm

The DeepCSV and DeepJet algorithms are compared using the same setup at both the trigger and analysis levels, as described in the previous paragraph. A comparison of the HLT efficiencies and pure rates between the DeepCSV and DeepJet algorithms is presented in Table 3.2. The DeepCSV and DeepJet discriminative scores are selected to be above 0.24 and 0.287 at the trigger level respectively, which correspond to a mis-id probability of 5% of b-jets versus udsg-jets. The DeepJet algorithm achieves a higher efficiency than that of the DeepCSV algorithm, which is consistent with the results discussed in section 3.2.2. Furthermore, its pure rate is lowered by 35%, resulting in a much higher rejection of the background, and leading to its selection by the B2G trigger group.

B-tagging algorithm	Efficiency	Pure rate (Hz)
DeepCSV	94.7%	13.7 ±0.9
DeepJet	95.5%	8.9 ±0.7

Table 3.2: Efficiencies and pure rates of the studied HLT path for the DeepCSV and DeepJet algorithms, computed using 2022 simulation with $m_{T'} = 650$ GeV.

Additional validations

The efficiencies are calculated as a function of the p_T of the fifth leading jet and the H_T, as presented in Figure 3.8. The p_T of the fifth leading jet must exceed 40 GeV in the offline selection and 30 GeV in the HLT selection, whereas the H_T must exceed 500 GeV in the offline selection and 400 GeV in the HLT selection. The turn-on curve is observed for both variables around the online thresholds, but the efficiency reaches a plateau for values beyond the offline thresholds. Despite these considerations, the efficiency is always above 95% within uncertainties, which is considered acceptable.

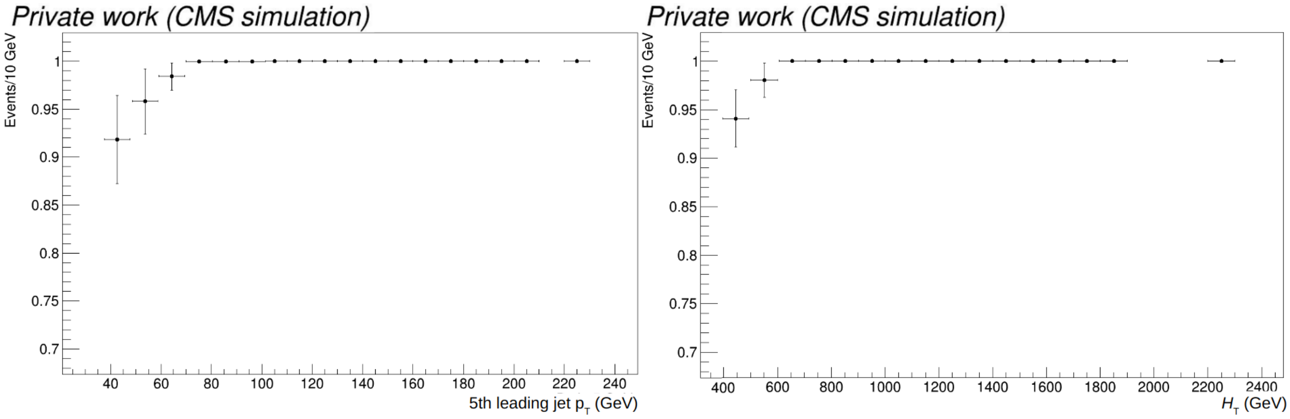


Figure 3.8: Trigger efficiencies of the studied HLT path as a function of the p_T of the fifth leading jet (left) and the H_T (right), computed using 2022 simulation with $m_{T'} = 650$ GeV.

Finally, the timing without and with the contribution of the studied HLT path is presented in Figure 3.9. The pure timing of the HLT path is 5.4 ms, which is completely acceptable. With all these evaluations, the HLT path was validated by the B2G trigger group, and has been deployed in 2022 and 2023 under the label 'HLT_PFHT400_FivePFJet_100_100_60_30_30_DoublePFBTagDeepJet_4p5'.

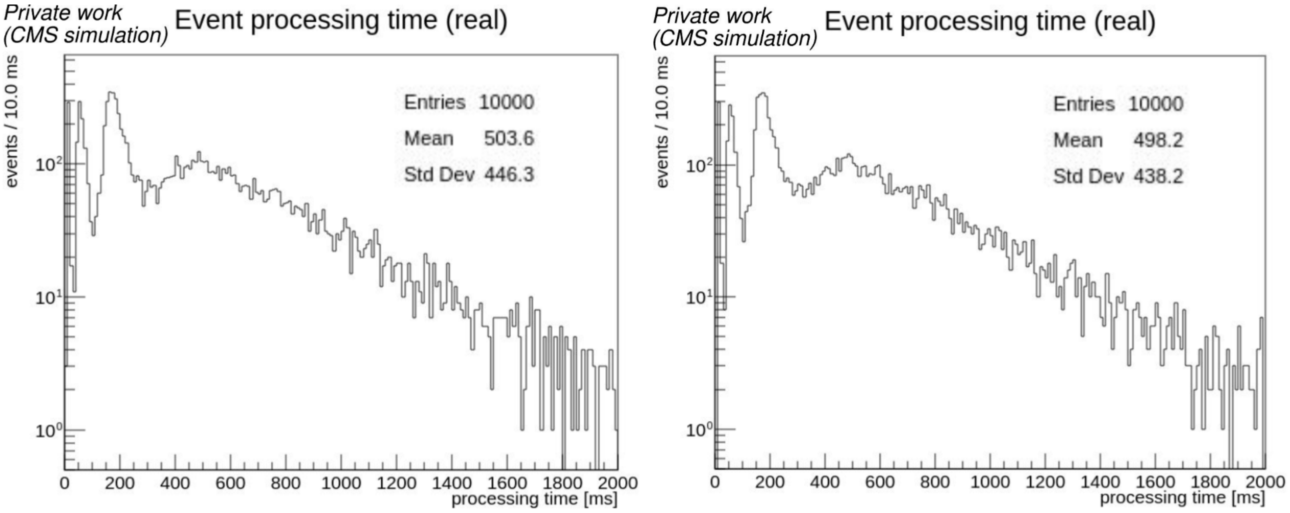


Figure 3.9: Timing of the HLT menu with and without the contribution of the studied HLT path, computed using 2022 simulation with $m_{T'} = 650$ GeV.

First results with the 2022-2023 data

The efficiency as a function of the p_T of the fifth leading jet is presented in Figure 3.10, with 2022 data samples. These results were performed by the B2G trigger group, using the same offline selection and HLT reference path as mentioned in section 3.2.3. The efficiency does not reach a plateau for all p_T values, and remains below 90% even for p_T values significantly exceeding the offline threshold. The average efficiency is 86.1% for the same data-taking period, which is 9.4% lower than the expected efficiency. Additionally, the pure rate is approximately 22 Hz for both 2022 and 2023 data, which is 2.5 times higher than the expected pure rate. This behavior has been observed across all the HLT paths using the DeepJet algorithm, and may be attributed to a malfunction of the algorithm at the trigger level.

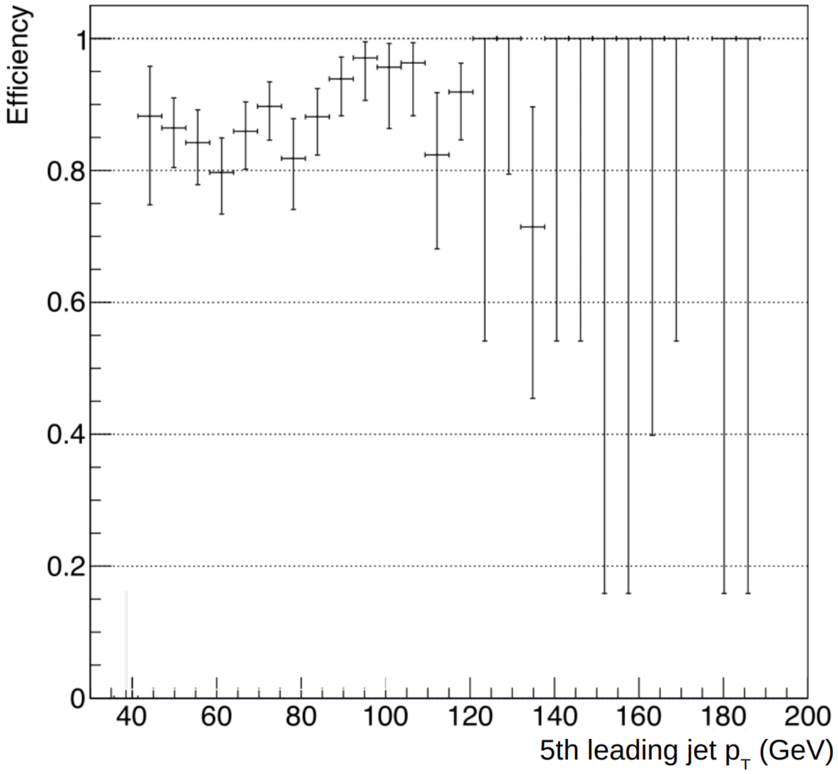


Figure 3.10: Trigger efficiency as a function of the p_T of the fifth leading jet of the HLT path HLT_PFHT400_FivePFJet_100_100_60_30_30_DoublePFBTagDeepJet_4p5, computed using 2022 data [244].

Improvement of the HLT path for the 2024-2025 period

To mitigate the potential saturation of the CMS processors, and increase the trigger efficiency, two solutions have been proposed for the studied HLT path.

- The DeepJet algorithm has been replaced by the ParticleNet algorithm [245]. The ParticleNet algorithm treats jets as 'particle clouds', where particles are grouped as unordered, permutation-invariant, but yet correlated sets of particles. The ParticleNet is a CNN, using approximately 15 input variables per jet. It is trained using QCD simulated events, with a splitting between the training, the validation, and the testing samples. A comparison of the DeepCSV, DeepJet, and ParticleNet algorithms performances is presented in Figure 3.11, with Run 3 $t\bar{t}$ simulated samples. The udsg-jets mis-id rate is evaluated as a function of the b-jet efficiency. For both the DeepJet and ParticleNet discriminators, the efficiencies are presented for the HLT and offline selections, where the offline selection requires jets with $p_T > 25$ GeV. In all cases, the ParticleNet discriminator reaches higher scores compared to those of the DeepJet discriminator. For a mis-id probability of 1%, the b-jet discrimination against udsg-jets increases by 7% with the HLT selection, and by 4% with the offline selection. With this enhanced performance, the ParticleNet algorithm has become the recommended algorithm for the 2024-2025 data-taking period. The efficiency as a function of the p_T of the fifth leading jet of the studied HLT path is presented in Figure 3.12, for both the DeepJet and ParticleNet algorithms. These results were again performed by the B2G trigger group, using the same offline selection and HLT reference path as before. The performance for the DeepJet algorithm remains consistent with 2022 results presented in Figure 3.10, but is lower than Run 3 predictions presented in Figure 3.8. This could come from potential issues in the training of the DeepJet algorithm with the PU increasing for Run 3. On the other hand, the ParticleNet algorithm shows a minimum efficiency of 90% at any value of the p_T of the fifth leading jet.

- The p_T thresholds for the first two jets have been increased from 100 GeV to 120 GeV, and the pseudorapidity range has been adjusted to $0 < |\eta| < 4.3$ to slightly improve the efficiency.

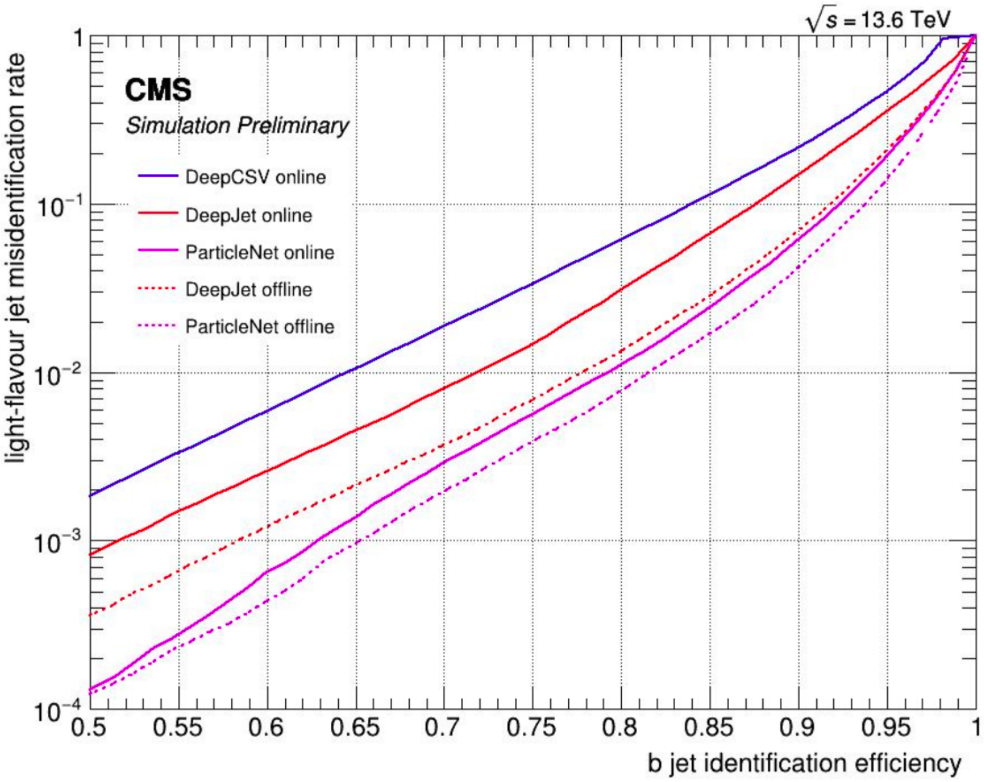


Figure 3.11: Udsg-jet misidentification rate as a function of the b-tagging efficiency for DeepCSV (blue), DeepJet (red) and ParticleNet (purple), computed using 2024 $t\bar{t}$ multijet simulation. The performance is shown for both offline selection (dashed lines) and HLT selection (solid lines) [246].

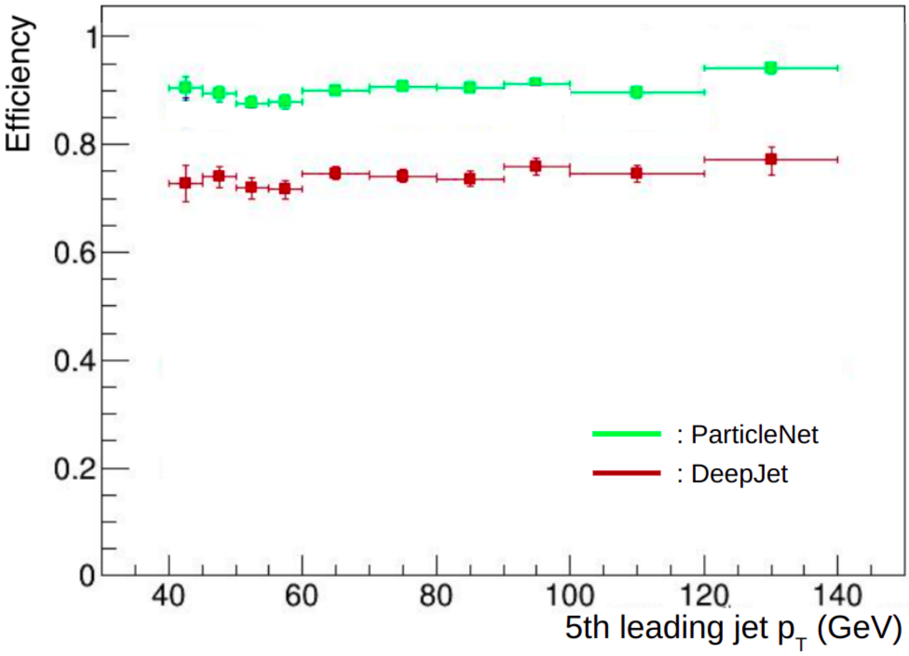


Figure 3.12: Trigger efficiency as a function of the p_T of the fifth leading jet for DeepJet (red) and ParticleNet (green) of the HLT path HLT_PFHT400_FivePFJet_120_120_60_30_30_PNet2BTag_4p3, computed using 2024 simulation with $m_{T'} = 650$ GeV [247].

Consequently, the new HLT path is labeled 'HLT_PFHT400_FivePFJet_120_120_60_30_30_PNet2BTag_4p3'. First results with 2024 data samples indicate a pure rate of approximately 12 Hz, reducing the previous pure rate by almost a factor of 2. Although this pure rate is still higher than the expected pure rate, all the HLT paths using the DeepJet algorithm have been removed. This reduction in the number of HLT paths has decreased the total event rate to approximately 1.5 kHz, which is the maximal rate authorized during the Run 3 period. Moreover, the modified HLT path can benefit from other analyses which already use the ParticleNet algorithm in their offline selection. For all these reasons, the updated HLT path was validated by the B2G trigger group, and will be used, and is being regularly monitored, until the end of the Run 3 period.

Chapter 4

T’ $\rightarrow$ tH same sign dilepton final state: presentation and strategy

The status of the searches for the VLQs to address the hierarchy problem in the Higgs sector has been explored in Chapter 1. Current results exclude the mass of a potential VLQ up to the TeV scale, for both the single and double VLQ production modes. However, an excess has been evidenced at the level of 3σ within the CMS experiment for the decay of a VLQ T’ into a Z or Higgs boson, accompanied with a top quark in the all-hadronic resolved final state, and corresponding to a nominal mass $m_{T'} = 680$ GeV with 2016 data [164]. While the same study was conducted with Run 2 data with negative results [165], other final states must be investigated as cross-check analyses to verify this result. In particular, the decay of a VLQ T’ into a Higgs boson accompanied with a top quark in the SS dilepton final state will be investigated. The analysis will be divided into two chapters: the overall presentation of the analysis and the analysis strategy will be discussed here, while the results and their interpretation will be the subject of Chapter 5.

This chapter will be divided into three sections. The first section will focus on the motivation for searching the VLQ T’ in the SS dilepton final state. It will include a discussion of results in other final states, and the importance of investigating the T’ in this specific final state. The second section will detail the object selection. Extensive studies on both the offline and online selections have been conducted to achieve optimal signal identification. Eventually, we will discuss the strategy used to discriminate the signal from the various background processes in the Signal Region (SR). This strategy is performed in two steps, using simulated events. The first step involves a cut-based selection aimed at reducing the total background contribution as much as possible. The second step optimizes this selection by using quantiles of the background distribution to preserve the shape of the background distribution after each selection criterion.

4.1 Presentation of the Vector-Like Quark T’ $\rightarrow Ht$ analysis in the same sign dilepton final state

The T’ $\rightarrow Ht$ and T’ $\rightarrow Zt$ decays have been studied across multiple channels in both the CMS and ATLAS experiments. These channels correspond to three distinct final states, excluding the SS dilepton final state: the all-hadronic, the diphoton, and the semileptonic final states.

4.1.1 Strategy employed in other T' → Ht final states

The all-hadronic resolved final state (2016 data)

The all-hadronic final state was first introduced in Chapter 3, and a typical Feynman diagram of this final state is presented in Figure 3.2. The event selection requires 5 jets, among which 3 b-jets, and the targeted nominal mass of the T' $m_{T'}$ ranges from 600 GeV to 1200 GeV. The analysis was initially conducted in the CMS experiment with 2016 data, using the invariant mass of the T' $M_{T'}$ as the discriminant observable. Detailed information on the analysis strategy can be found in Ref. [164]. The invariant mass of the T', presented in Figure 4.1 for both the T' → Zt and T' → Ht decays, also shows the signal distribution for a nominal mass of the T' $m_{T'} = 700$ GeV. An excess is observed in the tH channel, corresponding to a deviation from the SM at the level of 3σ for $M_{T'} = 680$ GeV, and with an observed cross section of nearly 2 pb. No excess is reported in the tZ channel. Further studies were conducted with Run 2 data in the same final state to either confirm or infirm this result.

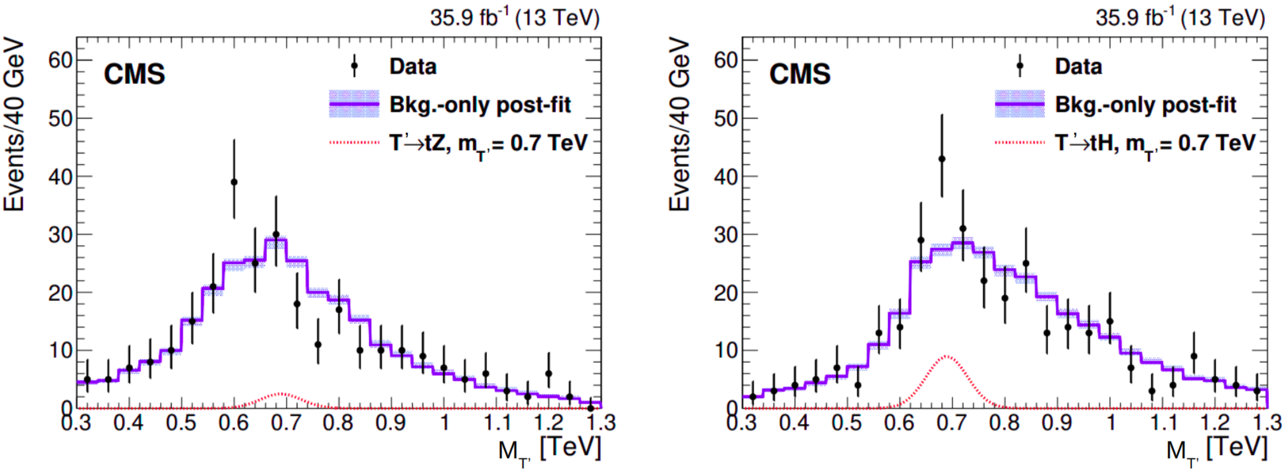


Figure 4.1: Fitted distributions of $M_{T'}$ of the T' → Zt (left) and the T' → Ht (right) decays in the all-hadronic resolved final state, comparing 2016 simulation (blue) and data (black points). The expected signal distribution (red) is shown at $m_{T'} = 700$ GeV [164].

The all-hadronic resolved final state (Run 2 data)

Detailed information on the analysis strategy with Run 2 data can be found in Ref. [165]. For nominal masses of the T' exceeding 800 GeV, the analysis strategy remains unchanged. For nominal masses of the T' below 800 GeV, the background estimation is refined by using quantiles of the background distribution to avoid shaping the $M_{T'}$ background distribution in applying the selection criteria. This refined selection ensures that no bump could possibly be artificially created in the background distribution under a potential signal (see section 4.3.4). The invariant mass of the T', presented in Figure 4.2 for the T' → Ht decay only, also shows the signal for both the nominal masses of the T' $m_{T'} = 700$ GeV and 900 GeV, each scaled by a factor of 10. No excess is reported with Run 2 data, especially for a nominal mass of the T' $m_{T'} = 680$ GeV, and the exclusion limit is set up to 700 GeV. No excess is reported in the ATLAS experiment in the same final state with Run 2 data either, setting up the exclusion limit up to 1200 GeV [166].

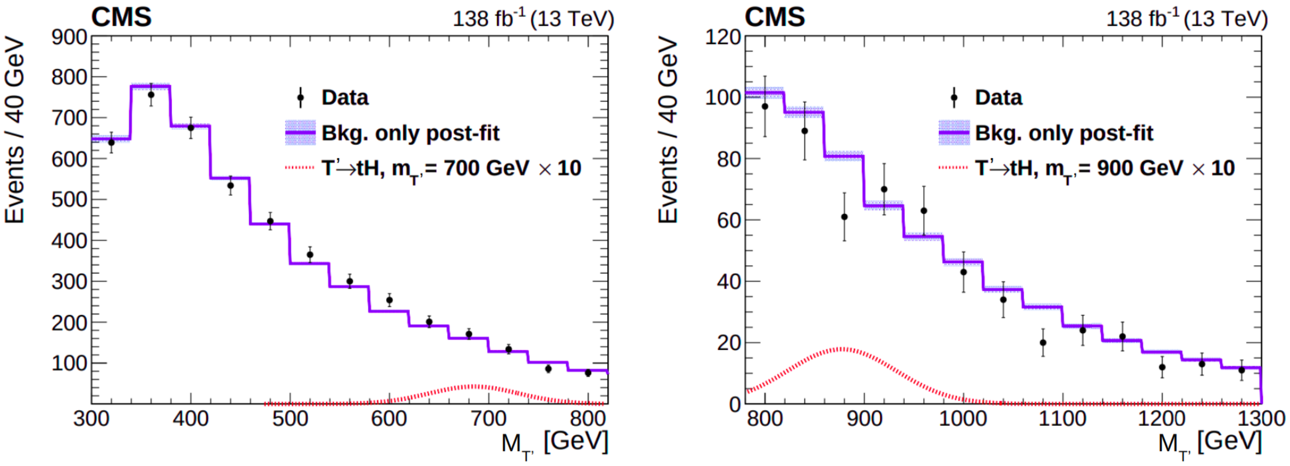


Figure 4.2: Fitted distributions of $M_{T'}$ of the T' → Ht decay in the all-hadronic resolved final state, comparing Run 2 simulation (blue) and data (black points). The expected signal distribution (red) is shown at $m_{T'}$ = 700 GeV (left) and 900 GeV (right) [164].

The diphoton final state (Run 2 data)

A typical Feynman diagram of the studied T' decay is shown in Figure 4.3. The production mode of the T' is the same electroweak production as discussed in section 3.2. The T' → Ht decay is considered. The top quark decays into one W boson and a b-quark in almost 100% of cases, and the W boson subsequently decays into a pair of quark-antiquark $q''\overline{q'''}$ in 67.4% of cases, or into an antilepton and its corresponding neutrino $\bar{l}\nu_l$ in 32.6% of cases [8]. The Higgs boson decays into a pair of photons $\gamma\gamma$ in 0.2% of cases. Despite this very small BR, the high photon reconstruction efficiency and mass resolution mentioned in section 2.2.2 results in a sensitivity comparable to that of the all-hadronic resolved final state. The event selection requires 2 photons, 1 b-jet, and either 2 additional jets (hadronic channel) or 1 lepton accompanied with MET (leptonic channel). The targeted nominal mass of the T' ranges from 600 GeV to 1200 GeV.

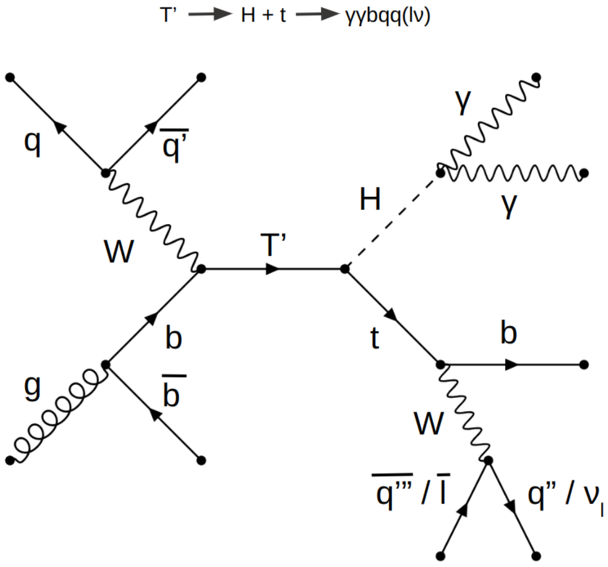


Figure 4.3: Example Feynman diagram for the single production of the VLQ T' and its decay into the diphoton final state via a Higgs boson accompanied with a top quark.

The analysis was conducted with Run 2 data in the CMS experiment, using the diphoton invariant mass $m_{\gamma\gamma}$ as the discriminant observable. Detailed information on the analysis strategy can be found in Ref. [169]. The diphoton invariant mass, presented in Figure 4.4, shows the signal distribution for both the nominal masses of the T' $m_{T'}$ = 600 GeV and 900 GeV, and combine both the hadronic and leptonic channels. No excess is reported, and the exclusion limit is set up to 720 GeV.

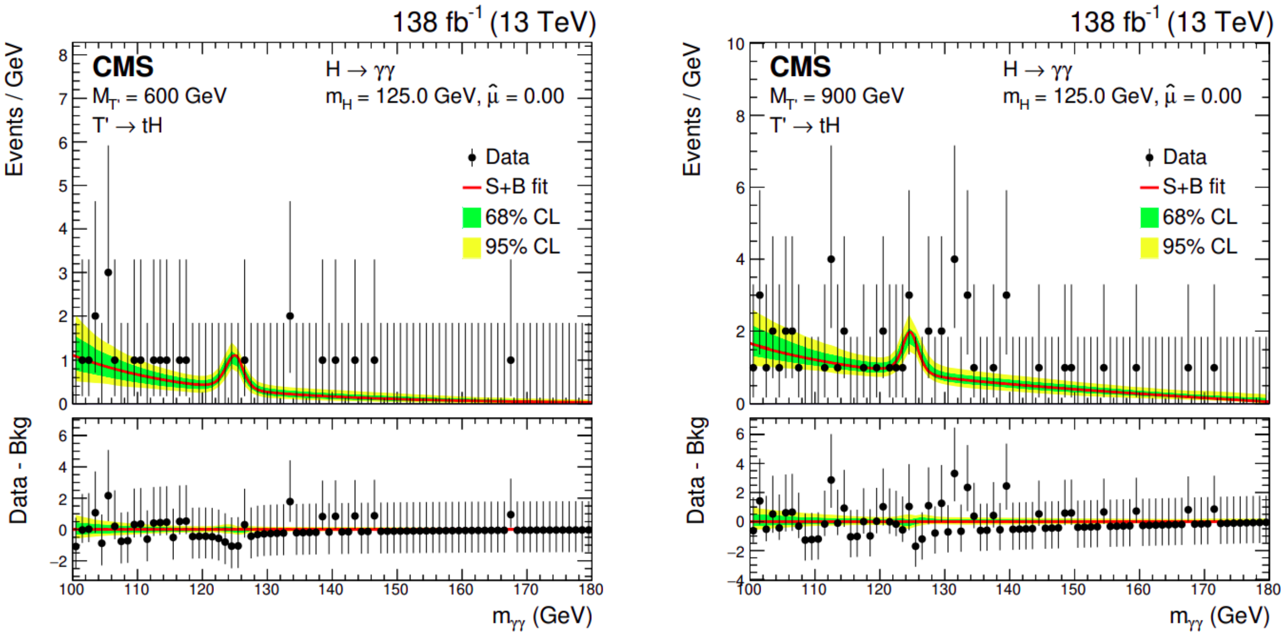


Figure 4.4: Fitted distributions of $m_{\gamma\gamma}$ of the T' → Ht decay in the diphoton final state with both hadronic and leptonic channels combined, comparing Run 2 simulation (red) and data (black points). The signal distribution is fitted at $m_{T'}$ = 600 GeV (left) and 900 GeV (right) [169].

The semileptonic final state (Run 2 data)

A typical Feynman diagram of the studied T' decay is shown in Figure 4.5. The production mode of the T' is the same electroweak production as discussed in section 3.2. The top quark decays into one W boson and a b-quark in almost 100% of cases, and the W boson subsequently decays into an antilepton and its corresponding neutrino $\bar{l}\nu_l$ in 32.6% of cases. The Higgs boson decays into a pair of b-quark-antiquark $b\bar{b}$ in 53% of cases, and the Z boson decays into a pair of quark-antiquark $q''\overline{q'''}$ in 69.9% of cases. The event selection requires 1 lepton accompanied with MET, and either 3 b-jets or 3 jets, among which 1 b-jet. The targeted nominal mass of the T' ranges from 600 GeV to 2200 GeV.

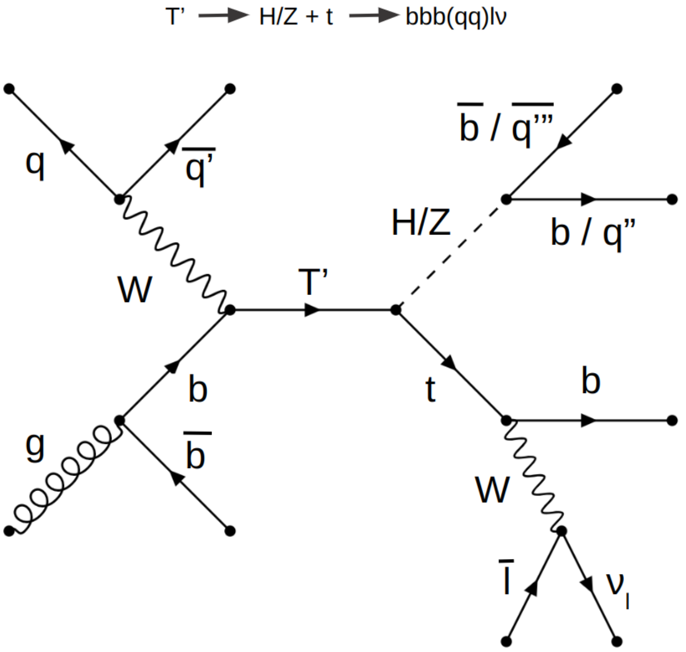


Figure 4.5: Example Feynman diagram for the single production of the VLQ T' and its decay into the semileptonic final state via a Higgs boson accompanied with a top quark.

The analysis was conducted with Run 2 data in the ATLAS experiment. The scalar sum of the p_T of the jets, the leptons, and the MET, known as the m_{eff} variable, defines the discriminant observable. Detailed information on the analysis strategy can be found in Ref. [171]. The effective mass is presented in Figure 4.6. The analysis targets both the T' → Ht and T' → Zt decays. No excess is reported, and the exclusion limit is set up to 1100 GeV.

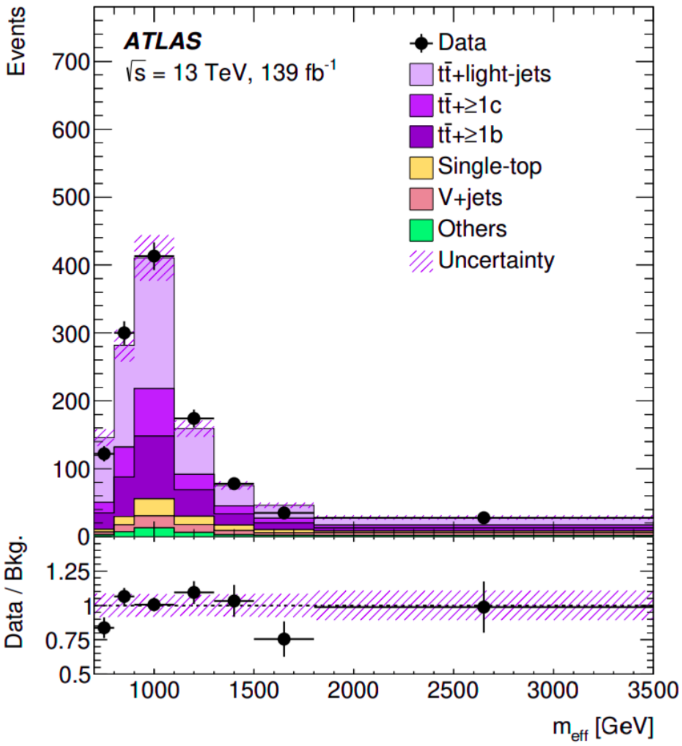


Figure 4.6: Fitted distribution of m_{eff} of the combined T' → Ht and T' → Zt decays in the semileptonic final state, comparing Run 2 simulation (purple) and data (black points) [171].

4.1.2 The same sign dilepton final state

By taking the minimal exclusion limit of all experimental results presented above, the potential exclusion of the VLQ T' in the tH channel is set up to 700 GeV. Nevertheless, all the final states must be considered simultaneously to confirm this result. The multilepton final states have not been investigated yet while the leptons benefit from an excellent identification within the CMS detector, and especially the muons. Thus, we will focus on the SS dilepton final state with Run 2 data and simulated samples from the so-called Ultra Legacy (UL) campaign, which benefits from the latest calibrations of the detector. It is to note that this analysis started before the all-hadronic resolved analysis with full Run 2 data was published.

Signal process

A typical Feynman diagram of the studied T' decay is shown in Figure 4.7. The production mode of the T' is the same electroweak production as discussed in section 3.2. The top quark decays into one W boson and a b-quark in almost 100% of cases, and the W boson subsequently decays into a lepton and its corresponding antineutrino $l\overline{\nu_l}$ in 32.6% of cases. The Higgs boson decays into a pair of W bosons W^+W^- in 21.4% of cases. One of the W boson decays into a pair of quark-antiquark $q''\overline{q'''}$ in 67.4% of cases, while the other W boson decays into a lepton and its corresponding antineutrino $l\overline{\nu_l}$ in 32.6% of cases. It is noteworthy that one of the two W bosons is produced off-shell. The event selection requires 2 SS leptons accompanied with MET, and 3 jets, among which 1 b-jet.

The leptons can be either electrons or muons. The taus are not targeted in this search because they are unstable particles, requiring dedicated algorithms for their reconstruction [248]. Three channels are therefore considered, according to the nature of the leptons: two muons ($\mu\mu$), one electron and one muon (eμ), or two electrons (ee). We do not consider the T' → Zt decay because there is no way of obtaining only 2 SS leptons in the final state, even with the Z boson decaying to 2 OS leptons or 2 neutrinos. There is no attempt in identifying the two jets produced in the initial state either, because they are mostly produced at low p_T in the forward region of the detector, where the jets are poorly reconstructed.

Background processes

The signature of the signal process in the CMS detector is, unfortunately, mimicked by numerous SM processes which constitute the background processes of the analysis, and must be discriminated from the signal. These background processes are typically categorized into the irreducible background, which shares the same final state as the signal, and the reducible background, which has a different final state from the signal, and is therefore easier to suppress.

- The $t\bar{t}X$ processes, where X can be a W, Z, or Higgs boson, constitute the irreducible background of the analysis. A typical Feynman diagram of the $t\bar{t}W$ process is shown on the left in Figure 4.8 as an illustration. The event selection requires 2 SS leptons, one arising from a top quark leptonic decay, and the other one arising from the X decay, accompanied with MET, and jets, among which at least 2 b-jets. There is a complete overlap between the $t\bar{t}X$ processes and the signal process final states. However, by contrast with the signal topology, one of the top quark decays hadronically for the $t\bar{t}X$ processes, indicating that the discrimination between the signal and the irreducible background remains doable.
- The $t\bar{t}$ process constitutes the largest reducible background of the analysis. A typical Feynman diagram of the $t\bar{t}$ process is shown in the middle in Figure 4.8, where both top

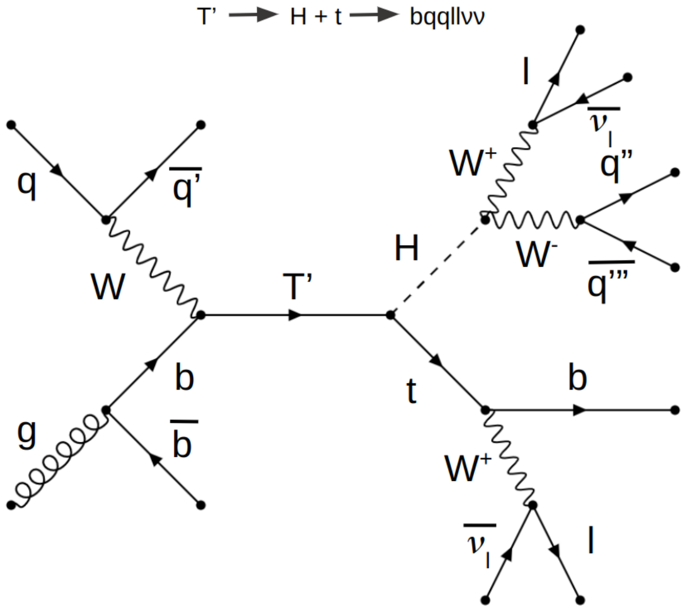


Figure 4.7: Example Feynman diagram for the single production of the VLQ T' and its decay into the SS dilepton final state via a Higgs boson accompanied with a top quark.

quarks decay leptonically ($t\bar{t}$2l process), as an illustration. The event selection requires 0, 1, or 2 leptons accompanied with MET, and jets, among which at least 2 b-jets. In order for $t\bar{t}$2l events to enter in the selected events of the analysis, the charge of one of the two leptons must be misidentified, such that the two leptons are reconstructed as being SS. We also define the $t\bar{t}$1l and $t\bar{t}$0l processes, where one or two leptons must be a jet misidentified as a lepton in a hadronic top quark decay. These background processes will be significantly reduced after the tight identification of the leptons (see section 4.2.2).

- The multiboson processes constitute the second largest reducible background of the analysis. A typical Feynman diagram of the WZ process is shown on the right in Figure 4.8 as an illustration. The event selection requires at least 2 SS leptons, each lepton arising from the decay of one boson, accompanied with MET. In this analysis, multiboson processes include the WW, WW Double parton Scattering (DS), WH, WZ, ZH, ZZ, WWW, WWZ, WZZ, and ZZZ processes as well, where additional jets may be expected in the final state.

- The single top $tX(W/q)$ processes, where X can be a W, Z, or Higgs boson, constitute the third largest reducible background of the analysis. The 2 SS leptons are arising from the top quark leptonic decay and the X or W boson decay, accompanied with MET, and at least 1 b-jet.

- Other processes regroup the $W + jets$ and Drell-Yan (DY) processes, which are considered as minor reducible backgrounds, and are expected to be negligible after the full selection (see section 4.3). In the case of the $W + jets$ process, the 2 SS leptons are arising from the W boson decay and a jet misidentified as a lepton, accompanied with MET, and jets. In the case of the DY processes, the 2 SS leptons are arising from the Z boson decays where the charge of one of the two leptons must be misidentified, accompanied with jets.

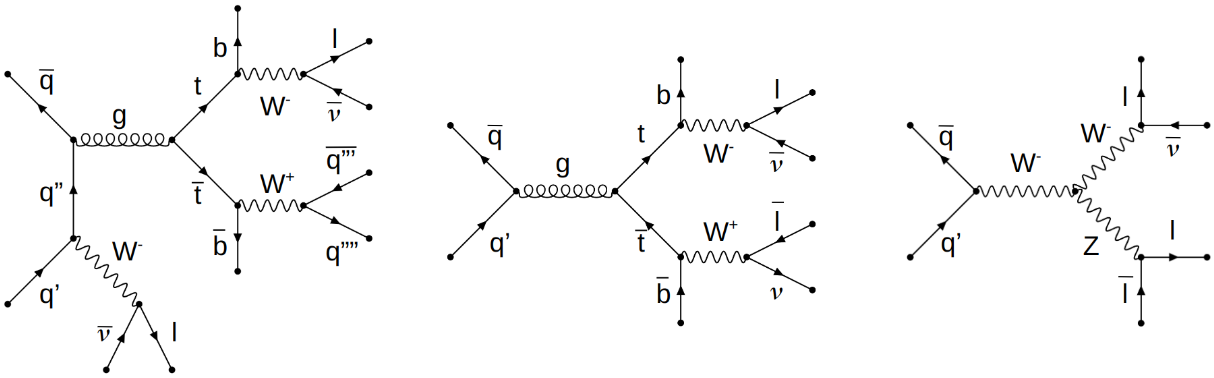


Figure 4.8: Example Feynman diagrams for the $t\bar{t}W$, $t\bar{t}2$l and WZ processes and their decay as main background processes in the SS dilepton analysis.

Data samples

The data samples are provided centrally by the CMS collaboration after the evaluation of data quality by both the trigger and DQM monitoring (see section 3.1.1). If data quality of certain runs is considered insufficient, those runs must be excluded from the data processing during the offline selection. The CMS collaboration supplies a list of all runs with an acceptable data quality which should be considered for offline selection, known as the 'golden' JSON file [249]. The data samples, or primary datasets, are presented in Table 4.1, in addition to the data-taking periods, or 'run eras', and the recorded integrated luminosity (see section 2.1.2). The primary datasets list all the events which pass the L1 trigger, based on the TPs.

The primary datasets are separated between 2016, 2017, and 2018. However, it is to note that 2016 data is split into two separate periods, known as the '2016 preVFP' and '2016 postVFP' periods, due to the substantial change in detector conditions between them. In the 2016 preVFP period, the strip tracker had a lower ratio between signal and noise, and fewer hits on tracks due to saturation effects in the readout chip under high-luminosity conditions. This was mitigated in the 2016 postVFP period by changing the Feedback Preamplifier bias Voltage (VFP) [250]. To enable direct comparison between data and simulation, simulated samples have been separately generated for both the 2016 preVFP and 2016 postVFP periods.

Cross sections and simulated samples of the signal process

The targeted nominal mass of the T' ranges from 600 GeV to 1200 GeV. The simulated samples for the signal, generated for each mass point value of the T', are presented in Table 4.2 with their respective cross section multiplied by the assumed BR of the T' → Ht decay. The cross section is calculated as follows [251]:

$$\begin{aligned} \sigma(m_{T'}, \Gamma_{T'}) &= \left(\frac{e}{\sqrt{2}\sin\theta_W} \kappa^{T'}_{W/Zb} \right)^2 \times \hat{\sigma}_{NWA}(m_{T'}) \\ &= \left(0.458486 \times \kappa^{T'}_{W/Zb} \right)^2 \times \hat{\sigma}_{NWA}(m_{T'}), \end{aligned} \tag{4.1}$$

where θ_W is the Weinberg angle, $\kappa^{T'}_{W/Zb}$ the coupling parameter between the T', the W or Z boson, and the b-quark, and $\hat{\sigma}_{NWA}$ the cross section for the production of the T' in the NWA regime considering all production modes. We assume that the T' is produced at LO in the NWA, which is considered as a good approximation if $\Gamma_T/M_T = 1\%$, and with a coupling to the third-generation SM quarks only. The values of $\kappa^{T'}_{W/Zb}$ and $\hat{\sigma}_{NWA}$ are unique for each mass point value of the T', and can be found in Ref. [252].

Primary dataset	Period	L_{int} (fb)	Era
Single Muon Single Electron Double Muon Double Electron Single Muon and Single Electron	2016 preVFP	19.5	B1 B2 C D E F
Single Muon Single Electron Double Muon Double Electron Single Muon and Single Electron	2016 postVFP	16.8	F G H
Single Muon Single Electron Double Muon Double Electron Single Muon and Single Electron	2017	41.5	B C D E F
Single Muon Double Muon Double Electron Single Muon and Single Electron	2018	59.8	A B C D

Table 4.1: Primary datasets recorded for the full Run 2 period with the periods, the recorded integrated luminosity and the eras [192].

The simulated samples for the signal are generated at LO, using the generator Madgraph5_aMC@NLO [219], and interfaced with the parton-shower generator Pythia8 [223] (see section 2.2.3). The underlying event is tuned using the CMS Pythia Tune 5 (CP5) tune [253]. Additional PU within the same or nearby bunch crossings are included in the simulated events, and events are reweighted according to the observed number of interactions per bunch crossing. The same procedure is applied to all the simulated samples for the background.

The number of generated events N_{GEN} is presented for each mass point value of the T' in Table 4.2. They are reweighted during their generation to take higher-order corrections into account, which is why some events may have a negative weight. To enable direct comparison between data and simulation, simulated events must be reweighted based on their cross section and the integrated luminosity as follows:

$$N_{\text{after selection, expected}} = \frac{N_{\text{after selection}} \times \sigma \times \text{BR} \times L_{int}}{N_{\text{GEN}}}, \tag{4.2}$$

where $N_{\text{after selection}}$ is the number of events after the selection. The integrated luminosity is $L_{int} = 137.6\ \text{fb}^{-1}$ for the full Run 2 period [192]. Before any selection, the number of expected generated events $N_{\text{GEN, expected}}$ is calculated as the product of the cross section, the branching ratio, and the integrated luminosity, and is presented in the last column of Table 4.2. If the selection efficiency and detector acceptance are both 100%, $N_{\text{after selection, expected}}$ is expected to be $N_{\text{GEN, expected}}$.

$m_{T'}$ (GeV)	σ (T' → Ht) × BR [fb] (order)	N_{GEN}	$N_{\text{GEN, expected}}$
600	176.4 (LO)	1,190,000	24,273
625	148.9 (LO)	1,200,000	20,489
650	121.3 (LO)	1,076,000	16,691
675	105.0 (LO)	1,198,000	14,448
700	88.6 (LO)	1,200,000	12,191
800	45.9 (LO)	1,192,000	6,316
900	25.1 (LO)	1,200,000	3,454
1000	14.5 (LO)	1,200,000	1995
1100	8.67 (LO)	1,200,000	1,193
1200	5.36 (LO)	1,190,000	738

Table 4.2: Simulated signal samples produced for the full Run 2 period, with the cross section multiplied by the BR of the T' → Ht decay and the number of generated events before and after the scaling to the integrated luminosity.

Cross sections and simulated samples of the background process

The simulated samples for the background, generated for each process, are presented in Table 4.3 with their respective matrix element generator (they are all interfaced with Pythia8), cross section, along with N_{GEN} and $N_{\text{GEN,expected}}$. It is important to note than even though the processes are generated at a given order, the assigned cross sections may include higher-order corrections to provide a more accurate representation of the theory. Before any selection criteria, signal-to-background ratio ranges from 5.50×10^{-6}% to 2.0×10^{-4}% for the considered mass point values of the T'. Consequently, a dedicated selection strategy is crucial to efficiently identify the signal signature, and maintain control over the background.

4.2 Particle reconstruction and identification

All the events in the MC and data samples must pass the same selection process. This selection is divided into two parts: the online selection at the HLT level and the offline selection.

4.2.1 Selection at the HLT level

The analysis uses the recommendations from the B2G trigger group of the CMS collaboration [254]. It consists of a logical 'OR' of single electron and single muon HLT trigger paths, with a prescale set to 1, as listed in Table 4.4. Although double lepton HLT trigger paths are provided centrally by the CMS collaboration, these paths require looser criteria for the lepton identification which are not relevant for a clear identification of the targeted signal signature. Events must first pass L1 paths which require at least one electron or muon with lower thresholds on p_T than the HLT selection. The same trigger selection is applied for both data and MC samples, and any eventual double counting of events entering into several separate trigger paths at once has been taken care of.

Background process	Generator (order)	$\sigma\times$ BR [fb] (order)	N_{GEN}	$N_{\mathrm{GEN,\,expected}}$
$t\bar{t}H$	Madgraph5 + FxFx (NLO) [255]	211.2 (NLO+NNLL) [256]	26,585,712	29,061
$t\bar{t}$W	Madgraph5 + FxFx (NLO) [255]	235 (NLO+NNLL) [257]	7,707,231	32,336
$t\bar{t}Z$	Madgraph5 + MLM (LO) [258]	859 (NLO+NNLL) [256]	95,439,728	118,198
$t\bar{t}$2l	Powheg (pQCD NLO) [259]	88,341.9 (NNLO+NNLL) [260, 261]	23,997,847,000	12,155,845
$t\bar{t}$1l	Powheg (pQCD NLO) [259]	365,457.4 (NNLO+NNLL) [260, 261]	330,797,424,000	50,286,938
$t\bar{t}$0l	Powheg (pQCD NLO) [259]	377,960.7 (NNLO+NNLL) [260, 261]	242,184,208,000	52,007,392
WW	Powheg (pQCD NLO) [262]	12,178 (NNLO) [263]	252,341,860	1,675,693
WW DS	Pythia8 (LO)	223.2 (NLO) [219]	2,008,000	30,712
WH	Powheg (pQCD NLO) [262]	31.3 (NLO) [219]	10,340,945	4,307
WZ	Madgraph5 + FxFx (NLO) [255]	4,429.7 (NLO) [219]	340,590,590	609,527
ZH	Powheg + JHUGen (NLO) [264]	185.8 (NLO) [219]	20,836,000	25,566
ZZ (2 leptons)	Powheg (pQCD NLO) [262]	564.4 (NLO) [219]	126,284,701	77,661
ZZ (4 leptons)	Powheg (pQCD NLO) [262]	1,256 (NLO) [219]	393,682,675	172,826
WWW	Madgraph5 (LO)	208.6 (NLO) [219]	118,693	28,703
WWZ	Madgraph5 (LO)	165.1 (NLO) [219]	95,406	22,718
WZZ	Madgraph5 (LO)	55.65 (NLO) [219]	50,038	7,657
ZZZ	Madgraph5 (LO)	14.76 (NLO) [219]	8,357	2,031
tW	Powheg (pQCD NLO) [220, 221, 265, 266]	39,650 (NNLO) [267, 268]	1,191,135,500	5,455,840
tHq	Madgraph5 (LO)	70.96 (NNLO) [269]	64,601,186	9,764
tHW	Madgraph5 (LO)	15.61 (NNLO) [269]	42,971,150	2,148
tZq	Madgraph5 + CKM NLO (NLO)	73.58 (NNLO) [269]	2,207,787	10,125
$W + jets$	Madgraph5 (LO)	61,334,900 (NNLO) [270, 271]	6,0296,666,000	8,439,682,240
DY (10-50 GeV)	Madgraph5 (LO)	18,610,000 (NNLO) [270, 271]	2,111,121,070	2,560,736,000
DY (50 GeV)	Madgraph5 (LO)	6,077,220 (NNLO) [270, 271]	676,716,338	836,225,472

Table 4.3: Simulated background samples produced for the full Run 2 period, with the cross section and the number of generated events before and after the scaling to the integrated luminosity.

- For 2016, the single electron trigger requires at least one electron, which must satisfy a tight cut-based ID WP and $p_T > 27$ GeV at the HLT level. The single muon trigger requires at least one muon, which must satisfy isolation requirements using tracking information, and $p_T > 24$ GeV at the HLT level.
- For 2017, the single electron trigger requires at least one electron, which must satisfy a tight ID WP and $p_T > 32$ GeV at the HLT level. The single muon trigger requires at least one muon, which must satisfy isolation requirements using both tracking and muon chambers information, and $p_T > 27$ GeV at the HLT level.
- For 2018, the single electron trigger requires at least one electron, which must satisfy a tight ID WP and $p_T > 32$ GeV at the HLT level. The single muon trigger requires at least one muon, which must satisfy isolation requirements using both tracking and muon chambers information, and $p_T > 24$ GeV at the HLT level.

Year	HLT trigger path name	Prescale
2016	HLT_Ele27_WPTight_Gsf	1
	HLT_IsoTkMu24	1
2017	HLT_Ele32_WPTight_Gsf_L1DoubleEG	1
	HLT_IsoMu27	1
2018	HLT_Ele32_WPTight_Gsf	1
	HLT_IsoMu24	1

Table 4.4: List of the single electron and single muon HLT trigger paths for 2016, 2017 and 2018.

The trigger efficiencies for both the single electron and single muon HLT trigger paths are presented in Figure 4.10 as a function of the p_T of the leading lepton, with 2018 simulated and data samples. The trigger MC efficiencies are calculated using the signal sample with a nominal mass of the T' $m_{T'} = 700$ GeV, while data efficiencies are calculated using the MET primary datasets provided centrally by the CMS collaboration. The p_T-spectrum of both electrons and muons for the signal sample are presented in Figure 4.9 before any selection, as a comparison with Figure 4.10. Both efficiencies are calculated requiring 2 SS leptons in the offline selection as the first criterion of the preselection discussed in section 4.3.1, and the HLT_PFMET120_PFMHT120_IDTight path defines the reference HLT path, which is orthogonal to the studied HLT trigger paths (see section 3.1.2). Efficiency reaches the 100% plateau for simulation for p_T values exceeding offline thresholds. Efficiency does not always reach the 100% plateau for data but is above 90% for p_T values exceeding offline thresholds, within statistical uncertainties, which is considered acceptable given the large statistical uncertainties. The results are similar for the 2016 preVFP, 2016 postVFP, and 2017 periods, and can be found in Appendix A.1. The HLT trigger paths are then validated for the analysis.

4.2.2 Offline selection

The targeted final state is reconstructed with different objects identified in various CMS subdetectors. Consequently, distinct identification criteria must be employed for each of these objects.

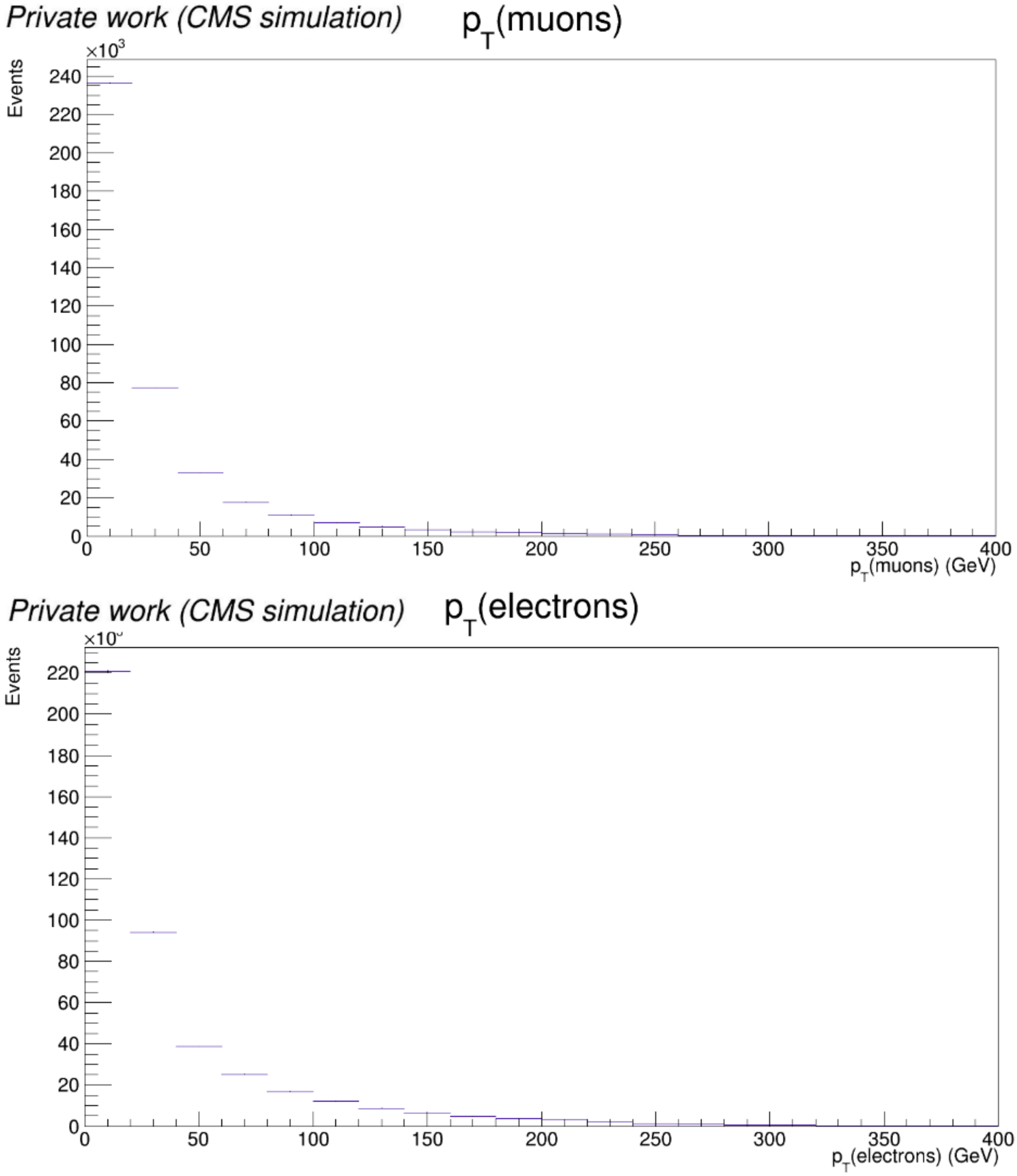


Figure 4.9: Distributions of the p_T for both electrons (left) and muons (right), computed using 2018 simulation with $m_{T'} = 700$ GeV.

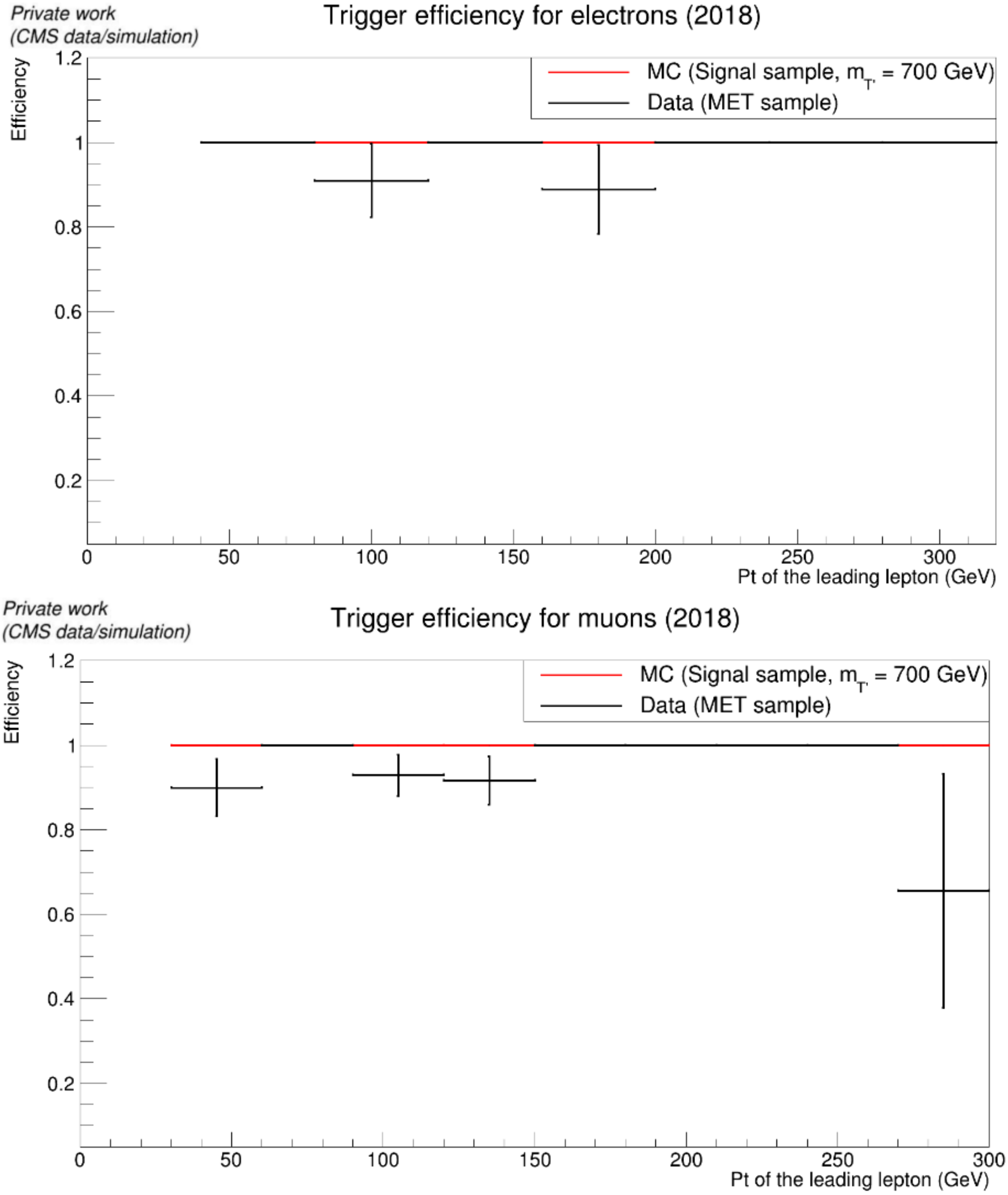


Figure 4.10: Trigger efficiencies for both electron (top) and muon (bottom) HLT trigger paths, comparing 2018 simulation (red) and data (black) samples with $m_{T'} = 700$ GeV.

Electron reconstruction and identification

Electron candidates are reconstructed by associating the ECAL clusters and GSF tracks, thanks to the PF algorithm (see section 2.2.2). They are required to satisfy $p_T > 40$ GeV, within the pseudorapidity region of $|\eta| < 2.5$. Because of the strong magnetic field, the energy deposits of the electrons are spread in the η-φ plane. Therefore, we reconstruct the electrons using ECAL clusters, known as superclusters [272]. These superclusters must lie outside of the transition between the two regions of the ECAL, satisfying $0 < |\eta| < 1.4442$ and $1.566 < |\eta| < 2.5$. The 3D Impact Parameter Significance SIP_{3D}, as well as the distance in the xy-plane d_{xy} and in the z direction d_z with respect to the PV, must satisfy $\mathrm{SIP}_{3D} < 4$ cm, $|d_{xy}| < 0.05$ cm, and $|d_z| < 0.1$ cm to reject electrons originating from taus and displaced vertices.

Electron candidates must satisfy the tight cut-based ID WP to reject the jets misidentified as electrons, where criteria are designed to achieve an efficiency of 70% [273]. The criteria include a selection on supercluster and track information in the ECAL. The ratio of the HCAL to ECAL energy is also employed to ensure that the electron's energy lies in the ECAL only, distinguishing it from jets.

To reject electrons originating from converted photons, additional requirements are made on the number of pixel layers which are not associated with any hit, and on the conversion probability during the vertex reconstruction. These requirements define the 'Conversion veto' and the 'Missing inner hits' criteria respectively. To suppress events with two OS leptons, where the charge of an electron is misreconstructed, it is required that three independent electron charge measurements, derived from different cluster information, are in agreement. This requirement defines the 'tight charge' criterion.

Finally, electrons are required to be isolated. The isolation is constructed using three FOMs [274–276].

- The relative mini isolation I^{rel}_{mini} is defined as follows [277]:

  $$I^{rel}_{mini} = \frac{\sum\limits_R \Big(p_T(h^\pm)\Big) + \max\left(0, \sum\limits_R \Big(p_T(h^0) + p_T(\gamma) - \rho A \tfrac{R^2}{0.3}\Big)\right)}{p_T(e)}, \tag{4.3}$$

  where ρ is the PU energy density, A the effective area correction, and where $\sum\limits_R \Big(p_T(h^\pm)\Big)$, $\sum\limits_R \Big(p_T(h^0)\Big)$, and $\sum\limits_R \Big(p_T(\gamma)\Big)$ refer respectively to the sum of the p_T of the charged hadrons, neutral hadrons, and photons, within a cone R depending on the electron p_T:

  $$R = \frac{10}{\min\left(\max\Big(p_T(e), 50\Big), 200\right)}. \tag{4.4}$$

  Requiring I^{rel}_{mini} below a given threshold ensures that the electron is locally isolated, even in topologies where its transverse momentum exceeds a hundred of GeV. The presence of the last term in the numerator mitigates the impact of PU.

- p_T^{ratio} is defined as follows:

  $$p_T^{ratio} = \frac{p_T(e)}{p_T(jet)}, \tag{4.5}$$

  where $p_T(e)$ is the p_T of the electron, and $p_T(jet)$ the p_T of the jet closer than $\Delta R < 0.4$ from the electron. Most times, this corresponds to the jet which contains the electron. If the electron is not inside of a jet, the jet closest to the electron is used. The use of the p_T^{ratio} allows to discriminate the electrons from the jets which include them, or which are misidentified as electrons.

- p_T^{rel} is defined as follows:

$$p_T^{rel} = \frac{|\left(\overrightarrow{p}(jet) - \overrightarrow{p}(e)\right) \times \overrightarrow{p}(e)|}{|\overrightarrow{p}(jet) - \overrightarrow{p}(e)|}. \tag{4.6}$$

  Requiring p_T^{rel} above a given threshold allows recovering electrons from accidental overlap with jets. If the jet and the electron are spatially separated, the numerator will be indeed large and the denominator small.

An electron is considered isolated if the following conditions are satisfied:

$$I_{mini}^{rel} < I_1 \text{ AND } \left(p_T^{ratio} > I_2 \text{ OR } p_T^{rel} > I_3\right). \tag{4.7}$$

The values I_i, i = 1, 2, 3 are summarized for both electrons and muons in Table 4.6. They define the tight WP of the isolation. Because the probability to misidentify a jet as an electron is higher than that of a muon, tighter isolation values are used for electrons.

A relaxed 'fakeable' electron selection is also defined. Electrons which pass the fakeable selection, but not the tight selection, are labeled as fakeable electrons. Fakeable electrons must satisfy the loose WP of the same ID criteria, designed to achieve an efficiency of 90%, and the loose WP of I_{mini}^{rel}, satisfying $I_{mini}^{rel} < 0.40$. The rest of the criteria is preserved. Both selections are summarized in Table 4.5.

Muon reconstruction and identification

Muon candidates are reconstructed by associating the tracks in both the tracker and muon chambers thanks to the PF algorithm. They are required to satisfy $p_T > 30$ GeV, within the pseudorapidity range of $|\eta| < 2.4$. The criteria on the SIP$_{3D}$, $|d_{xy}|$, and $|d_z|$ are identical to the electron criteria. The tight charge criterion for the muons is based on the ratio between the error on the p_T and the p_T of the best muon track. Muon candidates must satisfy the medium ID WP to reject the jets misidentified as muons, where criteria are designed to achieve an efficiency of 97% [278]. The criteria include a selection on track information in both the tracker and muon chambers. The tight ID WP was also tried but leads to reduced significance by roughly 20% after the full selection. Finally, muons are required to be isolated the same way as the electrons.

A relaxed 'fakeable' muon selection is also defined. Muons which pass the fakeable selection, but not the tight selection, are labeled as 'fakeable' muons. Fakeable muons must satisfy the loose WP of I_{mini}^{rel}, satisfying $I_{mini}^{rel} < 0.40$. The rest of the criteria is preserved. Both selections are summarized in Table 4.5.

Lepton reconstruction and identification

In summary, two electron and two muon selections are defined, as summarized in Table 4.5.

- A first category defines the SR and the $t\bar{t}$X (ttX) Control Region (CR), requiring exactly two tight SS leptons.

- A second category defines the $t\bar{t}$ dilepton (tt2l) Measurement Region (MR), requiring exactly two tight OS leptons.
- A third category defines the $t\bar{t}$ semileptonic (tt1l) MRs, requiring exactly one fakeable lepton and one tight lepton, with SS.

The SR will be discussed in section 4.3, while the CR and MRs will be discussed in Chapter 5.

Criterion	Muons		Electrons	
	Fakeable	Tight	Fakeable	Tight
p_T (GeV)	> 30	> 30	> 40	> 40
$\|\eta\|$	< 2.4	< 2.4	< 1.4442 OR (> 1.566 AND < 2.5)	
Identification	Medium ID	Medium ID	Loose ID	Tight ID
Isolation	Loose WP	Tight WP	Loose WP	Tight WP
$\|d_{xy}\|$ (cm)	< 0.05	< 0.05	< 0.05	< 0.05
$\|d_z\|$ (cm)	< 0.1	< 0.1	< 0.1	< 0.1
SIP$_{3D}$	< 4	< 4	< 4	< 4
Missing inner hits	-	-	$= 0$	$= 0$
Conversion veto	-	-	Yes	Yes
Tight charge	Yes	Yes	Yes	Yes

Table 4.5: Summary of the lepton selection for both electrons and muons.

Isolation	Muon Tight WP		Electron Tight WP	
Year	2016	2017-2018	2016	2017-2018
I_1	0.16	0.11	0.12	0.07
I_2	0.76	0.74	0.80	0.78
I_3	7.2	6.8	7.2	8.0

Table 4.6: Isolation WPs for both electrons and muons for 2016, 2017 and 2018.

Jet and Missing Transverse Energy reconstruction and identification

Jet candidates are reconstructed from the PF candidates, using the AK algorithm in a cone of a radius parameter $\mathcal{R} = 0.4$ (see section 2.2.2). They are required to satisfy $p_T > 30$ GeV, within the pseudorapidity range of $|\eta| < 2.4$. Jets candidates closer than $\Delta R < 0.4$ from the selected leptons are not considered to avoid double counting. This ΔR cleaning is applied to selected leptons. Jets candidates must satisfy the tight cut-based ID WP, where criteria are designed to achieve an efficiency of 98% [279]. The criteria include a selection on track information and energy deposits of all the CMS sub-detectors.

In 2018, an issue has been reported in the HCAL modules labeled 'HEM15/16', for which power supply died in the middle of the 2018B era until the end of the data taking period [280]. It is estimated that the energy has decreased by 35% for jets with $-1.57 < \varphi < -0.87$ and $-3.0 < \eta < -2.5$. For that reason, the jets reconstructed in this $\varphi - \eta$ region are rejected for 2018 data.

The b-jets are selected the same way as the jets, with an additional criterion on the DeepJet discriminator as recommended for the Run 2 period (see section 3.2.2). The b-jet candidates must exceed the medium WP threshold, which correspond to a mis-id probability of 5% of b-jets versus udsg-jets. These thresholds are 0.2598, 0.2489, 0.3040, and 0.2783 for the 2016 preVFP, 2016 postVFP, 2017, and 2018 periods respectively [281].

Lastly, the MET is calculated as the negative sum of the p_T of all the PF particles reconstructed in an event, as defined in equation 2.9.

4.2.3 Object corrections

The MC samples are generated with a complete simulation of the CMS detector (see section 2.2.3). However, this simulation is not perfect due to imperfect knowledge of the material budget for instance. The discrepancy between data and simulation necessitates the renormalization of the MC events using weights, known as the SFs, which must be calculated for each targeted object.

The tag and probe method

The tag and probe method is the standard technique developed in high-energy physics to measure any SF, by exploiting di-object resonances such as a Z boson or a J/Ψ meson [282].

- Step 1: resonances are reconstructed as pairs, with one leg passing a tight identification (tag), and the other one passing the loose identification under study (probe).
- Step 2: the efficiency ϵ is the ratio of the number of events passing both the tag and probe identifications, to the number of events passing the tag identification but failing the probe identification. This calculation is performed for both MC and data samples, thereby defining the SF as follows:

$$SF = \frac{\epsilon_{Data}}{\epsilon_{MC}}. \tag{4.8}$$

The procedure is repeated in bins of the η-p_T plane of the studied object to calculate the SFs as a function of those variables. These maps are provided centrally by the CMS collaboration for all the variables presented below.

Electron corrections

The electron SFs are provided centrally by the CMS collaboration, and divided into three parts: the trigger, the reconstruction, and the ID SFs [283]. The SFs introduce an absolute correction up to 7%, particularly in high-η regions where the electron reconstruction is less efficient. The results are similar for all the years, and can be found in Appendix A.2.1. It is important to note that other variables used in the electron reconstruction introduce minor corrections, which is considered acceptable to neglect (see section 5.2.3).

Muon corrections

The muon SFs are provided centrally by the CMS collaboration, and divided into four parts: the trigger, the reconstruction, the ID, and the isolation SFs [284–286]. Only low-p_T muons have an absolute reconstruction correction up to 20%, but these muons are rejected during the muon selection process. The rest of the SFs introduces an absolute correction of less than 5%. The results are similar for all the years, and can be found in Appendix A.2.2. It is important to note that other variables used in the muon reconstruction are neglected for the same reason as for the electrons.

Jet and Missing Transverse Energy corrections

To mitigate the effects of the detector miscalibration and PU contamination, the jet calibration, known as the Jet Energy Corrections (JEC), is divided into three steps, which do not systematically use the tag and probe method [287]. These corrections are provided centrally by the CMS collaboration, and rescale the p_T and mass of the jets.

- Step 1: the offset energy induced by PU and noise is calculated separately in data and MC samples, such that the ratio between the two defines the SF. For jets with $0 < |\eta| < 2.5$, this SF introduces an absolute correction up to 20%.
- Step 2: the detector non-uniformity is calculated for MC samples as the ratio of the p_T of the jet at the reconstruction level to the p_T of the jet at the particle level, and then applied to both MC and data. For jets with $p_T > 30$ GeV, this correction improves the jet response by up to 40%.
- Step 3: residual corrections, which accounts for minor differences between MC and data samples, are calculated for data only using the tag and probe method. For jets with $p_T > 30$ GeV, this correction improves the jet response by up to 4 %.

After applying the JEC to the jets, the last step involves computing the SF for Jet Energy Resolution (JER) as the ratio of JER in both data and MC samples. This correction introduces an absolute correction up to 15%. The JEC and JER corrections are provided centrally by the CMS collaboration for the 2016 preVFP, 2016 post VFP, 2017, and 2018 periods separately [288]. Finally, the JEC and JER are propagated to the MET, rescaling its p_T [289].

B-tagging corrections

As discussed in Chapter 3, the b-jet discrimination against the c-jets and the udsg-jets using the DeepJet algorithm is performed on a statistical basis, leading to a potential misidentification where jets from b-quarks can be misclassified as non b-tagged and vice versa. Consequently, two separate corrections are necessary for the jet collection: one for the jets passing the identification criterion, and another for the jets failing this criterion. These corrections are computed following the procedure described below [290].

- Step 1 : we define the two efficiencies ϵ_{b-tag} and ϵ_{light}. The efficiency ϵ_{b-tag} is the ratio of the number of selected b-jets matched to a b-quark at the particle level, to the number of jets matched to a b-quark at the particle level. The efficiency ϵ_{light} is the ratio of the number of selected b-jets matched to a jet which does not contain a b-quark at the particle level, to the number of jets matched to a jet which does not contain a b-quark at the particle level. These efficiencies are calculated using the cut-based selection detailed in section 4.3.2, and are presented in Figure 4.11 in the p_T-η plane of the jets, with 2018 simulated samples. The efficiency ϵ_{b-tag} is above 60% across the entire detector, except for high-p_T and high-$|\eta|$ regions where jet reconstruction efficiency is lower, and statistical uncertainties are important. The efficiency ϵ_{light} is below 10% across the entire detector, except for high-p_T and high-$|\eta|$ regions for the same reasons as before. The results for the 2016 preVFP, 2016 postVFP, and 2017 periods are similar, and can be found in Appendix A.2.3.

- Step 2: the probability that all the jets within the jet collection are identified as b-jets, in both data and simulation, is defined as follows:

$$\begin{aligned} P(MC) &= \prod_{i=tagged} \epsilon_i \prod_{j=not\ tagged} (1-\epsilon_j) \\ P(Data) &= \prod_{i=tagged} SF_i\ \epsilon_i \prod_{j=not\ tagged} (1-SF_j\ \epsilon_j). \end{aligned} \tag{4.9}$$

  The probabilities $P(MC)$ and $P(Data)$ are calculated for both ϵ_{b-tag} and ϵ_{light}. The SFs are the ratio of the b-tagging efficiency between data and MC samples here, and are provided centrally by the CMS collaboration [291]. For the corrections applied to the b-jets, the SFs are calculated using data and b-enriched QCD simulated samples. For the correction applied to the non b-jets, the SFs are calculated using data and standard QCD simulated samples.

- Step 3: the weight w is defined as follows:

$$w = \frac{P(Data)}{P(MC)}. \tag{4.10}$$

  This weight is applied to MC events only, similar to a standard SF.

In summary, selected MC events will be reweighted by multiplying all the corrections detailed in this section. Data samples are not corrected to enable direct comparison with simulation.

4.3 Event selection in the Signal Region

Identifying the objects composing the SS dilepton final state of the T' → Ht decay is necessary, but insufficient for distinguishing the signal signature from background processes. A dedicated selection must be designed to maximize background reduction, and the signal must be extracted using a specific variable known as the discriminant observable of the analysis. This procedure defines the SR.

4.3.1 Preselection

In the DY and the ttZ background processes, a Z boson decays into two leptons. Therefore, events with a pair of SS leptons and with an invariant mass within a Z boson mass window of 80 GeV $< M_{ll} <$ 100 GeV are rejected. Additionally, events failing the filters known as the 'noise filters', designed to discard events affected by detector miscalibration issues, are rejected. They are provided centrally by the CMS collaboration [292]. This selection defines the preselection or cut 0, and is summarized in Table 4.7.

Preselection	p_T threshold (GeV)	$\|\eta\|$ threshold
HLT selection	Section 4.2.1	Section 4.2.1
Exactly two SS tight leptons ($\mu\mu$/eμ/ee)	Muon: > 30 Electron: > 40	Muon: < 2.4 Electron: < 1.4442 OR (> 1.566 AND < 2.5)
At least three jets	> 30	< 2.4
At least one b-jet	> 30	< 2.4
$M_{ll} \notin [80, 100]$ GeV	-	-
Noise filters	-	-

Table 4.7: Preselection of the events.

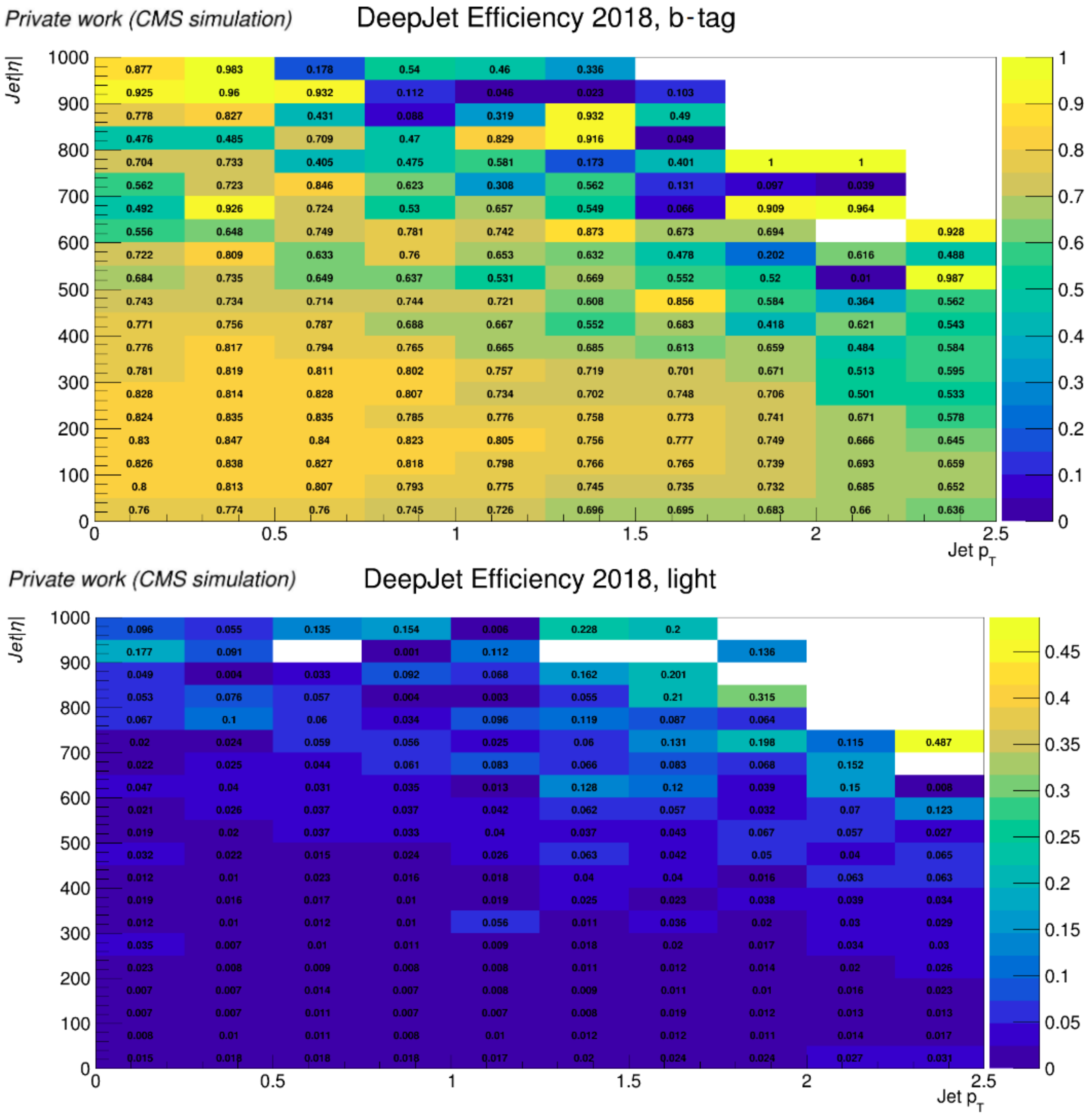


Figure 4.11: Efficiencies ϵ_{b-tag} (top) and ϵ_{light} (bottom), computed using in 2018 simulation with $m_{T'} = 700$ GeV.

The number of selected signal and background events after each criterion of the preselection are presented in Table 4.8 for 2018. The $\mu\mu$, eμ, and ee channels are combined, and all the values are scaled to the integrated luminosity as presented in equation 4.2. The 'Others' column includes the $t\bar{t}$0l, DY, and $W + jets$ processes. The signal efficiency ϵ_{signal} is the ratio of the number of selected events where the leptons are matched to the two leptons arising respectively from the top quark and the Higgs boson at the particle level, to the total number of events where the leptons are matched to the two leptons arising respectively from the top quark and the Higgs boson at the particle level. It is important to note that the signal sample includes all the possible top quark and Higgs boson decays, while the efficiency is computed by considering the studied final state only.

$$\epsilon = \frac{\text{Particle level + Offline selection}}{\text{Particle level selection}} \tag{4.11}$$

All the background processes are reduced by at least 99% when requiring exactly 2 tight SS leptons. The multiboson processes are again reduced by 99% when requiring at least 3 jets among which at least 1 b-jet. Other background processes which present a Z boson in the final state are almost entirely suppressed when applying the Z boson veto criterion. The signal efficiency is considered acceptable after the preselection. The signal-to-background ratio S/B is 0.89% after the preselection, and increases by 4 orders of magnitude relative to no cut.

Criterion	Signal	ϵ_{signal} (%)	$t\bar{t}$X	$t\bar{t}$1l	$t\bar{t}$2l	Multibosons	$tX(W/q)$	Others	S/B (%)
No criterion	5.30×10^{3}	-	9.93×10^{4}	2.18×10^{7}	5.22×10^{6}	1.10×10^{6}	2.37×10^{6}	5.16×10^{9}	1.02×10^{-4}
HLT selection	888.57	78.6	2.31×10^{4}	7.06×10^{6}	2.84×10^{6}	4.96×10^{5}	1.04×10^{6}	1.30×10^{8}	6.30×10^{-4}
= 2 SS leptons	9.95	27.0	443.58	362.84	202.21	2.99×10^{3}	100.79	1.24×10^{3}	0.19
$\geq$ 3 jets	5.11	13.9	321.28	190.21	50.64	223.57	28.86	16.49	0.61
$\geq$ 1 b-jet	4.46	12.3	287.12	145.78	42.60	32.08	22.90	3.23	0.83
Z boson veto	4.29	11.9	260.91	127.63	38.89	29.61	20.96	0.19	0.89
Noise filters	4.28	11.8	260.57	127.39	38.89	29.57	20.94	0.19	0.89

Table 4.8: Number of events for the signal and background processes after each criterion of the preselection, computed using 2018 simulation with $m_{T'}$ = 700 GeV. The signal efficiency ϵ_{signal} is also provided.

4.3.2 Cut-based selection

The analysis strategy is built on a 'cut-based' approach, which is a series of sequential criteria. In a first attempt, the strategy is optimized at the nominal mass of the T' $m_{T'}$ = 700 GeV, since the all-hadronic resolved final state reported an excess for $M_{T'}$ = 680 GeV with 2016 data [164]. To build the set of selection criteria, we use the modified Punzi FOM defined as follows [293, 294]:

$$\text{Punzi}_{\text{FOM},i} = \frac{S_i}{\sqrt{B_i} + \frac{A}{2} + 0.2 \times B_i}, \tag{4.12}$$

where S_i and B_i are the numbers of signal and background events in the bin i of the distribution. The factor A is the statistical significance, aimed for a deviation from the SM at the level of 3σ, and the term proportional to the background contribution accounts for a 20% systematic uncertainty on the background knowledge. The Punzi FOM is plotted as a function of the cut value on all the variables listed in Table 4.9 with 2018 simulated samples. The i^{th} bin which maximizes the Punzi FOM across all the variables is chosen to define the cut. However, a sufficient amount of selected signal events must be maintained, which is why the maximal value of the Punzi FOM is not always chosen as the final cut, as illustrated in Figure 4.12. The signal efficiency ϵ_{signal} defined in equation 4.11, and used in Table 4.10, is employed. The optimized criteria forming the cut-based selection are presented below.

- Cut 1: $\Delta R(l_1, l_2) > 1.8$. The Punzi FOM for this criterion is shown on top in Figure 4.12. As the top quark and the Higgs boson are mostly produced back-to-back, the two leptons must be spatially separated. This criterion tends to reduce all types of background processes.

- Cut 2: $p_{T,l_1} + p_{T,l_2} > 160$ GeV. The Punzi FOM for this criterion is shown on bottom in Figure 4.12. The large mass of the T' tends to boost both the top quark and the Higgs boson decays, such that the two leptons will gain a large amount of momentum. This criterion tends to reduce all types of background processes which have low-p_T leptons.

- Cut 3: $\min\left(|M_{inv}(3j) - m_{top}|\right) > 34\,\mathrm{GeV}$. There is no hadronic top quark decay in signal events. In this formula, $M_{inv}(3j)$ is computed with the three jets of the selected jets collection among jet permutations such that the closest value of $M_{inv}(3j)$ to the nominal top quark mass m_{top} is chosen. The threshold is chosen to be equal to the one determined in Ref. [165], which reconstructs the top quark mass, such that $M_{inv,3j} \notin m_{top} \pm 2\sigma$. The Punzi FOM is not used for this cut. In this formula, $M_{inv}(3j)$ is computed with the three jets of the selected jets collection among jet permutations such that the closest value of $M_{inv}(3j)$ to the nominal top quark mass m_{top} is chosen.

When computing the b-tagging efficiencies illustrated in Figure 4.11 considering events in all the regions simultaneously (see section 5.1), a simplified selection must be applied. In that case, only the preselection, the cut 1, and the cut 2 are applied to construct the map of the b-tagging efficiencies.

Variable	Description
$p_T(l_1)$	p_T of the leading lepton
$p_T(l_2)$	p_T of the subleading lepton
$p_{T,l_1} + p_{T,l_2}$	Scalar sum of p_T of the two leptons
$\Delta\eta(l_1, l_2)$	$\Delta\eta$ between the two leptons
$\Delta\varphi(l_1, l_2)$	$\Delta\phi$ between the two leptons
$\Delta R(l_1, l_2)$	ΔR between the two leptons
$p_T(l_1 + l_2)$	p_T of the vectorial sum of the two leptons
$M_{inv}(l_1 + l_2)$	Invariant mass of the two leptons
$N(jets)$	Number of jets
$p_T(j_1)$	p_T of the leading jet
$p_T(j_2)$	p_T of the subleading jet
$p_T(j_3)$	p_T of the subsubleading jet
$N(bjets)$	Number of b-jets
$p_T(b_{j_1})$	p_T of the leading b-jet
H_T	Scalar sum of p_T of all jets in the event
S_T^{lep}	Scalar sum of p_T of the leptons and all jets in the event
S_T	Scalar sum of p_T of the leptons, all jets in the event and MET

Table 4.9: List of all the studied variables for the computation of the Punzi FOM.

Cut	Signal	ϵ_{signal} (%)
No cut	5.30×10^3	-
Cut 0	4.28	11.8
Cut 1	4.17	11.6
Cut 2	3.00	8.4
Cut 3	2.49	7.4

Table 4.10: Number of events for the signal after each cut of the cut-based selection, computed using 2018 simulation with $m_{T'} = 700$ GeV. The signal efficiency ϵ_{signal} is also provided.

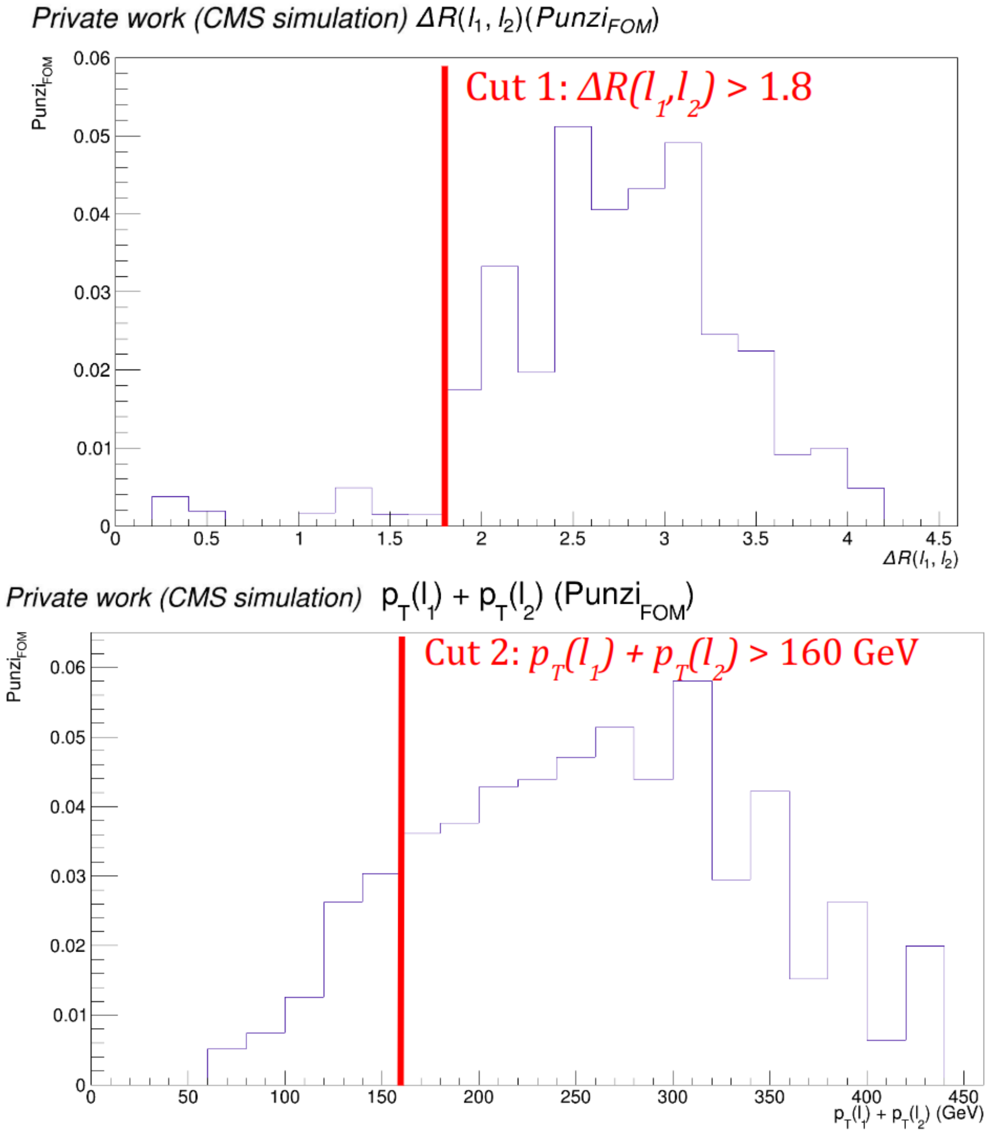


Figure 4.12: Punzi FOMs for the $\Delta R(l_1, l_2)$ (top) and $p_{T,l_1} + p_{T,l_2}$ (bottom) criteria and the chosen cuts of the cut-based selection, computed using 2018 simulation with $m_{T'} = 700$ GeV.

4.3.3 Discriminant observable in the Signal Region

The discriminant observable is the variable chosen for distinguishing the signal distribution from the background distribution in the likelihood fit (see Chapter 5). Initially, we used the variable S_T^{lep} as defined in Table 4.9, and as illustrated in Figure 4.13 with 2018 simulated samples. To quantify the signal discrimination from the background, we use the significance S, expressed in terms of the number of standard deviations σ, and defined as follows [177]:

$$S = \sqrt{2 \ln Q} = \sqrt{2 \sum_{bins} (S_i + B_i) \ln \left(1 + \frac{S_i}{B_i}\right) - S_i}, \tag{4.13}$$

where $\ln Q$ is the Log-Likelihood Ratio (LLR) [295]. S_i and B_i correspond to the respective number of signal and background events in the bin i. This formula is valid if the number of events is large, and if the different bins of the studied variable are independent, which is the case here. The higher the significance, the easier it is to discriminate the signal from the background. Even though the Punzi FOM takes systematic effects into account, it is computed separately in each bin and cannot be used in a global combination of the histogram bins, which is why the significance defined in equation 4.13 will be used here. With 2018 simulated samples, the

$\mu\mu$, eμ and ee channels combined, and $m_{T'} = 700$ GeV, the significance is $S = 0.246\sigma$ for the variable S_T^{lep}.

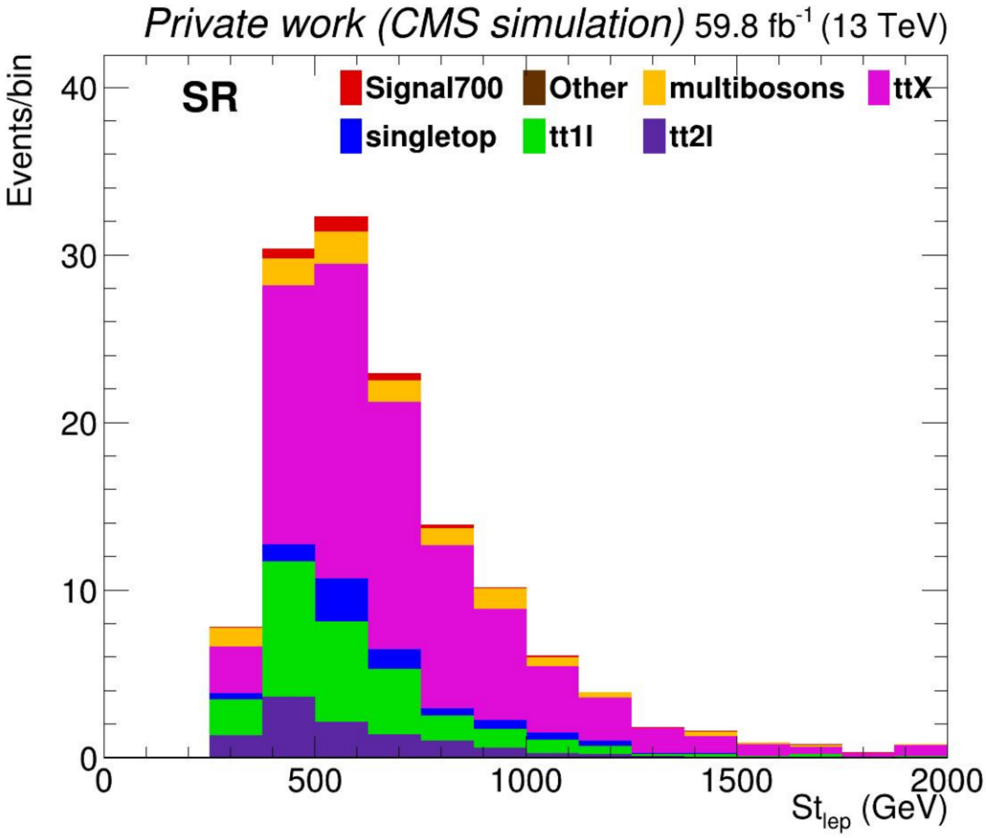


Figure 4.13: Distribution of S_T^{lep} after the cut-based selection, computed using in 2018 simulation with $m_{T'} = 700$ GeV.

We will now attempt at constructing a discriminant observable which could improve the significance relative to the use of the variable S_T^{lep}. We propose the transverse mass of the T' $M_{T,T'}$ as the new discriminant observable in the SR, defined as follows:

$$\begin{aligned} M_{T,T'}^2 =& \left(\sum_{particles} E_T\right)^2 - \left(\sum_{particles} \overrightarrow{p_T}\right)^2 \\ =& \Big(E_{T,lep,Higgs} + E_{T,\nu,Higgs} + E_{T,j1,Higgs} + E_{T,j2,Higgs} + E_{T,lep,top} + E_{T,\nu,top} + E_{T,b,top}\Big)^2 \\ & - \Big(\overrightarrow{p_{T,lep,Higgs}} + \overrightarrow{p_{T,\nu,Higgs}} + \overrightarrow{p_{T,j1,Higgs}} + \overrightarrow{p_{T,j2,Higgs}} + \overrightarrow{p_{T,lep,top}} + \overrightarrow{p_{T,\nu,top}} + \overrightarrow{p_{T,b,top}}\Big)^2 . \end{aligned} \tag{4.14}$$

Since neutrinos cannot be directly reconstructed within the CMS detector, we must assume additional information regarding their transverse energy and transverse momentum, which satisfy the following relations:

$$\begin{aligned} E_{T,\nu,Higgs} + E_{T,\nu,top} =& \sqrt{\overrightarrow{p_{T,\nu,Higgs}}^2 + M_{T,\nu,Higgs}^2} + \sqrt{\overrightarrow{p_{T,\nu,top}}^2 + M_{T,\nu,top}^2} \\ =& \left|\overrightarrow{p_{T,\nu,Higgs}}\right| + \left|\overrightarrow{p_{T,\nu,top}}\right| \\ =& \sqrt{\left(\left|\overrightarrow{p_{T,\nu,Higgs}}\right| + \left|\overrightarrow{p_{T,\nu,top}}\right|\right)^2} \\ =& \sqrt{\left|\overrightarrow{p_{T,\nu,Higgs}}\right|^2 + \left|\overrightarrow{p_{T,\nu,top}}\right|^2 + 2\left|\overrightarrow{p_{T,\nu,Higgs}}\right|\left|\overrightarrow{p_{T,\nu,top}}\right|} . \end{aligned} \tag{4.15}$$

$$\begin{aligned}\overrightarrow{p_{T,MET}}^2 &= \left(\overrightarrow{p_{T,\nu,Higgs}} + \overrightarrow{p_{T,\nu,top}}\right)^2 \\ &= \left|\overrightarrow{p_{T,\nu,Higgs}}\right|^2 + \left|\overrightarrow{p_{T,\nu,top}}\right|^2 + 2\left|\overrightarrow{p_{T,\nu,Higgs}}\right|\left|\overrightarrow{p_{T,\nu,top}}\right|\cos(\nu_H,\nu_t).\end{aligned} \tag{4.16}$$

- Hypothesis 1: the two neutrinos are anticollinear $\Big(\cos(\nu_H,\nu_t) = -1\Big)$. This may be a good approximation since the top quark and the Higgs boson are mostly produced back-to-back, such that:

$$\Rightarrow E_{T,\nu,Higgs} + E_{T,\nu,top} = \sqrt{\overrightarrow{p_{T,MET}}^2 + 4 \times \left|\overrightarrow{p_{T,\nu,Higgs}}\right|\left|\overrightarrow{p_{T,\nu,top}}\right|}. \tag{4.17}$$

$$\Rightarrow M_{T,T'}=\sqrt{\left(E_{T,lep,Higgs}+E_{T,j1,Higgs}+E_{T,j2,Higgs}+E_{T,lep,top}+E_{T,b,top}+\sqrt{\overrightarrow{p_{T,MET}}^2+4\times\left|\overrightarrow{p_{T,\nu,Higgs}}\right|\left|\overrightarrow{p_{T,\nu,top}}\right|}\right)^2-\left(\overrightarrow{p_{T,ljjlb}}+\overrightarrow{p_{T,MET}}\right)^2}. \tag{4.18}$$

- Hypothesis 2: the value of the $\left|\overrightarrow{p_{T,\nu,Higgs}}\right|$ is taken as the average value of $p_{T,\nu,Higgs}$ at particle level, computed at a given nominal mass value of the T'. We have two cases.
  - Case 1: if $\left|\overrightarrow{p_{T,\nu,Higgs}}\right| > \left|\overrightarrow{p_{T,\nu,top}}\right|$,

$$\begin{aligned}\left|\overrightarrow{p_{T,MET}}\right| &= \left|\overrightarrow{p_{T,\nu,Higgs}}\right| - \left|\overrightarrow{p_{T,\nu,top}}\right|. \\ \Leftrightarrow \left|\overrightarrow{p_{T,\nu,top}}\right| &= \left|\overrightarrow{p_{T,\nu,Higgs}}\right| - \left|\overrightarrow{p_{T,MET}}\right|.\end{aligned} \tag{4.19}$$

  - Case 2: if $\left|\overrightarrow{p_{T,\nu,Higgs}}\right| < \left|\overrightarrow{p_{T,\nu,top}}\right|$,

$$\begin{aligned}\left|\overrightarrow{p_{T,MET}}\right| &= \left|\overrightarrow{p_{T,\nu,top}}\right| - \left|\overrightarrow{p_{T,\nu,Higgs}}\right|. \\ \Leftrightarrow \left|\overrightarrow{p_{T,\nu,top}}\right| &= \left|\overrightarrow{p_{T,\nu,Higgs}}\right| + \left|\overrightarrow{p_{T,MET}}\right|.\end{aligned} \tag{4.20}$$

In summary, the final expression of the $M_{T,T'}$ is as follows:

$$M_{T,T'}=\sqrt{\left(E_{T,lep,Higgs}+E_{T,j1,Higgs}+E_{T,j2,Higgs}+E_{T,lep,top}+E_{T,b,top}+\sqrt{\overrightarrow{p_{T,MET}}^2+4\times\left|\overrightarrow{p_{T,\nu,Higgs}}\right|\times\left||\overrightarrow{p_{T,\nu,Higgs}}|\pm|\overrightarrow{p_{T,MET}}|\right|}\right)^2-\left(\overrightarrow{p_{T,ljjlb}}+\overrightarrow{p_{T,MET}}\right)^2}. \tag{4.21}$$

First, to validate the two hypotheses, the mean value and absolute standard deviation of the variables $\cos(\nu_H,\nu_t)$ and $\left|\overrightarrow{p_{T,\nu,Higgs}}\right|$ are presented in Table 4.11, for each mass point value of the T', and with 2018 simulated samples. Only the preselection is applied. The first hypothesis is valid with a negative value of $\cos(\nu_H,\nu_t)$, especially at high masses. For the second hypothesis, the standard deviation of $\left|\overrightarrow{p_{T,\nu,Higgs}}\right|$ is large. To mitigate the statistical fluctuations from a given nominal mass of the T' to another, this distribution of $\left|\overrightarrow{p_{T,\nu,Higgs}}\right|$ is fitted as a function of $m_{T'}$ with a first-degree polynomial in $m_{T'}$, as presented in Figure 4.14, and as follows:

$$\left|\overrightarrow{p_{T,\nu,Higgs}}\right| = -3.136 + 0.103 m_{T'}. \tag{4.22}$$

It is important to note that as $\left|\overrightarrow{p_{T,\nu,Higgs}}\right|$ depends on $m_{T'}$, the $M_{T,T'}$ formula must be rederived for each mass point value of the T'.

$m_{T'}$ (GeV)	Mean $\cos(\nu_H, \nu_t)$	$\left|\sigma\left(\cos(\nu_H, \nu_t)\right)\right|$	Mean $\left|\overrightarrow{p_{T,\nu,Higgs}}\right|$ (GeV)	$\left|\sigma\left(\left|\overrightarrow{p_{T,\nu,Higgs}}\right|\right)\right|$ (GeV)
600	-0.487	0.483	59.40	48.61
625	-0.522	0.458	61.48	49.27
650	-0.522	0.458	63.51	50.50
675	-0.564	0.421	66.15	53.11
700	-0.572	0.415	68.42	55.51
800	-0.644	0.350	78.74	64.20
900	-0.703	0.295	89.87	71.28
1000	-0.744	0.256	100.7	80.71
1100	-0.777	0.223	110.1	90.04
1200	-0.807	0.193	120.2	97.03

Table 4.11: Mean value and standard deviation of $\cos(\nu_H, \nu_t)$ and $\left|\overrightarrow{p_{T,\nu,Higgs}}\right|$, computed using 2018 simulation per mass point value of the T'.

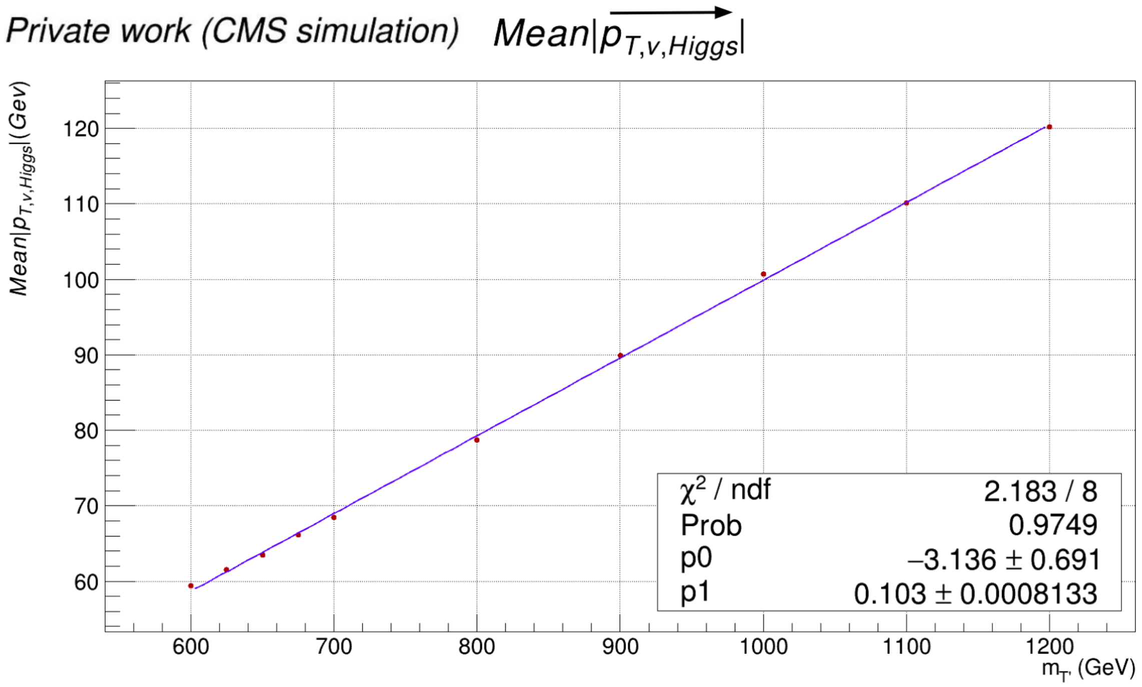


Figure 4.14: Mean $\left|\overrightarrow{p_{T,\nu,Higgs}}\right|$ distribution, computed using 2018 simulation. A first-degree polynomial in $m_{T'}$ is used to fit the distribution.

Secondly, from all the jets in the collection of selected jets passing the preselection, we need to select the 3 jets to be assigned as arising from the top quark and the Higgs boson decay, and which enter in the formula of $M_{T,T'}$.

- Three cases are considered to select the b-jet to be assigned to the top quark decay, as presented below. The case with the largest efficiency ϵ_{bjet} and purity P_{bjet}, as defined in equations 4.11 and 4.23, will be selected. The efficiency ϵ_{bjet} characterizes the number of selected b-jets matched to a b-quark arising from the top quark decay at the particle level, and the purity P_{bjet} the number of b-jets matched to a b-quark arising from the top quark decay at the particle level which are selected. These are the three cases.
  - $\min(\Delta R(b, l))$: the b-jet providing the minimal ΔR value with one of the two leptons.
  - The b-jet with the highest p_T.
  - the b-jet with the highest DeepJet score.

$$P = \frac{\text{Particle level + Offline selection}}{\text{Offline selection}} \tag{4.23}$$

The results are presented in Table 4.12, with 2018 simulated samples. The first case provides the highest values of ϵ_{bjet} and P_{bjet}, and will be used for the reconstruction of the top quark. The lepton coming from the top quark is selected to be the one used for the computation of $\min \Delta R(b, l)$, and the lepton coming from the Higgs boson is assigned to be the other selected lepton. To confirm this assignation, we can define a new efficiency ϵ_{lepton} and purity P_{lepton} for the leptons. The efficiency ϵ_{lepton} characterizes the number of selected leptons matched to the lepton arising from the top quark decay at the particle level, and the purity P_{lepton} the number of leptons matched to the lepton arising from the top quark decay at the particle level which are selected. The leptons are correctly assigned in most cases, with $\epsilon_{lepton} = 78.9\%$ and $P_{lepton} = 90.3\%$.

Selection of the b-jet	$\min(\Delta R(b, l))$	Highest p_T	Highest Deepjet score
ϵ_{bjet}	78.9%	77.5%	74.2%
P_{bjet}	93.5%	93.5%	87.1%

Table 4.12: Efficiency ϵ_{bjet} and purity P_{bjet} values for the assignation of the b-jet coming from the top quark, computed using 2018 simulation with $m_{T'} = 700$ GeV.

- Three cases are considered to select the jets (without considering the b-jet assigned to the top quark decay) to be assigned to the Higgs boson decay, as presented below. The case with the higher efficiency ϵ_{jets} and purity P_{jets} will be selected. The efficiency ϵ_{jets} characterizes the number of selected jets matched to a quark arising from the Higgs boson at the particle level, and the purity P_{jets} the number of jets matched to a quark arising from the Higgs boson at the particle level which are selected. These are the three cases.
  - $\min(\Delta R(j_1, j_2))$: the minimal ΔR value between two jets among jets permutations.
  - $\min(M(j_1 + j_2) - m_W)$: the invariant mass closest to the nominal W boson mass m_W for two jets among jet permutations.
  - $\min(M(j_1 + j_2))$: the invariant mass closest to 0 for two jets among jet permutations.

The results are presented in Table 4.13, with 2018 simulated samples. The values of ϵ_{jets} are lower here, which is expected as at least one of the two jets coming from the Higgs boson does not usually pass the p_T threshold at the analysis level. The first and third strategies have the largest values of ϵ_{jets} and P_{jets}. However, the W boson which decays hadronically is produced on- or off-shell in 50% of cases each. In other words, selecting the jets with the second (on-shell) or third (off-shell) case will be relevant in 50% of cases at maximum. For this reason, the first case will be used for the reconstruction of the Higgs boson.

Selection of the jets	$\min(\Delta R(j_1, j_2))$	$\min(M(j_1 + j_2) - m_W)$	$\min(M(j_1 + j_2))$
ϵ_{jets}	9.8%	7.6%	9.8%
P_{jets}	80.6%	64.5%	80.6%

Table 4.13: Efficiency ϵ_{jets} and purity P_{jets} values for the assignation of the jets coming from the Higgs boson, computed using 2018 simulation with $m_{T'} = 700$ GeV.

It is noteworthy that the selected b-jet assigned to the top quark and the selected jets assigned to the Higgs boson will be now used when computing the variable used in the cut 3 of the selection.

Finally, a decision needs to be made between the two cases proposed in the second hypothesis for calculating $M_{T,T'}$ ($\left|\overrightarrow{p_{T,\nu,Higgs}}\right| > \left|\overrightarrow{p_{T,\nu,top}}\right|$ or $\left|\overrightarrow{p_{T,\nu,Higgs}}\right| < \left|\overrightarrow{p_{T,\nu,top}}\right|$). These two cases are compared using the significance S defined earlier. The distributions of the $M_{T,T'}$ for the two strategies are shown in Figure 4.15. With 2018 simulated samples, the $\mu\mu$, eμ and ee channels combined and $m_{T'} = 700$ GeV, the significance is $S = 0.294\sigma$ for the first case, and $S = 0.282\sigma$ for the second case. Hence, the first case has a better signal discrimination than the second case and than the variable S_T^{lep}, and is selected for the definition of the discriminant observable $M_{T,T'}$ as follows:

$$M_{T,T'}=\sqrt{\left(E_{T,lep,Higgs}+E_{T,j1,Higgs}+E_{T,j2,Higgs}+E_{T,lep,top}+E_{T,b,top}+\sqrt{\overrightarrow{p_{T,MET}}^2+4\times\left|\overrightarrow{p_{T,\nu,Higgs}}\right|\times\left|\left|\overrightarrow{p_{T,\nu,Higgs}}\right|-\left|\overrightarrow{p_{T,MET}}\right|\right|}\right)^2-\left(\overrightarrow{p_{T,ljjlb}}+\overrightarrow{p_{T,MET}}\right)^2}. \tag{4.24}$$

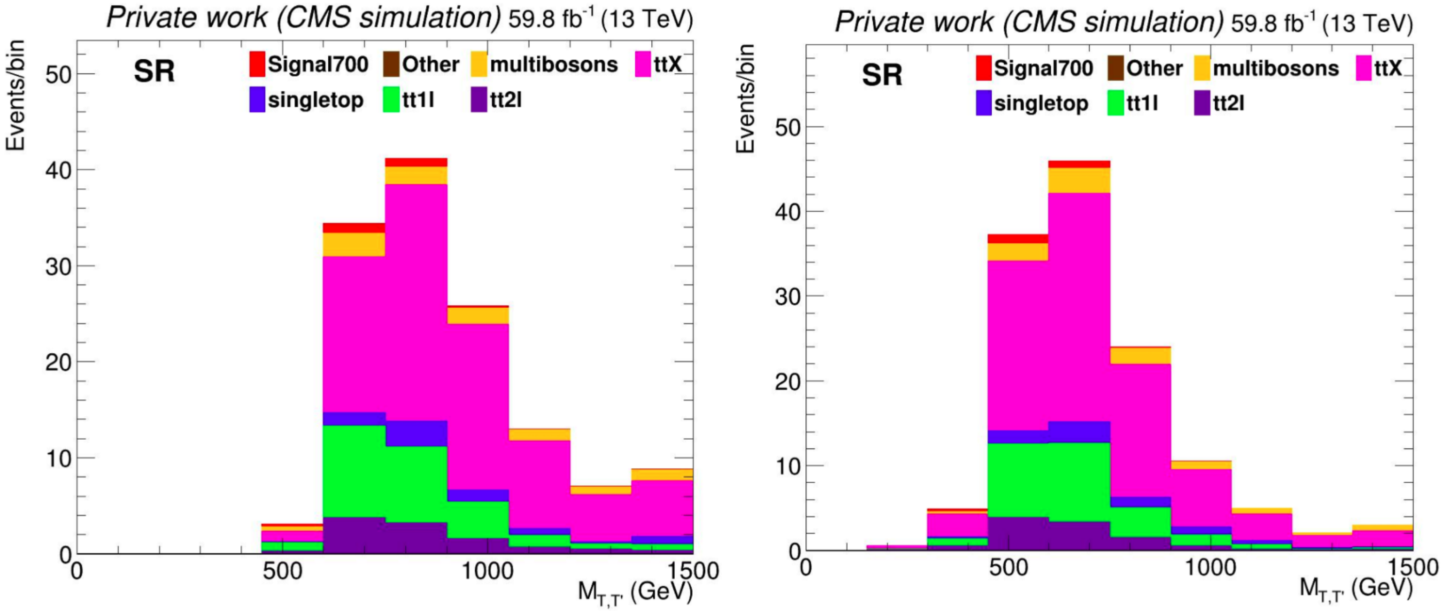


Figure 4.15: Distributions of $M_{T,T'}$ with the first (left) and second (right) case of the second hypothesis after the cut-based selection, computed using 2018 simulation with $m_{T'} = 700$ GeV.

4.3.4 Optimized selection using the getQuantiles method

The last optimization of the selection addresses a significant issue of the cut-based selection. If a given cut would sculpt the background shape of $M_{T,T'}$ in the SR so much that a bump would be created in the background distribution, it would be difficult to rely on a fit for the signal superimposed on this bump. Consequently, the background efficiency in each $M_{T,T'}$ bin must remain similar after each selection criterion to preserve the background shape after the full selection. This will be achieved by deriving each cut as a function of $M_{T,T'}$, using the 'getQuantiles' method in ROOT [296]. The procedure was first introduced in the all-hadronic resolved final state analysis [165], and is presented below [297, 298]. The cut 2 will serve as an illustrative example, with 2018 simulated samples and $m_{T'} = 700$ GeV. We are looking for an optimized selection criterion in the form of $p_{T,l_1} + p_{T,l_2} > f(M_{T,T'})$.

- Step 1: The discriminant observable $M_{T,T'}$ of the SR and the variable from which we want to define a selection criterion are plotted on the respective x- and y-axes of a 2D-histogram for background only. The preselection and the cut 1 have been applied. This is illustrated in Figure 4.16 by combining the $\mu\mu$, eμ, and ee channels.

- Step 2: A cut is applied on the variable under study for each bin on the x-axis in the 2D-histogram. The value of the cut may vary to keep the same given fraction, or quantile, of the background yield, for each bin. The quantile value is scanned in 5% increments in the background yield, such that the final value of 30% of the background yield is retained for each bin of $M_{T,T'}$. This approach maintains a relatively high signal efficiency, as presented in Table 4.16. The 2D-histogram is then converted into a 1D-histogram, as illustrated in Figure 4.17.

- Step 3: The distribution of the calculated value of the cut (chosen at a quantile of 30%) as a function of $M_{T,T'}$ is fitted with polynomial functions with given degrees of freedom to mitigate statistical fluctuations. The polynomial degree with the lowest χ^2/ndf is selected. This is represented by the blue curve in Figure 4.17, where the distribution is fitted with a second-degree polynomial in $M_{T,T'}$. The optimized cut 2 is finally defined as follows:
$$p_{T,l_1} + p_{T,l_2} > 6.468 + 0.3081 M_{T,T'} - 9.901 \times 10^{-5} M_{T,T'}^2. \tag{4.25}$$

The same procedure is repeated for each selection criterion. The distribution of $M_{T,T'}$ after the full selection is illustrated in Figure 4.20 with 2018 simulated events. The selection criteria are consistently applied for 2016, 2017, and 2018 to ensure uniformity across the years, as illustrated in Figure 4.21. Additional details regarding the optimized selection are provided below.

- The cut 1 does not require optimization, as illustrated in Figure 4.18. The 1D-histogram for the calculated value of $\Delta R(l_1, l_2)$ (chosen at a quantile of 30%) shows an almost flat distribution, and a constant value of $\Delta R(l_1, l_2) > 1.8$ will be retained to define the cut 1.

- The cut 3 requires optimization, as illustrated in Figure 4.19. The distribution is fitted with a fourth-degree polynomial in $M_{T,T'}$, such that the optimized cut 3 is defined as follows:

$$|M_{inv}(3j) - m_{top}| > 143 - 0.7486 M_{T,T'} + 1.583 \times 10^{-3} M_{T,T'}^2 - 1.324 \times 10^{-6} M_{T,T'}^3 + 3.835 \times 10^{-10} M_{T,T'}^4. \tag{4.26}$$

This procedure must be repeated for each mass point value of the T', since $\left|\overrightarrow{p_{T,\nu,Higgs}}\right|$ increases with $m_{T'}$ in the definition of $M_{T,T'}$. The results for the cuts 2 and 3 can be found for each mass point value of the T' in Appendix B.1.

4.3.5 Expected yield in the Signal Region

The number of selected signal and background events after each cut are presented in Table 4.14, 4.15 and 4.16 for 2016, 2017, and 2018 respectively. The $\mu\mu$, eμ, and ee channels are combined, and all the values are scaled to the integrated luminosity as presented in equation 4.2. The 'Others' column includes the $t\bar{t}0l$, DY, and $W + jets$ processes, and the signal efficiency ϵ_{signal} is defined in equation 4.11.

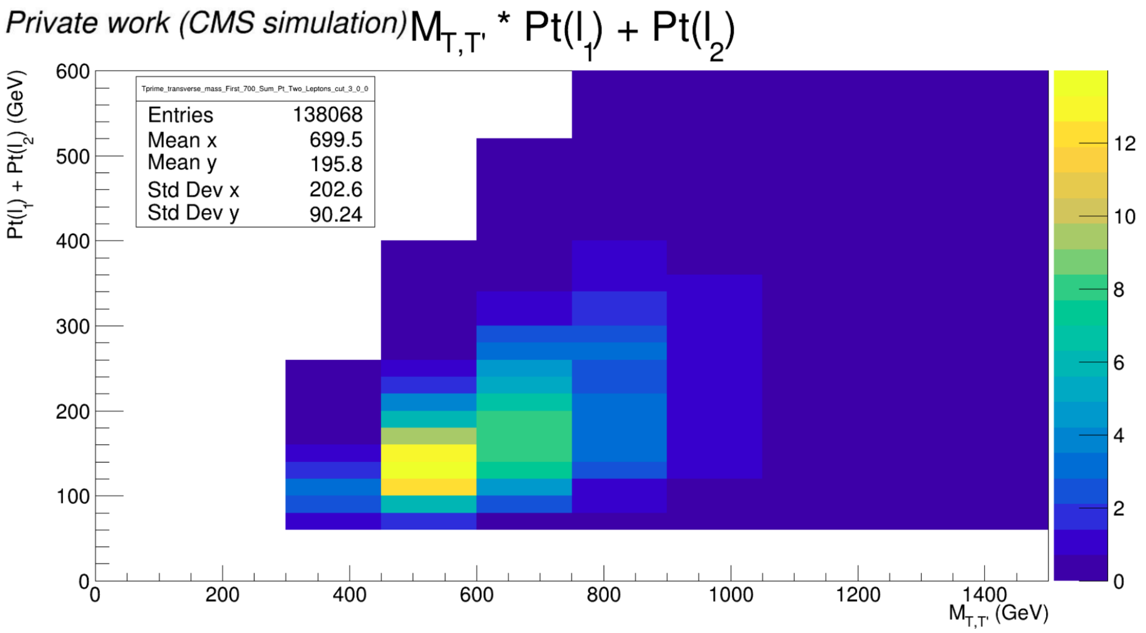


Figure 4.16: Distributions of $p_{T,l_1} + p_{T,l_2}$ vs $M_{T,T'}$ for the background processes, computed using 2018 simulation, with $m_{T'} = 700$ GeV and the $\mu\mu$, eμ and ee channels combined.

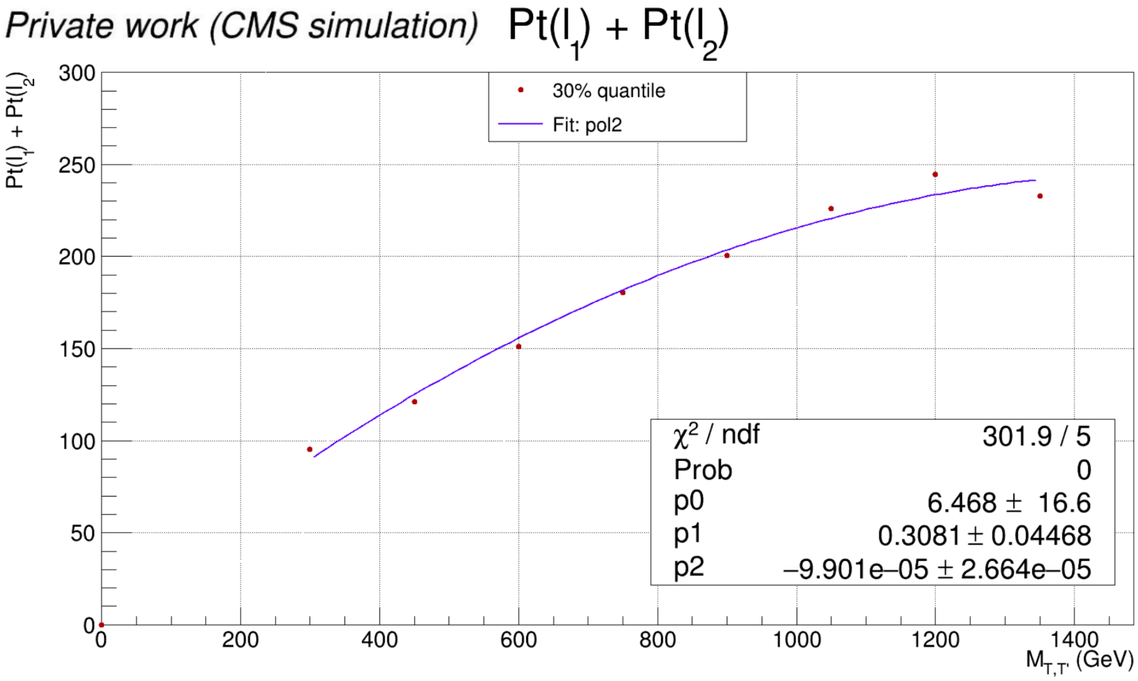


Figure 4.17: Distribution of the calculated value of $p_{T,l_1} + p_{T,l_2}$, chosen at a fraction of 30% of the background yield, as a function of $M_{T,T'}$ with $m_{T'} = 700$ GeV. A second-degree polynomial in $M_{T,T'}$ (blue) is used to fit the distribution (red dots).

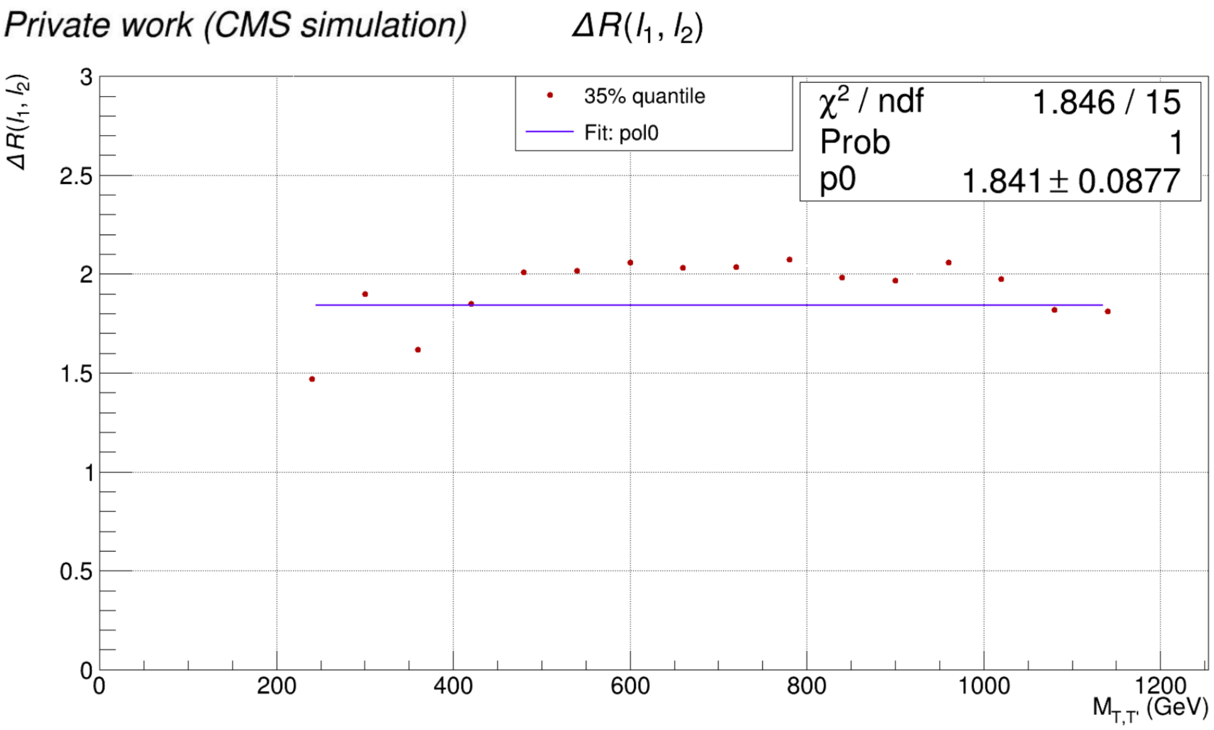


Figure 4.18: Distribution of the calculated value of $\Delta R(l_1, l_2)$, chosen at a fraction of 30% of the background yield, as a function of $M_{T,T'}$ with $m_{T'} = 700$ GeV. The distribution (blue) is flat in $M_{T,T'}$ around 1.8.

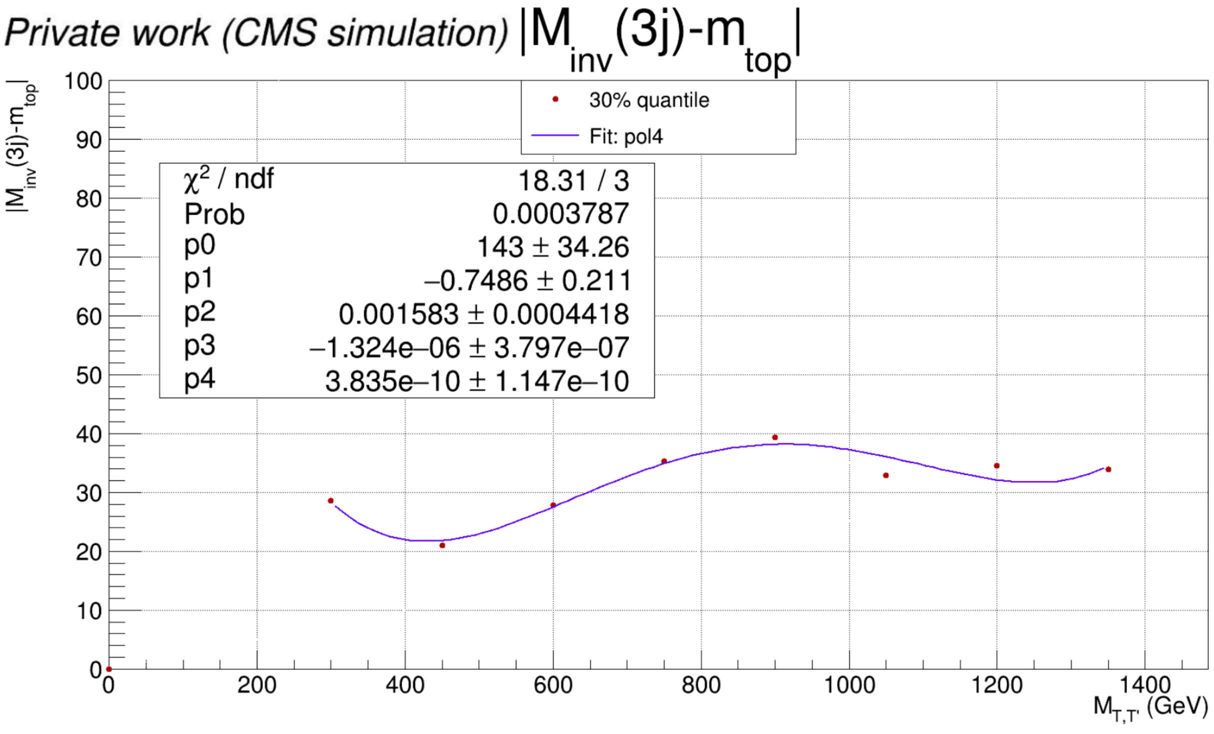


Figure 4.19: Distribution of the calculated value of $\left(|M_{inv}(3j) - m_{top}|\right)$, chosen at a fraction of 30% of the background yield, as a function of $M_{T,T'}$ with $m_{T'} = 700$ GeV. A fourth-degree polynomial in $M_{T,T'}$ (blue) is used to fit the distribution (red dots).

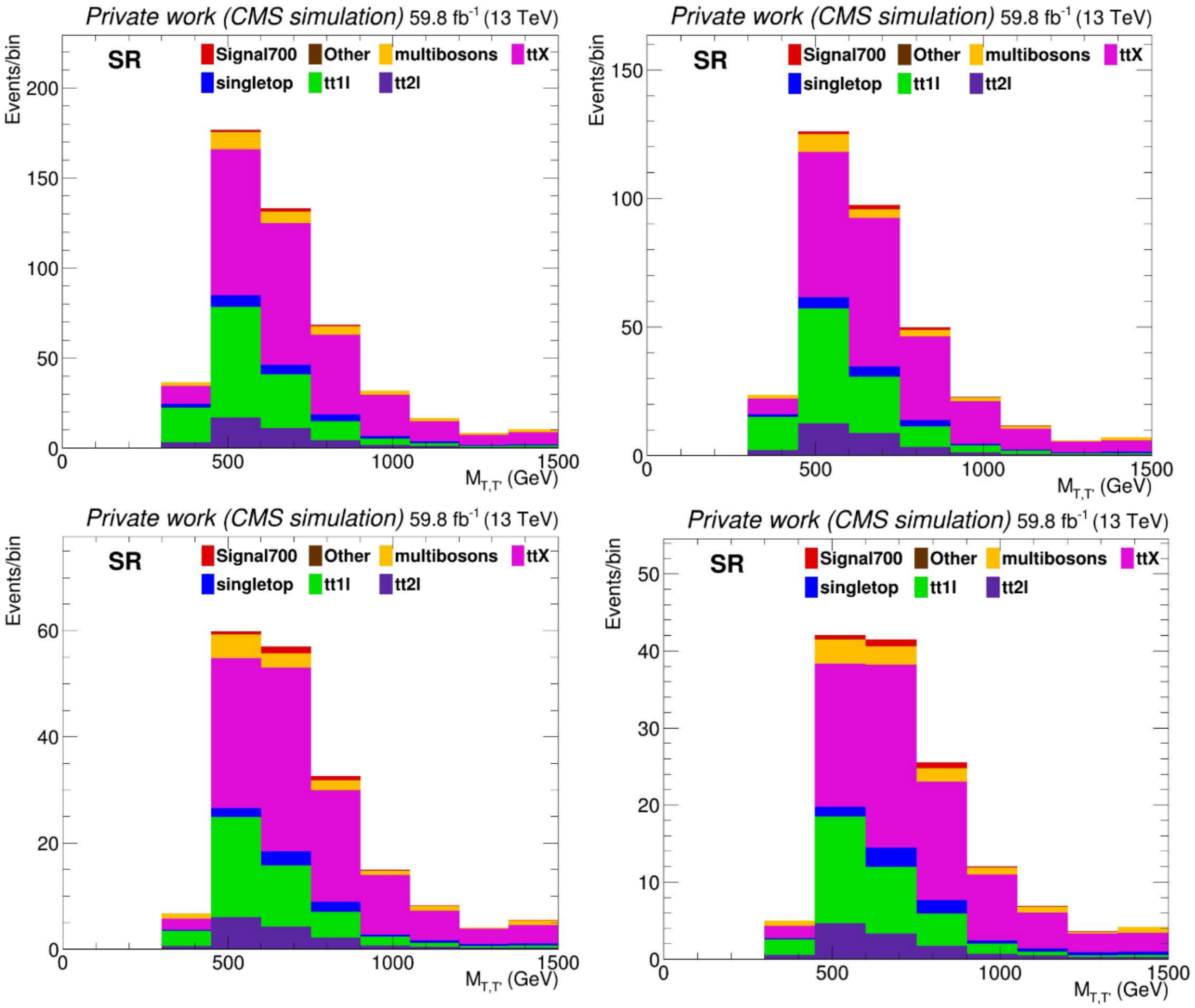


Figure 4.20: Distribution of $M_{T,T'}$ after the preselection (top left), the cut 1 (top right), the cut 2 (bottom left) and the cut 3 (bottom right) of the optimized selection, computed using 2018 simulation with $m_{T'} = 700$ GeV.

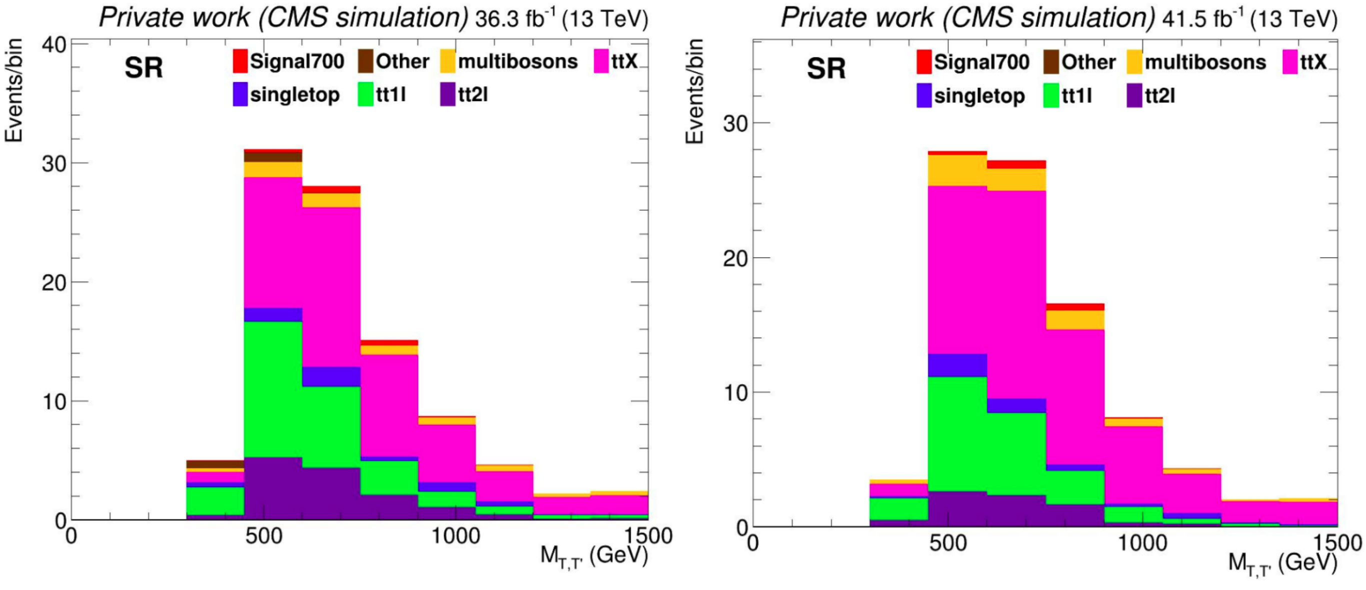


Figure 4.21: Distribution of $M_{T,T'}$ after the full selection, computed using 2016 (left) and 2017 (right) simulation with $m_{T'} = 700$ GeV.

After the full selection, the total background yield is reduced by at least 99.9%. The major remaining background processes are the irreducible $t\bar{t}$X processes, as expected. The $t\bar{t}$1l and $t\bar{t}$2l processes remain significant due to their large cross sections. Other background processes are minor or entirely suppressed. Although the signal yield is small after the full selection, ϵ_{signal} is better than after the cut-based selection, and considered acceptable. The signal-to-background ratio S/B is 1.47%, 1.63%, and 1.68% for 2016, 2017, and 2018 respectively after the full selection, and increases by a factor 2 relative to the preselection. The results for all the mass point values of the T' can be found in Appendix B.2. To improve the background yield estimation, a dedicated method using additional regions, using background CRs and background MRs, is required though. This will be the subject of the next chapter.

Cut	Signal	ϵ_{signal} (%)	$t\bar{t}$X	$t\bar{t}$1l	$t\bar{t}$2l	Multibosons	$tX(W/q)$	Others	S/B (%)
No cut	3.22×10^3	-	6.03×10^4	1.32×10^7	3.17×10^6	6.70×10^5	1.44×10^6	3.13×10^9	1.02×10^{-4}
Cut 0	2.72	13.8	146.06	101.02	41.57	15.79	14.82	3.98	0.83
Cut 1	2.56	12.9	104.29	73.93	33.05	9.36	9.88	2.23	1.09
Cut 2	1.65	7.6	61.65	32.51	17.02	6.34	6.18	1.57	1.30
Cut 3	1.43	6.4	44.24	25.82	13.81	5.28	4.86	1.57	1.47

Table 4.14: Number of events for the signal and background processes after each cut of the optimized selection in the SR, computed using 2016 simulation with $m_{T'} = 700$ GeV. The signal efficiency ϵ_{signal} is also provided.

Cut	Signal	ϵ_{signal} (%)	$t\bar{t}$X	$t\bar{t}$1l	$t\bar{t}$2l	Multibosons	$tX(W/q)$	Others	S/B (%)
No cut	3.68×10^3	-	6.89×10^4	1.51×10^7	3.62×10^6	7.66×10^5	1.65×10^6	3.58×10^9	1.02×10^{-4}
Cut 0	2.76	12.3	165.20	76.20	24.95	18.63	13.93	2.90	0.90
Cut 1	2.63	11.8	118.17	55.34	18.80	12.19	9.78	2.90	1.20
Cut 2	1.80	7.4	70.68	26.45	9.79	8.67	5.56	0.00	1.46
Cut 3	1.49	6.3	50.71	20.51	7.63	7.13	4.14	0.00	1.63

Table 4.15: Number of events for the signal and background processes after each cut of the optimized selection in the SR, computed using 2017 simulation with $m_{T'} = 700$ GeV. The signal efficiency ϵ_{signal} is also provided.

Cut	Signal	ϵ_{signal} (%)	$t\bar{t}$X	$t\bar{t}$1l	$t\bar{t}$2l	Multibosons	$tX(W/q)$	Others	S/B (%)
No cut	5.30×10^3	-	9.93×10^4	2.18×10^7	5.22×10^6	1.10×10^6	2.37×10^6	5.16×10^9	1.02×10^{-4}
Cut 0	4.28	11.8	260.57	127.39	38.89	29.57	20.94	0.19	0.89
Cut 1	4.17	11.6	186.06	92.17	28.66	18.24	14.05	0.07	1.21
Cut 2	2.81	7.9	109.02	41.00	14.53	12.82	8.30	0.07	1.49
Cut 3	2.36	7.0	77.28	31.04	11.65	10.63	7.49	0.07	1.68

Table 4.16: Number of events for the signal and background processes after each cut of the optimized selection in the SR, computed using 2018 simulation with $m_{T'} = 700$ GeV. The signal efficiency ϵ_{signal} is also provided.

Chapter 5

T' $\rightarrow$ tH same sign dilepton final state: background estimation and results

The previous chapter introduced the analysis strategy in the SR to discriminate the signature of the T' $\rightarrow Ht$ decay in the SS dilepton final state from the background processes. Although the signal is targeted mainly in the SR, additional regions must be defined to estimate the yield of every background process. These regions include CRs, meant to control the main background, and MRs, where the background distributions are estimated from data. By combining the SR and CR, the expected exclusion limit at 95% CL for the cross section is computed as a function of the nominal mass for the VLQ T' production. This represents the final result of the analysis before comparison with data.

This chapter will be divided into three sections. The first section will redefine the background with two new processes: the non prompt leptons, or 'fake' leptons, and the charge misidentified leptons, or 'flip' leptons. Along with the $t\bar{t}$X processes, they become the main background processes. The distributions for these processes will be measured in dedicated MRs, while a dedicated CR will be used to control the $t\bar{t}$X processes. The second section will evaluate the impact of the uncertainties, treated as nuisance parameters in the fit. This evaluation will include all systematic and statistical uncertainties included in the signal extraction. The last section will focus on the expected results. The measure of the inclusive signal strength r and its expected exclusion limit at 95% CL as a function of the nominal mass of the VLQ T' will be discussed, and compared to other T' $\rightarrow Ht$ final states presented in Chapter 4. Possible improvements and extensions of this analysis will conclude the chapter.

5.1 Background estimation

For a reducible background process ($t\bar{t}$, multiboson, or $tX(W/q)$ process) to pass the selection criteria discussed in the previous chapter, at least one of the two leptons must be misidentified. This misidentification may happen in every background sample, necessitating a redefinition of the background processes.

5.1.1 Regrouping background processes

Some of the background events which enter the SR can be categorized into two types.

- The 'fake leptons' are non prompts leptons, meaning they do not originate from the primary collision in the event. They typically arise from heavy-flavor hadron decays, muons from light meson decays, and electrons from unidentified photon conversions. The largest contributor to this category is the $t\bar{t}$1l process (63.4% of the fake leptons in the SR, with a contribution of the $t\bar{t}$X, $t\bar{t}$2l, $tX(W/q)$, and multiboson processes of 16.9%, 10.1%, 7.6%, and 2.0% respectively). Events with fake leptons will be referred to as 'non prompt' events.

- The 'flip leptons' are leptons whose charge has been misidentified. This misidentification can occur, for example, in cases where OS leptons are incorrectly identified as SS due to bremsstrahlung in the detector material. Since the muon charge is measured with high efficiency in both the tracker and muon chambers, this background is negligible for muons, and is only considered for electrons. The largest contributor to this category is the $t\bar{t}$2l process (98.8% of the flip leptons in the SR). Events with flip leptons will be referred to as 'charge flip' events.

For the rest of the chapter, events identified as non prompt or charge flip events will be removed from other processes, while the two new categories of flip and fake leptons are introduced. All other events, which rarely produce two real SS leptons, will be referred to as 'other' events.

The distribution of $M_{T,T'}$ with the regrouped background processes is presented in the SR in Figure 5.1, with 2018 simulated samples. The $t\bar{t}$X processes remain the major background processes, accounting for 48.3% of the total background yield. The non prompt and charge flip processes constitute the second and third largest background processes, accounting for 33.0% and 7.6% of the total background yield respectively. The regrouping of the background processes is therefore justified, necessitating the establishment of additional regions to accurately estimate the background distributions for the $t\bar{t}$X, non prompt, and charge flip processes. In contrast, the contribution from the multiboson and other events are considered negligible (5.7% and 2.7% of the total background yield respectively), and will be estimated from simulation.

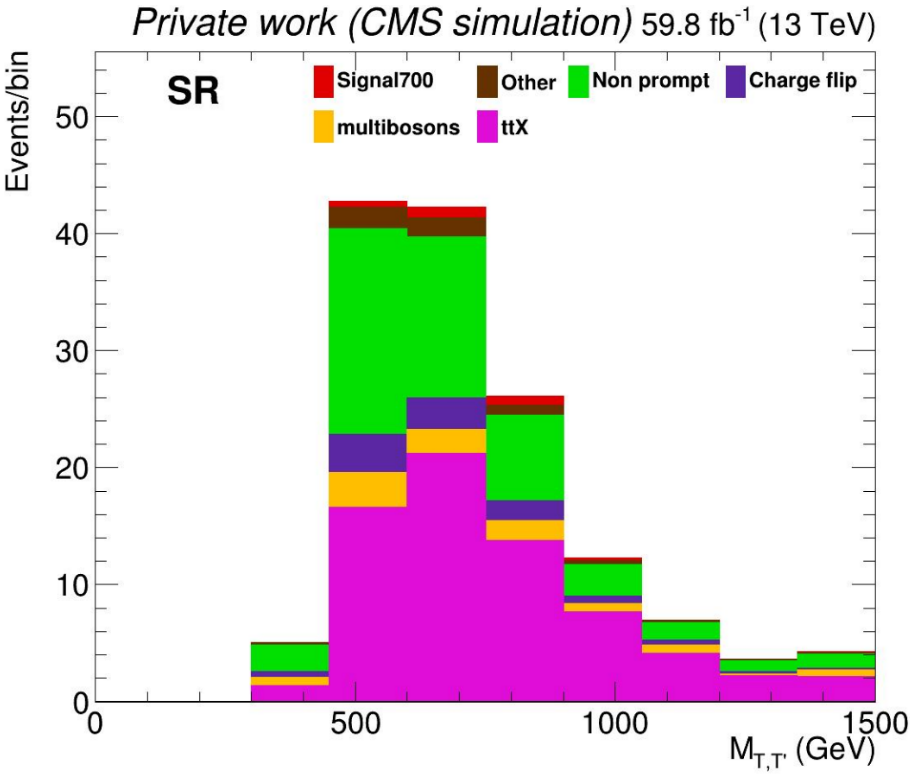


Figure 5.1: Distribution of $M_{T,T'}$ in the SR after the regrouping of the background processes, computed using 2018 simulation with $m_{T'} = 700$ GeV.

5.1.2 Control and Measurement regions

All the CRs and MRs are defined such that the targeted process gets the highest event yield among all processes in the related region. This procedure must be done for every background process which is not negligible. Each CR allows a large constraining power in the SR by combining information of the process distributions in both regions. Conversely, each MR allows to decrease the statistical uncertainty and improve the modeling by replacing the targeted process in the SR with the data distribution in the MR (see section 5.1.3). Using this data-driven approach avoids relying on simulation to estimate those processes where simulation is known to have limitations (see section 5.2.2). The targeted process must be almost pure in the MR for this method to be effective.

The ttX Control Region

The ttX region targets the $t\bar{t}$X processes, where a top quark pair is produced in association with a massive boson, with one of the two top quarks decaying hadronically. The cut 3 of the selection is reverted to require an hadronic top quark decay, and the rest of the selection is preserved. The comparison between data and MC is presented in Figure 5.2 for the $M_{T,T'}$ variable, with 2018 samples and $m_{T'}$ = 700 GeV. The $t\bar{t}$X processes account for the majority of the background yield (58.1% of the total yield), with a signal contamination of 0.9%. But other background processes are not negligible, particularly the non prompt process (29.4% of the total yield). Moreover, the MC underestimates data by roughly 40%, as already observed in dedicated $t\bar{t}X$ analyses [299–301]. There are indeed low statistics in this region, and higher-order electroweak corrections are not taken into account in the NLO samples, leading to events not accounted for in the sample (see section 5.2.2). The results are similar for all the years and mass point values of the T', and can be found in Appendix B.3.1. Thus, the distribution of the $t\bar{t}$X processes will be estimated from simulation, with a large normalization uncertainty of 40%, and the ttX region is used as a CR.

The tt2l Measurement Region

The tt2l region targets the charge flip process, where a top quark pair is produced with both top quarks decaying leptonically. The cut 0 of the selection is modified to require 2 OS leptons, and the rest of the selection is preserved. The comparison between data and MC is presented in Figure 5.3 for the $M_{T,T'}$ variable, with 2018 samples and $m_{T'}$ = 700 GeV. Since charge flip events are mostly composed of $t\bar{t}$2l events in the SR and ttX CR, the tt2l region is used to replace the charge flip process. Furthermore, the DY and $tX(W/q)$ processes are identified as distinct background processes here as they are not negligible in this region. On the contrary, multiboson processes are categorized as 'other' events as they are negligible in this region. The $t\bar{t}$2l process accounts for the majority of the background yield (85.5% of the total yield), with a signal contamination of 0.009%. Other background processes, like the non prompt and $tX(W/q)$ processes, are small (6.6% and 5.0% of the total yield respectively). Data agrees with MC, within data statistical uncertainties. The results are similar for all the years and mass point values of the T', and can be found in Appendix B.3.2. Thus, the distribution of the charge flip process will be estimated from data, and the tt2l region is used as a MR.

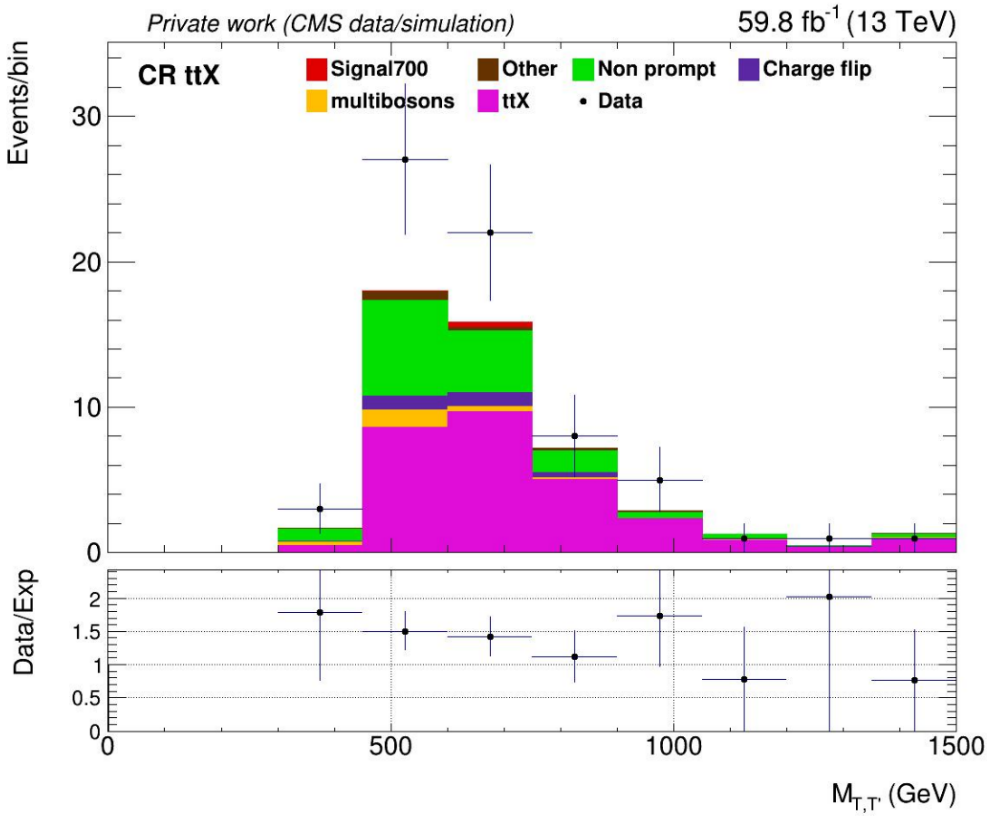


Figure 5.2: Distribution of $M_{T,T'}$ in the ttX CR, comparing 2018 simulation and data with $m_{T'} = 700$ GeV. Data is represented by the black points with statistical error bars.

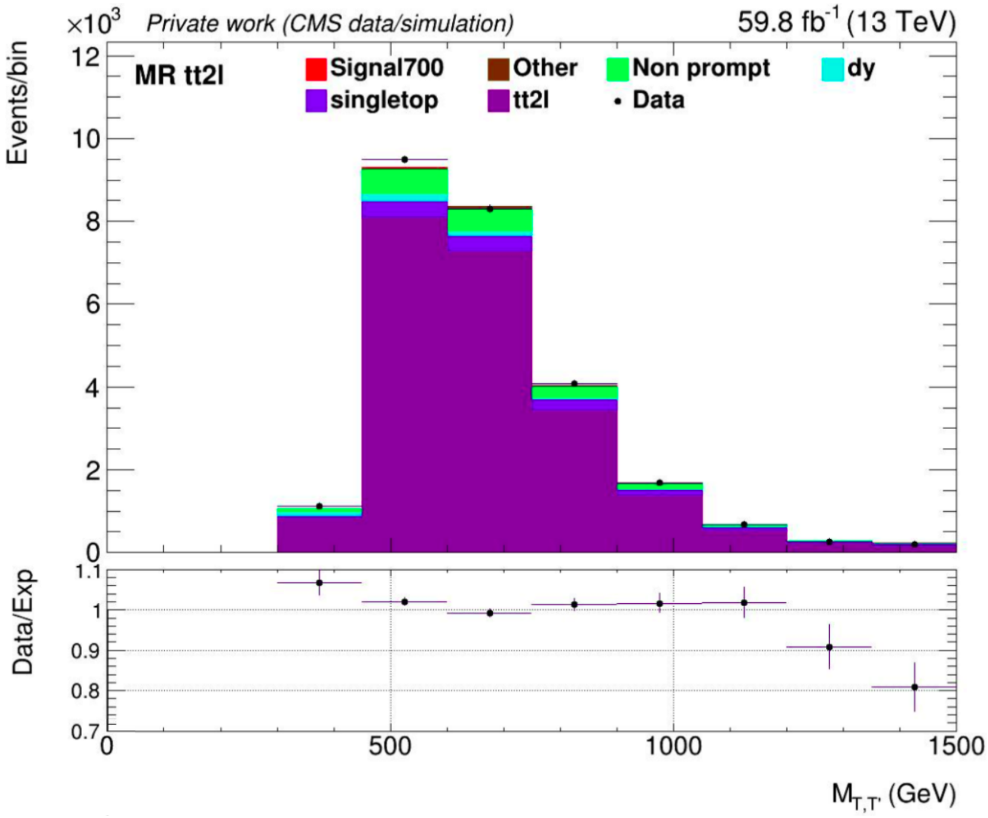


Figure 5.3: Distribution of $M_{T,T'}$ in the tt2l MR, comparing 2018 simulation and data with $m_{T'} = 700$ GeV. Data is represented by the black points with statistical error bars.

The tt1l Measurement Regions

The tt1l region targets the non prompt process, where a top quark pair is produced, with one of the two top quarks decaying hadronically and the other one leptonically. The cut 0 of the selection is modified to still require 2 SS leptons, except that one lepton must pass the tight requirements, and one lepton must pass the fakeable requirements but fail the tight requirements (see section 4.2.2). Events with two fakeable leptons are rejected. As we will see in section 5.1.3, two regions must be defined here because the distribution of the non prompt process is different in the SR and ttX CR. The first one preserves the rest of the selection, referred to as the tt1l region of the SR. The second one reverts the cut 3 of the selection to require an hadronic top quark decay, referred to as the tt1l region of the ttX CR. The comparison between data and MC is presented in Figures 5.4 and 5.5 for the $M_{T,T'}$ variable in both regions, with 2018 samples and $m_{T'}$ = 700 GeV. Multiboson processes are categorized as 'other' events as they are negligible in this region. The non prompt process accounts for the majority of the background yield (94.9% and 93.9% of the total yields), with a signal contamination of 0.2% and 0.04%. Other background processes, like the $t\bar{t}$X processes, are small (3.4% and 4.7% of the total yields). Data agrees with MC, within data statistical uncertainties. The results are similar for all the years and mass point values of the T', and can be found in Appendix B.3.3. Thus, the distribution of the non prompt process will be estimated from data, and the tt1l regions are used as MRs.

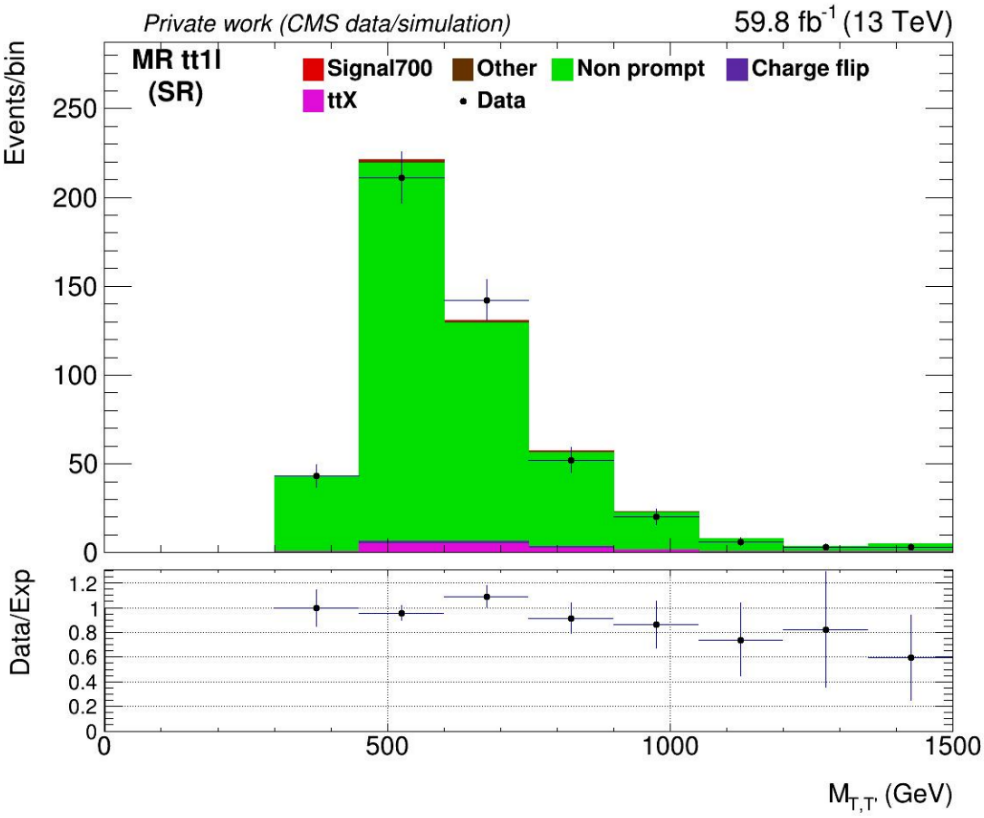


Figure 5.4: Distribution of $M_{T,T'}$ in the tt1l MR of the SR, comparing 2018 simulation and data with $m_{T'}$ = 700 GeV. Data is represented by the black points with statistical error bars.

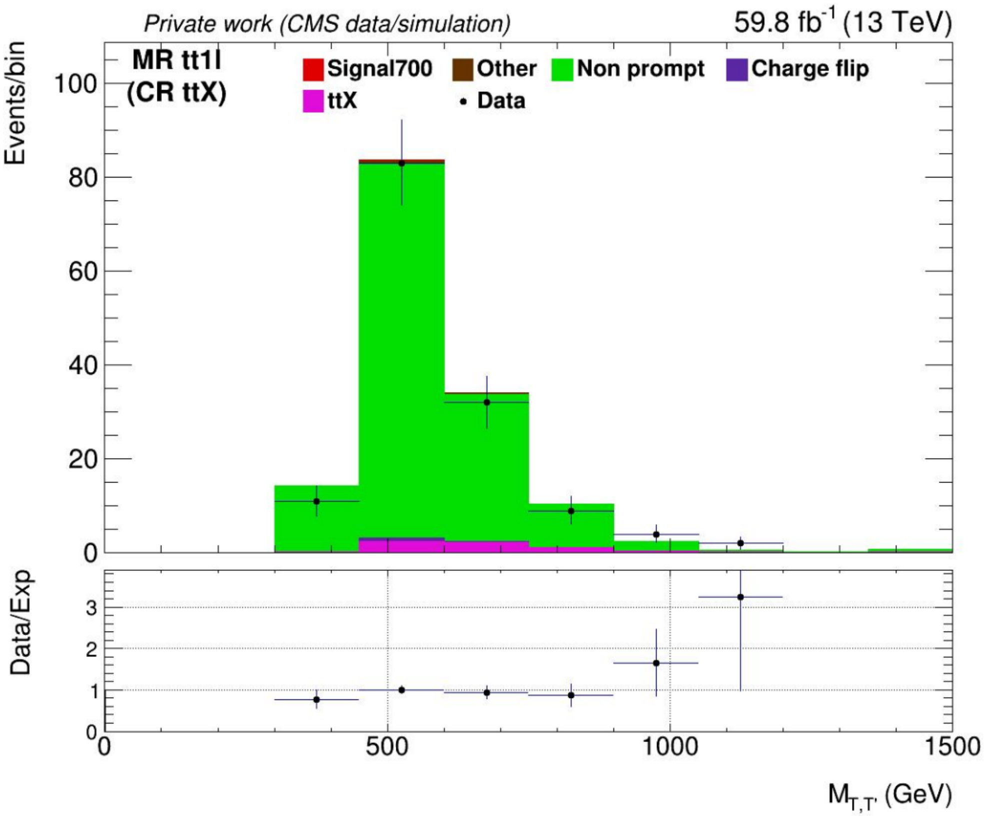


Figure 5.5: Distribution of $M_{T,T'}$ in the tt1l MR of the ttX CR, comparing 2018 simulation and data with $m_{T'} = 700$ GeV. Data is represented by the black points with statistical error bars.

5.1.3 Estimate of the charge flip and non prompt background distributions from the Measurement Regions

The MRs discussed in the previous section show a good agreement between data and MC. Since the charge flip and non prompt processes are dominating in their respective MR, those regions can be used to estimate the charge flip and non prompt distributions in the SR and ttX CR. To ensure that the overall quantity of events is preserved while modifying only the shape, the data normalization will be set to the expected number of events in the MC for the process in the respective SR or CR.

Estimate of the charge flip distribution from the tt2l Measurement Region

In the SR and ttX CR, charge flip events mostly arise from the $t\bar{t}$2l process. Therefore, to accurately estimate the relevant data distribution, contributions from processes other than the $t\bar{t}$2l process are subtracted in the tt2l MR, as illustrated in Figure 5.6. The normalized distribution of the charge flip process is compared in Figure 5.7 for the $M_{T,T'}$ variable in the SR, ttX CR, and tt2l MR, with 2018 samples and $m_{T'} = 700$ GeV, along with the corresponding data distribution (after subtraction) in the tt2l MR. The distributions agree within data statistical uncertainties, which justifies using the background-subtracted data distribution from the tt2l MR, referred to as the data-driven charge flip distribution, to replace the charge flip distribution in the SR and ttX CR.

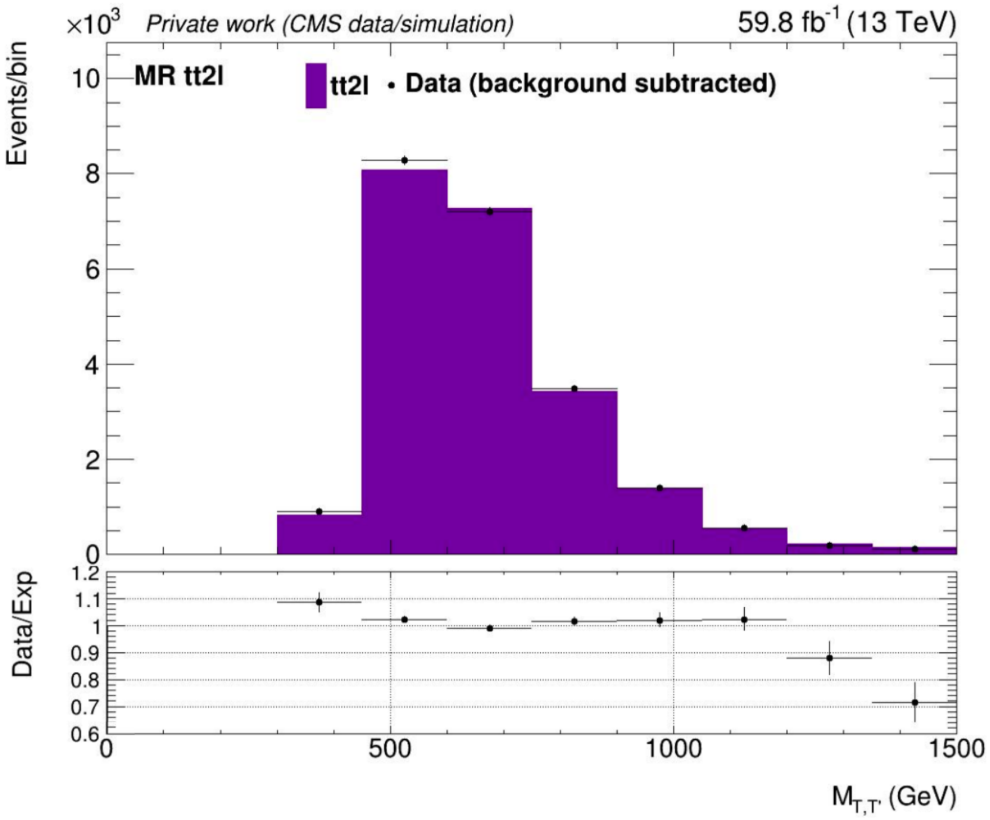


Figure 5.6: Distribution of $M_{T,T'}$ in the tt2l MR after the subtraction of processes other than the $t\bar{t}$2l process, comparing 2018 $t\bar{t}$2l simulation and data. The data-driven charge flip distribution is represented by the black points with statistical error bars.

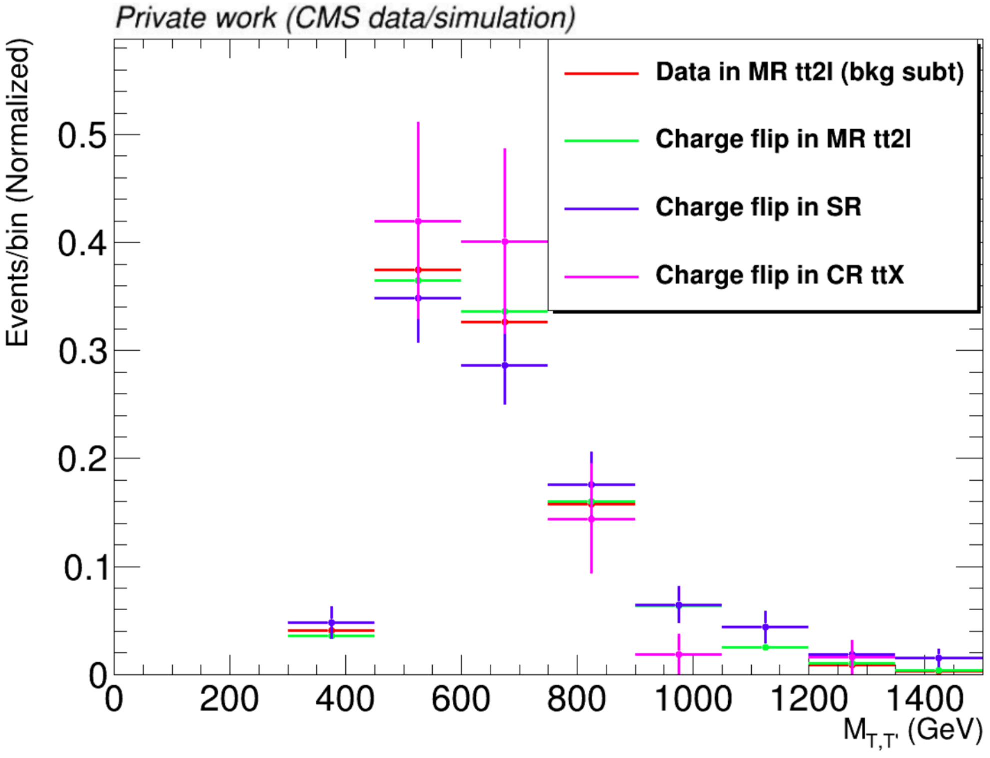


Figure 5.7: Normalized distributions of $M_{T,T'}$ of data in the tt2l MR (red), and of the charge flip process in the tt2l MR (green), SR (blue) and ttX CR (magenta), comparing 2018 simulation and data.

Estimate of the non prompt distribution from the tt1l Measurement Regions

In the SR and ttX CR, to accurately estimate the relevant data distribution, contributions from processes other than the non prompt process are subtracted in the tt1l MRs, as illustrated in Figures 5.8 and 5.9. The normalized distribution of the non prompt process is compared in Figure 5.10 for the $M_{T,T'}$ variable in the SR and tt1l MR of the SR, with 2018 samples and $m_{T'} = 700$ GeV, along with the corresponding data distribution (after subtraction) in the tt1l MR of the SR. Similarly, the normalized distribution of the non prompt process is compared in Figure 5.11 in the ttX CR and tt1l MR of the ttX CR, with 2018 samples and $m_{T'} = 700$ GeV, along with the corresponding data distribution (after subtraction) in the tt1l MR of the ttX CR. The distributions do not fully agree within data statistical uncertainties, and the difference will be used as shape uncertainties (see section 5.2.2). Additionally, the discrepancies between the non prompt distributions in the SR (blue curve in Figure 5.10) and the ttX CR (magenta curve in Figure 5.11) justify the definition of two distinct tt1l MRs. The background-subtracted data distribution in each tt1l MR, referred to as the data-driven non prompt distribution, is then used to replace the non prompt distribution in their respective region.

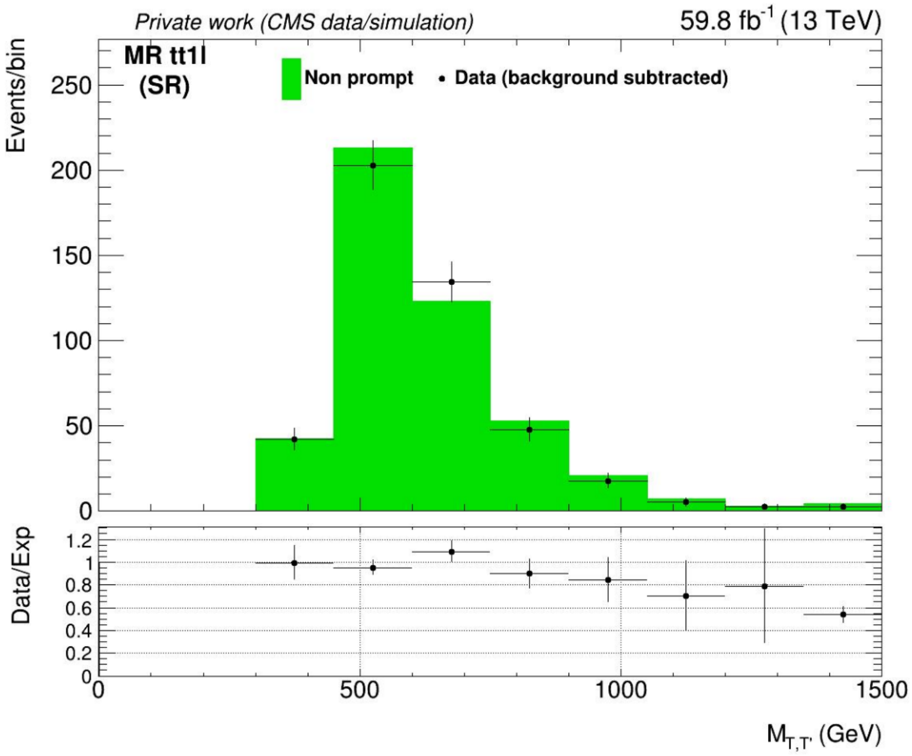


Figure 5.8: Distribution of $M_{T,T'}$ in the tt1l MR of the SR after the subtraction of processes other than the non prompt process, comparing 2018 simulation and data. The data-driven non prompt distribution is represented by the black points with statistical error bars.

Distributions of the transverse mass using data-driven background estimates in the Signal Region and ttX Control Region

The distribution of $M_{T,T'}$ after the data-driven estimation of the charge flip and non prompt contributions is presented in the SR in Figure 5.12. While the expected number of events for both processes is taken from simulation in this region, the shape of the distribution is modified, particularly for low-$M_{T,T'}$ non prompt events. This behavior is expected since non prompt events generally exhibit lower $M_{T,T'}$ in the tt1l MR of the SR, as illustrated in Figure 5.10. Similar results are observed in the ttX CR, as presented in Figure 5.13. The data-driven estimation of the charge flip and non prompt processes results in a maximum improvement of

20% in the agreement between data and MC in the ttX CR. Thus, this method is expected to provide an accurate estimation of the background distribution in the SR.

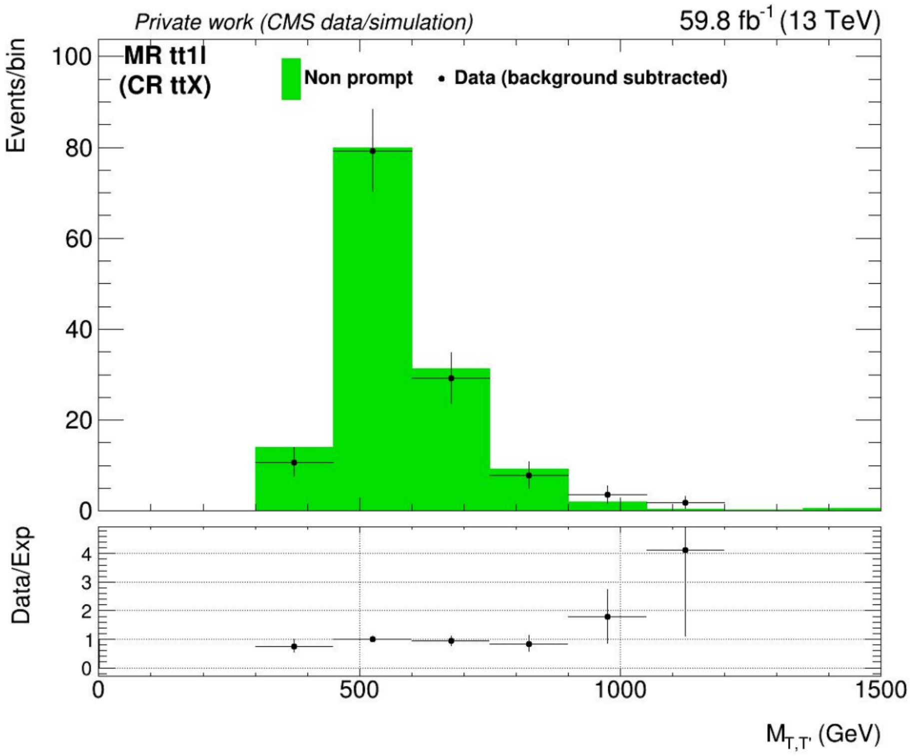


Figure 5.9: Distribution of $M_{T,T'}$ in the tt1l MR of the ttX CR after the subtraction of processes other than the non prompt process, comparing 2018 simulation and data. The data-driven non prompt distribution is represented by the black points with statistical error bars.

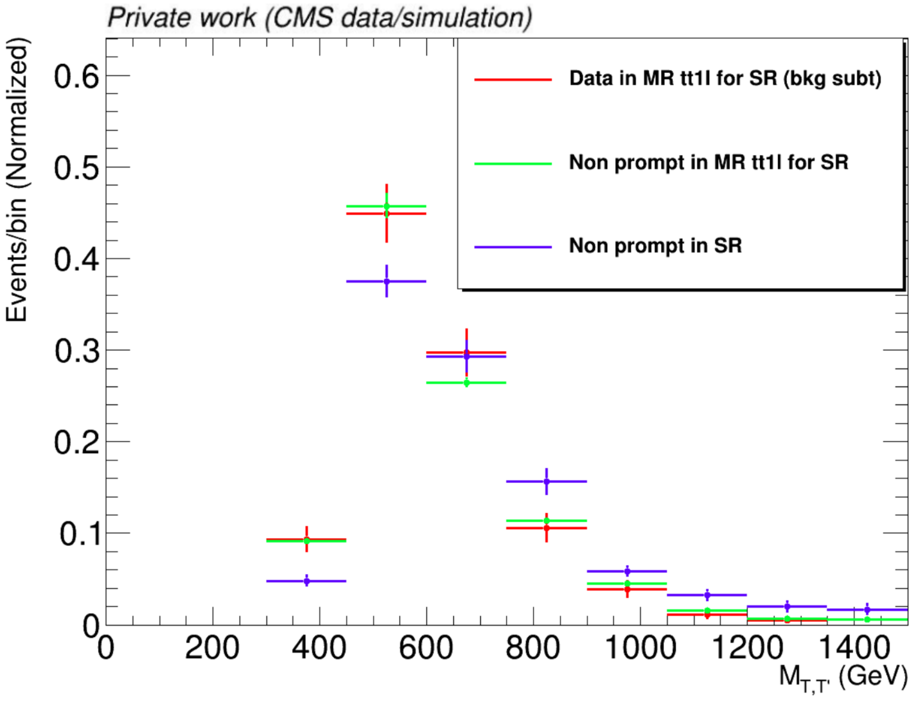


Figure 5.10: Normalized distributions of $M_{T,T'}$ of data in the tt1l MR of the SR (red), and of the non prompt process in the tt1l MR of the SR (green) and SR (blue), comparing 2018 simulation and data.

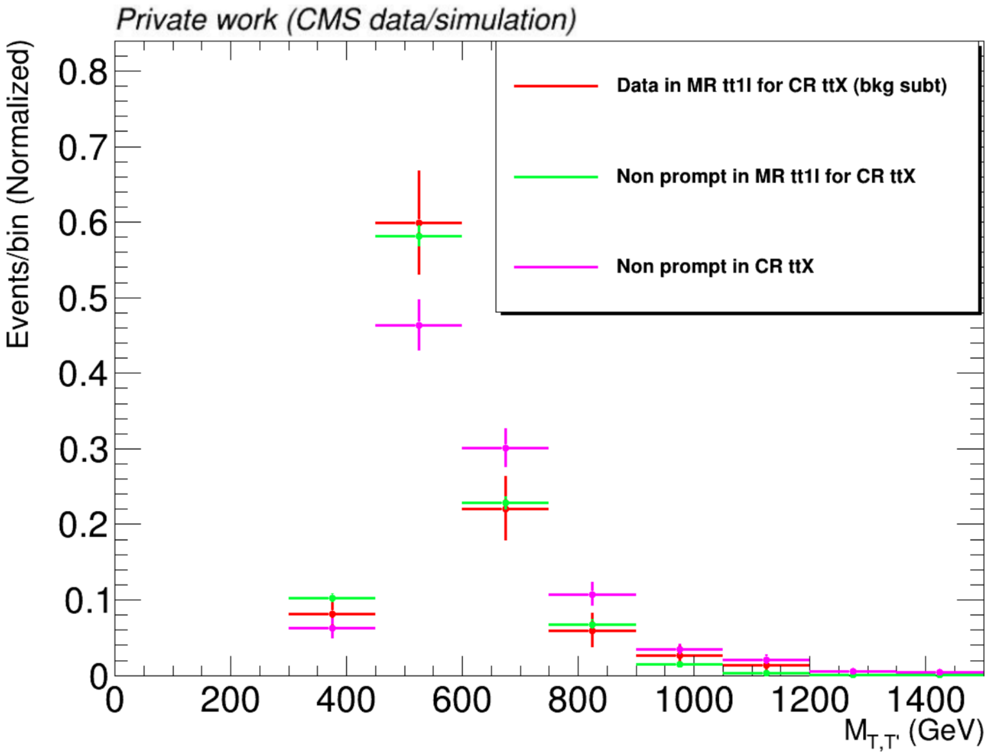


Figure 5.11: Normalized distributions of $M_{T,T'}$ of data in the tt1l MR of the ttX CR (red), and of the non prompt process in the tt1l MR of the ttX CR (green) and ttX CR (magenta), comparing 2018 simulation and data.

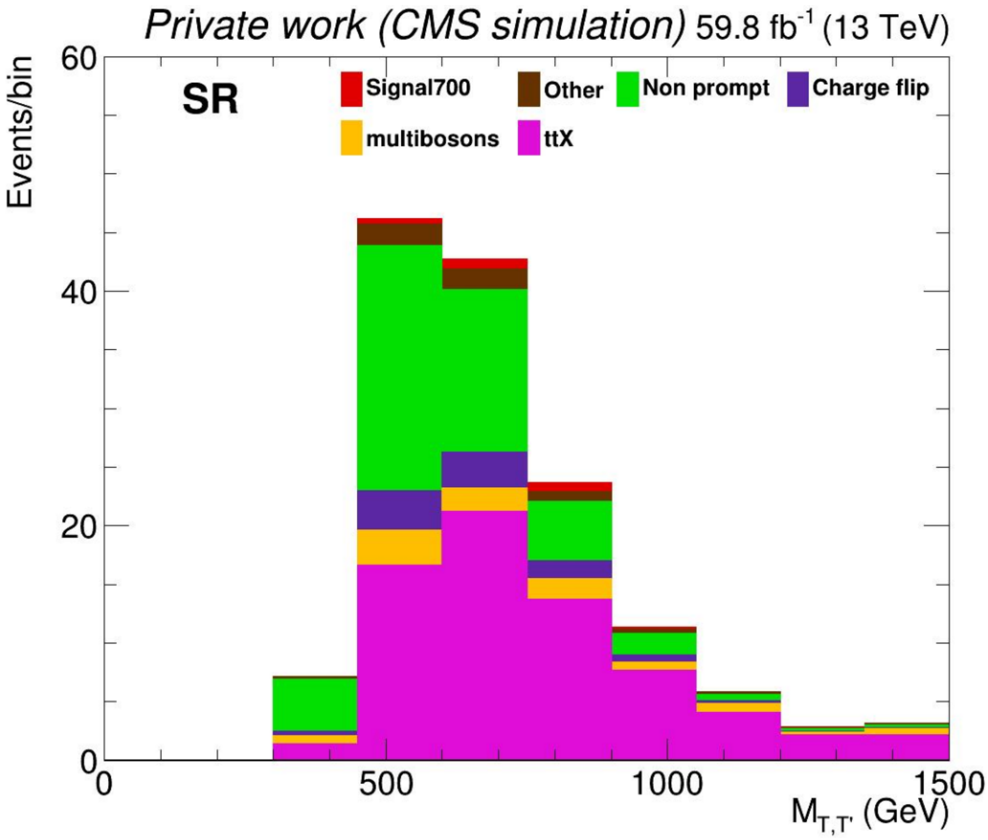


Figure 5.12: Distribution of $M_{T,T'}$ in the SR after the data-driven estimation of the charge flip and non prompt processes, computed using 2018 simulation with $m_{T'} = 700$ GeV.

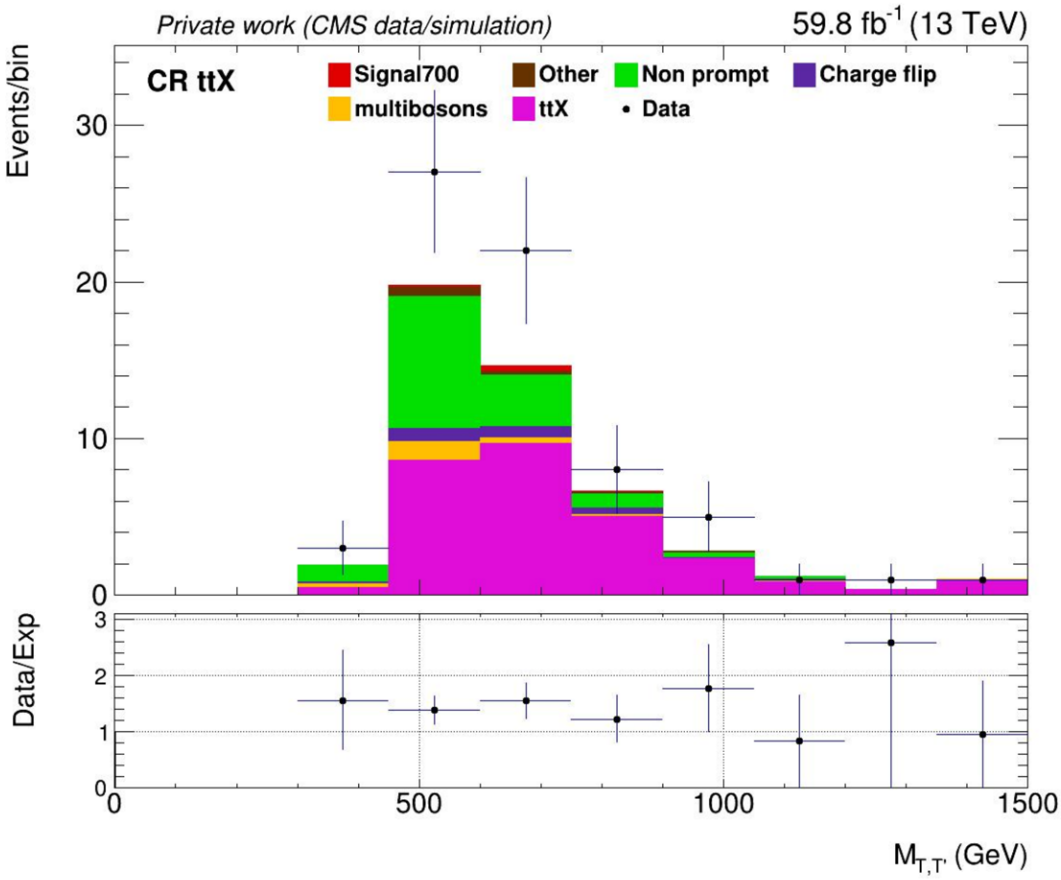


Figure 5.13: Distribution of $M_{T,T'}$ in the ttX CR after the data-driven estimation of the charge flip and non prompt processes, comparing 2018 simulation and data with $m_{T'} = 700$ GeV. Data is represented by the black points with statistical error bars.

5.2 Uncertainties

The definition of the CR and MRs is expected to provide an improved modeling of the background distribution in the SR. For a precise determination of the signal strength, it is essential to combine both the SR and ttX CR in the fit while accounting for all the systematic and statistical uncertainties.

5.2.1 Other event corrections

Despite the corrections applied to the objects to prevent any miscalibration in the CMS detector (see section 4.2.3), additional corrections are needed, regarding data taking or the modeling of some processes.

The Level 1 trigger ECAL prefiring

In 2016 and 2017 data, the gradual timing shift in the ECAL endcap was not properly propagated to the L1 TPs. This resulted in a significant fraction of high-$|\eta|$ TPs being mistakenly associated to the previous bunch crossing. A correction for this effect is provided centrally, and applied as event weights [302].

Top quark p_T reweighting

In many $t\bar{t}$ analyses at the LHC, it has been observed that the top quark p_T spectrum in data is softer than the top quark p_T spectrum predicted in MC. To correct for this effect, several schemes were proposed [303]. We are correcting events according to the top quarks p_T evaluated at parton level, after radiation and before decay. The individual top weight is a SF corresponding to the ratio of $t\bar{t}$ unfolded data over Powheg NLO predictions, as recommended by the CMS collaboration to improve the modelling of the detector response, and as follows:

$$SF(p_T) = \mathrm{e}^{0.0615-0.0005\times p_T}. \tag{5.1}$$

The overall event weight is calculated as: $SF = \sqrt{SF(p_{T,top})^2 + SF(p_{T,antitop})^2}$. In this analysis, we are never measuring the top quark p_T: thus, such a reweighting of top quark p_T is a relevant correction to be applied, in order to model the $t\bar{t}$ process to the best of our knowledge. It is to note that since charge flip and non prompt processes are estimated from data, the top quark p_T reweighting of MC $t\bar{t}$ events only affects 1.1% of the background yield in both the SR and ttX CR.

Charge flip process reweighting

An additional SF is defined for charge flip events to account for the differences between data and simulation, as documented in Ref. [304]. A dedicated CR requires exactly two tight OS leptons in the eμ and ee channels, which are identified the same way as in section 4.2.2, and where the $Z \to ee$ peak is reconstructed via a tag and probe method. The SF values are then measured from the data over MC ratio of this resonance, as presented per year in Table 5.1, and applied as weights to charge flip events.

Year	Flip lepton SF
2016	1.11
2017	1.44
2018	1.40

Table 5.1: Corrections applied on MC charge flip events rates per year.

5.2.2 Systematic uncertainties

The impacts of many sources of systematic uncertainties on the signal and SM backgrounds are evaluated. The $M_{T,T'}$ distributions are redetermined by computing the effect of each uncertainty up and down, and evaluating the yield in each bin of the discriminant observable for each process. This is performed for each region. The uncertainty is reported as the ratio between the variation and nominal yields. The systematic uncertainties (rate or shape) are included in the signal extraction as nuisance parameters in the likelihood fit. The correlation between the regions, years, and processes is summarized for each source of systematic uncertainty in Table 5.4.

Experimental uncertainties

All the experimental uncertainties are applied in a correlated way to the signal and background processes, in all the regions. Unless specified otherwise, these uncertainties do not apply on the charge flip and non prompt processes since they are estimated from data. Except for the Jet

Energy Scale (JES), the shape uncertainties introduce a variation of the distributions by 15% at maximum, as illustrated for the signal with $m_{T'} = 700$ GeV in Appendix C.1.

- Luminosity (rate): this uncertainty accounts for the imperfect knowledge of the luminosity recorded by CMS in each year. The recipe used to apply this uncertainty uses the recommended correlations for Run 2 from the CMS Luminosity group [305]. The CMS Luminosity group recommends implementing five total uncertainties: three uncorrelated uncertainties for each year, and two more which are correlated for 2016, 2017, and 2018, and for 2017 and 2018. These values are reported in Table 5.2.

Year	2016	2017	2018
Uncorrelated 2016	1.0	0.0	0.0
Uncorrelated 2017	0.0	2.0	0.0
Uncorrelated 2018	0.0	0.0	1.5
Correlated 2017-2018	0.0	0.6	0.2
Correlated 2016-2018	0.6	0.9	2.0

Table 5.2: Uncertainty values for the luminosity per year (uncorrelated) and between years (correlated).

- Pileup (rate + shape): simulation is reweighted to match the expected data distribution for the number of collisions per event. The CMS Luminosity group recommends this uncertainty to be fully correlated for 2016, 2017, and 2018, with a value of the total inelastic cross section varied by 4.6% [306].
- Level 1 Trigger ECAL prefiring (rate + shape): the SFs are used to correct for the prefiring issue discussed in section 5.2.1. The uncertainties on the scale factors are computed, for each p_T-η bin, as the squared sum of 20% of the object prefiring probability, considering as well the statistical uncertainty associated to the considered bin in the prefiring map for both electrons and muons [302].
- B-tagging (rate + shape): the SFs which are used to calculate the efficiency corrections of the DeepJet b-tagging algorithm are varied up and down within their uncertainties. From these up and down varied SFs, up and down shifted efficiency corrections are calculated, and applied to simulation. The uncertainties are split according to the flavor of the jet into a bottom flavor and a light flavor component. The CMS B-tagging group recommends implementing four total uncertainties for bottom and light flavors: three uncorrelated uncertainties for each year, and one more which is correlated for 2016, 2017, and 2018.
- Electron efficiency (rate + shape): simulated events are weighted by SFs to account for the differences between data and simulation in the electron trigger, reconstruction, and ID [283]. For the electron trigger, the CMS Electron and Photon group recommends implementing two uncertainties: one uncorrelated uncertainty for each year (statistical component), and one more which is correlated for 2016, 2017, and 2018 (systematic component). For the electron reconstruction and ID, the CMS Electron and Photon group recommends implementing one uncertainty for each of those sources, which is correlated for 2016, 2017, and 2018.
- Muon efficiency (rate + shape): simulated events are weighted by SFs to account for the differences between data and simulation in the muon trigger, reconstruction, ID, and isolation [284–286]. The CMS Muon group recommends implementing two uncertainties for each of those sources: one uncorrelated uncertainty for each year (statistical component), and one more which is correlated for 2016, 2017, and 2018 (systematic component).

- JES (rate + shape): the reconstructed jet energy differs from the particle-level jet energy due to contamination from PU and detector response effects. Therefore, all reconstructed jet four-momenta are simultaneously varied in simulation according to p_T- and $|\eta|$-dependent uncertainties [307]. The CMS JetMET group recommends implementing seven JES uncertainties [308], with or without correlation between the years, as summarized in Table 5.3. These variations are also propagated to the MET.

JES uncertainty source	Description	Correlation
Absolute	Variation in the CMS detector	Correlated + Uncorrelated
BBEC1	Variation in the pseudorapidity region $0 < \|\eta\| < 2.5$	Correlated + Uncorrelated
FlavorQCD	Variation between b-jets, c-jets and udsg-jets	Correlated
RelativeSample	Variation in residual corrections of the jets	Uncorrelated
RelativeBal	Variation in jet correction method	Correlated

Table 5.3: Uncertainty sources of the JES.

- MC modeling in the MR (shape): this uncertainty is applied exclusively to the data-driven charge flip and non prompt processes. The up variation of the uncertainty is determined by the discrepancy between the data and MC distributions in the MR, as shown in Figures 5.6, 5.8, and 5.9. This variation is then symmetrized to establish the down variation. This uncertainty is uncorrelated for each year.

- MC modeling between the MR and the targeted region (shape): this uncertainty is applied exclusively to the data-driven charge flip and non prompt processes. The up variation of the uncertainty is determined by the discrepancy of the MC distributions in the MR and the respective SR or CR, as shown in Figures 5.6, 5.8 and 5.9 for the MRs, and in Figures 5.1 and 5.2 for the SR and CR respectively. This variation is then symmetrized to establish the down variation. This uncertainty is uncorrelated for each year, plus uncorrelated between the SR and ttX CR for the non prompt process.

Theoretical uncertainties

The first part of the theoretical uncertainties are applied the same way as the experimental uncertainties to the signal and background processes, in all the regions and years. These uncertainties introduce a variation of the distributions by 15% at maximum, as illustrated for the signal with $m_{T'} = 700$ GeV in Appendix C.1.

- Top quark p_T reweighting (shape): in differential measurements of the top quark p_T in $t\bar{t}$ simulated events, the predicted p_T spectrum is found to be harder than the observed spectrum, as discussed in section 5.2.1. To account for this mismodeling, the results derived using the default simulation for the $t\bar{t}$ process are compared to the results using $t\bar{t}$ simulated events, which are reweighted according to the observed difference between data and MC. This results in one-sided variation of the nominal $t\bar{t}$ template.

- PDF (shape): this uncertainty is considered for all the processes except the WW DS process, for which it is not provided in the samples of the CMS collaboration. The systematic uncertainties due to the choice of the PDFs are studied by reweighting with eigenvector variations of the PDF set used in the sample, which is the NNPDF3.1 PDF set at NNLO here [309]. The systematically varied templates are constructed following the PDF4LHC recommendations [310].

- Renormalization and factorization scale (shape): this uncertainty is considered for all the processes except the WW DS process, for which it is not provided in the samples of the CMS collaboration. The envelope of all variations in which the renormalization and factorization scale are varied independently or simultaneously up or down is taken as uncertainty.

- ISR and FSR (shape): varied templates which correspond to a factor two up and down shifted scale are constructed. The FSR uncertainty is separated per process.

The normalization rate uncertainties are applied in a correlated way, in all the regions and years.

- $t\bar{t}$: a rate uncertainty of 6% is applied, taken from the uncertainty computed by the LHC Top group at NNLO [261].

- $t\bar{t}X$: a rate uncertainty of 40% is applied to be consistent with dedicated $t\bar{t}X$ analyses [299–301].

- Non prompt: a rate uncertainty of 30% is applied, as also used in [304, 311].

- Charge flip: a rate uncertainty of 30% is applied, as also used in [304, 311].

- Multibosons: a rate uncertainty of 10% is applied, as also used in [312, 313].

- Other processes: a rate uncertainty of 25% is applied to all other processes in the SR, which are the DY and $W + jets$ processes [304].

5.2.3 Validation with combination of the regions

Before presenting the results of the analysis in section 5.3, we perform an inclusive fit by injecting a signal strength of $r = 1$ in the pseudo-data, to confirm that the signal extraction related to the discriminant observable is working as expected, and that the nuisance parameters behave as expected. There is one parameter of interest in the fit, r, which is the signal strength for the VLQ production, and all the nuisance parameters related to the systematic uncertainties described in section 5.2.2 are included.

The combine framework and the Barlow-Beeston-lite method

All the results presented for the rest of the chapter use the software developed by the CMS collaboration, known as 'combine' [314]. The combine tool serves as the primary software framework for statistical model building in CMS physics analyses. It is built on top of the ROOT framework, enabling the encapsulation of the statistical model into a readable-human configuration file, known as a datacard [315]. Each bin of every distribution is treated independently, with the systematic uncertainties allowed to vary to estimate the optimal likelihood function. All regions and their respective likelihoods can then be combined to enhance sensitivity in searches or measurements. This approach assumes that the bins are statistically independent, which is valid here as all the regions have been defined orthogonally.

The simulated statistical uncertainties will be included using the Barlow-Beeston-lite method [316, 317]. This approach accounts for statistical uncertainties in the simulated samples by assigning a unique nuisance parameter to the sum of all processes, within each bin of the discriminant observable. The statistical uncertainty for each bin is calculated by considering the event weights associated with that bin. This method simplifies the model by reducing the number of statistical nuisance parameters to one per bin, thereby decreasing the processing time by a factor of 25 [317].

Systematic uncertainty source	Correlation between regions	Correlation between years	Correlation between processes
Luminosity, corr. component	✓	✓	partially
Luminosity, uncorr. component	✓	-	partially
Pileup	✓	✓	partially
L1 Trigger ECAL prefiring	✓	✓	partially
B-tagging (bottom), corr. component	✓	✓	partially
B-tagging (bottom), uncorr. component	✓	-	partially
B-tagging (light), corr. component	✓	✓	partially
B-tagging (light), uncorr. component	✓	-	partially
Electron trigger, syst. component	✓	✓	partially
Electron trigger, stat. component	✓	-	partially
Electron reconstruction	✓	✓	partially
Electron ID	✓	✓	partially
Muon trigger, syst. component	✓	✓	partially
Muon trigger, stat. component	✓	-	partially
Muon reconstruction, syst. component	✓	✓	partially
Muon reconstruction, stat. component	✓	-	partially
Muon ID, syst. component	✓	✓	partially
Muon ID, stat. component	✓	-	partially
Muon isolation, syst. component	✓	✓	partially
Muon isolation, stat. component	✓	-	partially
JES Absolute, corr. component	✓	✓	partially
JES Absolute, uncorr. component	✓	-	partially
JES BBEC1, corr. component	✓	✓	partially
JES BBEC1, uncorr. component	✓	-	partially
JES FlavorQCD	✓	✓	partially
JES RelativeSample	✓	-	partially
JES RelativeBal	✓	✓	partially
MC modeling MR	✓	-	partially
MC modeling MR/Target region	partially	-	partially
Top quark p_T reweighting	✓	✓	partially
PDF	✓	✓	partially
Renormalization and factorization scale	✓	✓	partially
ISR	✓	✓	partially
FSR	✓	✓	partially
Normalization of backgrounds	✓	✓	-

Table 5.4: Correlation scheme for the systematic uncertainties. Correlation between all processes except the non prompt and charge flip processes is total for all the systematic uncertainties, except for the normalization rate uncertainty.

Impact of the nuisance parameters

The impact of the nuisance parameters for the fits are evaluated with the combine tool, using the ’Impacts’ method [318]. The 60 largest impacts are presented in Figures 5.14 and 5.15 for 2016, 2017, and 2018 combined. The fit includes the $\mu\mu$, eμ, and ee channels, as well as the SR and ttX CR, with $m_{T'}$ = 700 GeV. Only Asimov datasets are considered, which means that data are replaced by the sum of the background processes, with a signal injection at $r = 1$. It is important to note that only the SR and ttX CR are included in the fit. The MRs are only used to estimate the charge flip and non prompt processes through a data-driven approach, and to define the systematic uncertainties associated with these processes. The variable $M_{T,T'}$ serves as the discriminant observable in both regions. The naming convention used in the Figures is as follows.

- lumi_cor_*: the luminosity uncertainty, correlated between years.
- lumi_uncor_*: the luminosity uncertainty, uncorrelated per year.
- syst_pu: the pileup uncertainty.
- syst_prefiring: the prefiring uncertainty.
- syst_b_correlated/syst_l_correlated: the uncertainty in the b-tagging SF for both the bottom and udsg-jets, correlated between years.
- syst_b_uncorrelated_*/syst_l_uncorrelated_*: the uncertainty in the b-tagging SF for both the bottom and udsg-jets, uncorrelated per year.
- syst_muon_*/syst_elec_*: the lepton systematic uncertainties, correlated between years.
- stat_muon_*/stat_elec_*: the lepton statistical uncertainties, uncorrelated per year.
- syst_JES_*: the JES systematic uncertainties, correlated or uncorrelated between years.
- syst_mr_flip_*/syst_mr_fake_*: the MC modeling uncertainty estimated with data and MC distributions in the MR, uncorrelated per year plus per region for the non prompt process.
- syst_tr_flip_*/syst_tr_fake_*: the MC modeling uncertainty estimated with MC distributions in the MR and the targeted region, uncorrelated per year plus per region for the non prompt process.
- syst_pt_top: the top quark p_T reweighting uncertainty.
- syst_pdfas: the PDF uncertainty.
- syst_qcdscale: the renormalization and factorization scale uncertainty.
- syst_isr: the ISR uncertainty.
- syst_fsr: the FSR uncertainty.
- r*: the normalization uncertainty, uncorrelated per process.
- prop_binch*_bin*: the statistical uncertainty, uncorrelated per channel and per bin.

The majority of the nuisance parameters remain constrained at their expected level after the fit. This is highlighted in the central region of the impact plots, where the best-fit value is centered at 0 with a standard deviation of 1σ. Additionally, their impact is moderate, with variations in the signal strength r of less than 0.2, as illustrated on the right side of the impact plots.

The normalization uncertainty of the $t\bar{t}X$ processes is the most important nuisance parameter. This uncertainty is constrained by 40% at maximum, and introduces variations in the signal strength of approximately 3 at maximum. This observation is corroborated by the correlation matrix, showing a strong level of anti-correlation (approximately -0.6) between the normalization uncertainty of the $t\bar{t}X$ processes and the signal strength. This result is expected, given that other processes, besides $t\bar{t}X$, are not negligible in the ttX CR, complicating the fit's ability to differentiate both the $t\bar{t}X$ and the signal processes.

The observations are similar for the nuisance parameters associated with the non prompt process. The normalization and data-driven uncertainties of the non prompt process are constrained by 30% at maximum, and introduce variations in the signal strength of approximately 1 at maximum. This observation is corroborated by the correlation matrix, showing a relatively strong level of anti-correlation (approximately -0.6) between the nuisance parameters associated with the non prompt process and the signal strength. This result is expected, given the significant discrepancies between the non prompt distributions in the MRs and the respective SR or CR. The results are similar for all the mass point values of the T', and can be found in Appendix C.2.

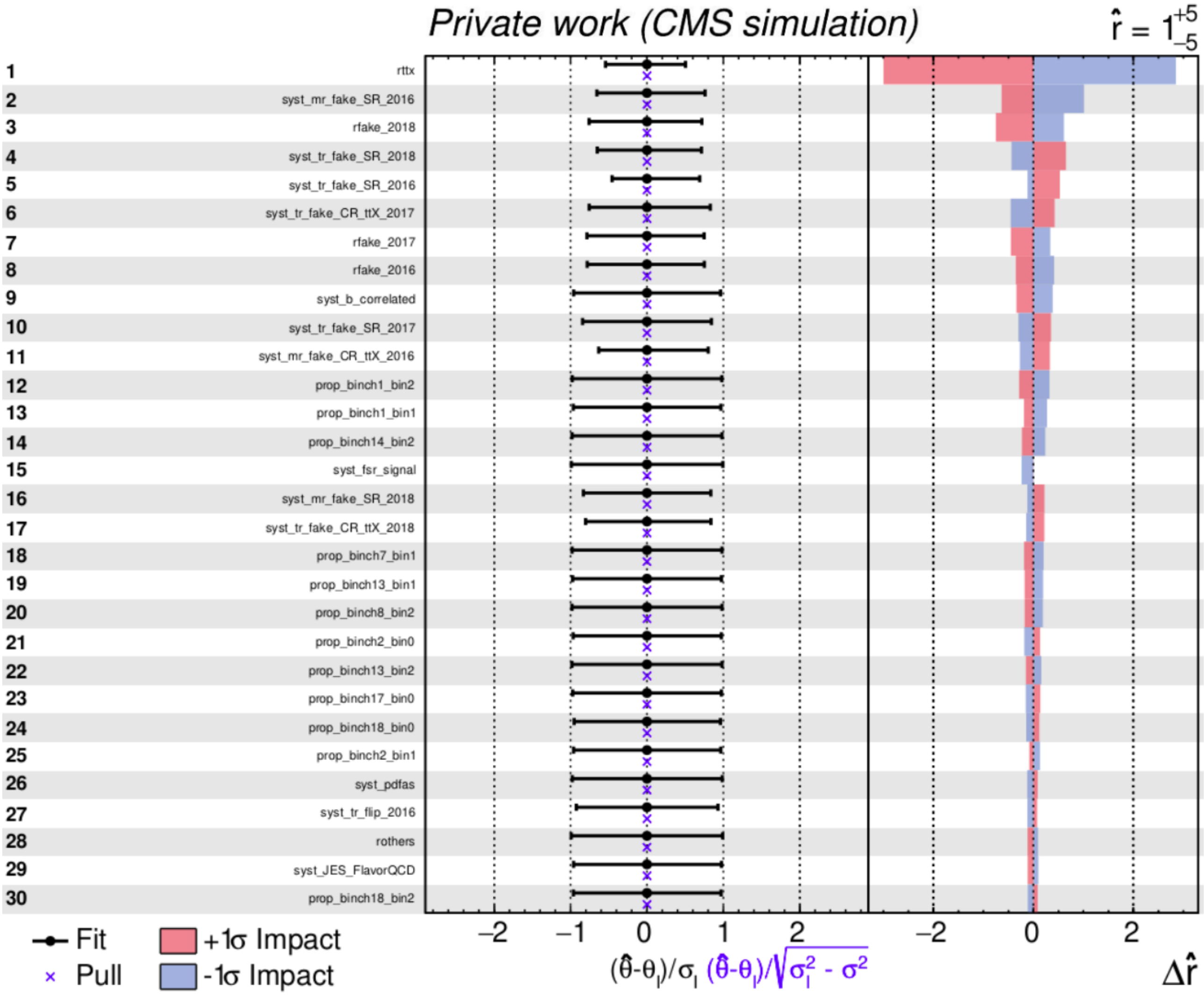


Figure 5.14: Impact of the nuisance parameters for the Run 2 period with all the channels and regions combined, computed using Run 2 simulation with a signal injection of $r = 1$ and $m_{T'} = 700$ GeV. The 30 first most important impacts are included.

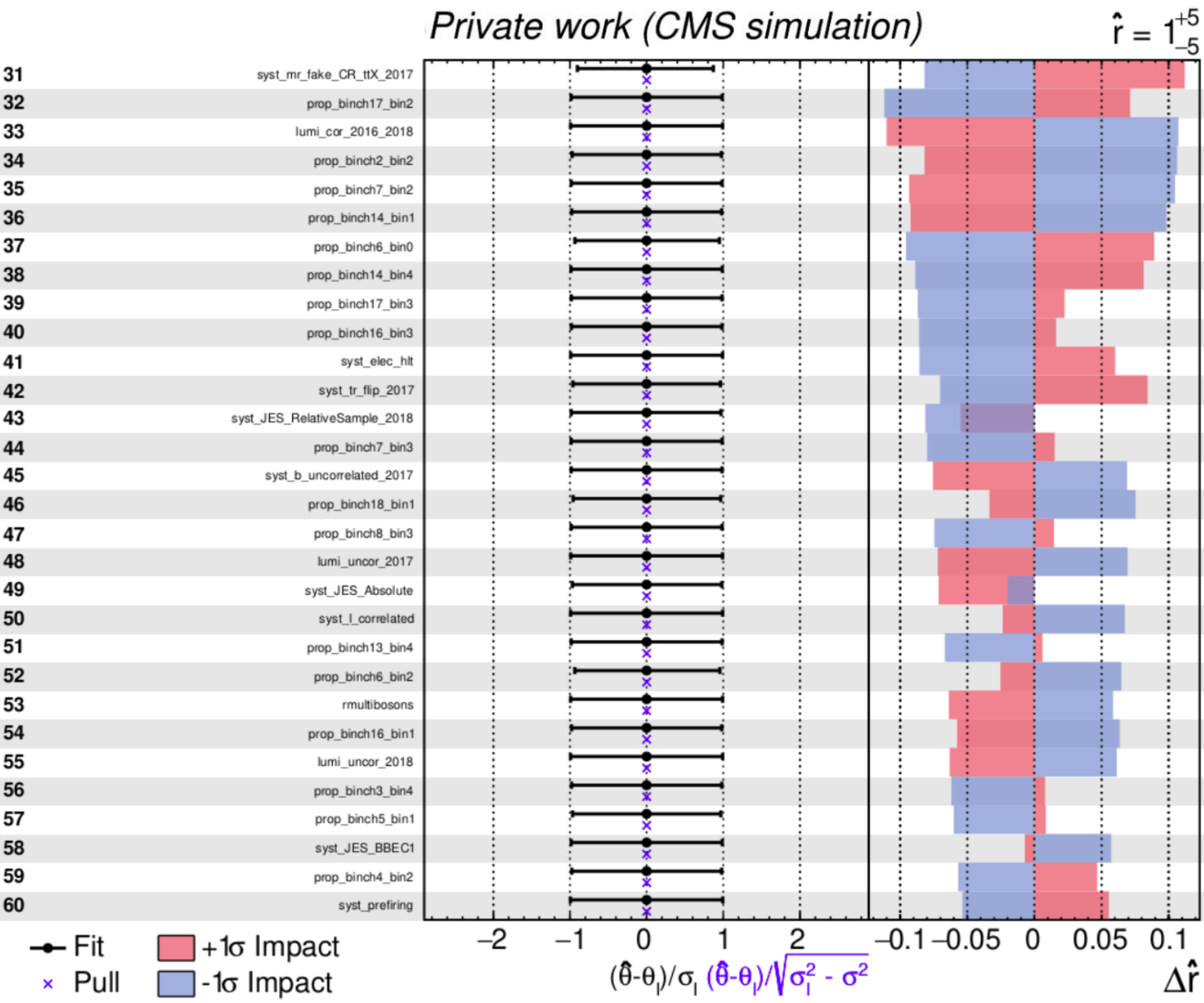


Figure 5.15: Impact of the nuisance parameters for the Run 2 period with all the channels and regions combined, computed using Run 2 simulation with a signal injection of $r = 1$ and $m_{T'} = 700$ GeV. The 30 second most important impacts are included.

5.3 Results

We have previously outlined the stages of the analysis strategy, including the object selection, the signal selection, the definition of the discriminant observable, the background estimation, and the evaluation of the uncertainties. We can now present the expected results based on simulated samples. Although a comparison with real data is currently not authorized before thorough review within the collaboration, it will be conducted in the future as the concluding phase of this analysis.

5.3.1 Signal extraction

Likelihood fit

The profiled likelihood fit is performed with the combine tool, using the 'MultiDimFit' method [319], to determine the signal strength r. The fit includes the $\mu\mu$, eμ, and ee channels, as well as the SR and ttX CR, and $m_{T'} = 700$ GeV. Only Asimov datasets are considered, with a signal injection at $r = 1$. The variable $M_{T,T'}$ serves as the discriminant observable. The results are presented for 2016, 2017, and 2018 separately, then combined for the three years.

- Asimov 2016: $r = 1.0^{+8.6}_{-8.1} = 1.0^{+6.6}_{-6.5}$ (Syst) $^{+5.5}_{-4.9}$ (Stat).
- Asimov 2017: $r = 1.0^{+8.3}_{-8.1} = 1.0^{+6.4}_{-6.5}$ (Syst) $^{+5.2}_{-4.8}$ (Stat).
- Asimov 2018: $r = 1.0^{+7.0}_{-6.9} = 1.0^{+5.6}_{-5.7}$ (Syst) $^{+4.2}_{-3.9}$ (Stat).
- Asimov Run 2: $r = 1.0^{+4.8}_{-4.7} = 1.0^{+4.0}_{-4.0}$ (Syst) $^{+2.8}_{-2.6}$ (Stat).

The analysis is systematic-dominated for the Run 2 period. This result is expected due to the large constraints and impacts of the normalization uncertainties of the $t\bar{t}X$ and non prompt processes, along with the MC modeling uncertainties associated with the non prompt process. Although statistical uncertainties are not negligible, they are more pronounced than systematic uncertainties on a per-year basis, as expected. The uncertainties with the largest impact are observed in 2016 and 2017, primarily due to their lower integrated luminosity compared to 2018. The results are similar for each mass point value of the T', and can be found in Appendix C.3.

Pre-fit and background-only fit plots in the Signal Region and ttX Control Region

The pre-fit distributions of $M_{T,T'}$ in the SR and ttX CR are presented in Figure 5.16. The distributions combine all the years, channels, and regions, with simulated and data samples. A signal injection at $r = 1$ is considered, with $m_{T'} = 700$ GeV. The variable $M_{T,T'}$ serves as the discriminant observable. The uncertainties presented here correspond to the values of the nuisance parameters before the fit. They introduce a global uncertainty of 10% in the SR where the signal is expected to be identified, and 40% in the ttX CR in the same $M_{T,T'}$ range. Data agrees with predictions, within uncertainties.

To improve the background estimation before the profiled likelihood fit, a first background-only fit is performed with the SR masked and the ttX CR 'unblinded'. The distributions of $M_{T,T'}$ after the background-only fit in the SR and ttX CR are presented in Figure 5.17. The $t\bar{t}$X rate is increased by 41.5%, as observed in dedicated $t\bar{t}$X analyses. Data agrees with predictions, within uncertainties, and the new shapes using post background-only fit results will be the input for the combined fit in both the unblinded SR and ttX CR.

5.3.2 Expected exclusion limits

The expected limits for the T' production cross section are computed for 2016, 2017, and 2018 combined, with all the channels and regions combined. It is important to note that the limits presented here are not using the post background-only fit nuisances, even though it will be the case in the future. They are performed with the combine tool, using the 'AsymptoticLimits' method [320]. The limit calculation is an approximation of the CL_s method [321], which sets upper limits on the signal strength r. The profiled likelihood fit is performed under the signal and background only hypotheses, to compute the respective p-values p_μ and p_b. The p-value is the probability that a statistical model produces results at least as extreme as those observed, under the null hypothesis. In other words, a small p-value indicates that observing such extreme results would be very unlikely under the null hypothesis. The expected limits are computed as a ratio between these two p-values, as follows [314]:

$$CL_s = \frac{p_\mu}{1 - p_b}. \tag{5.2}$$

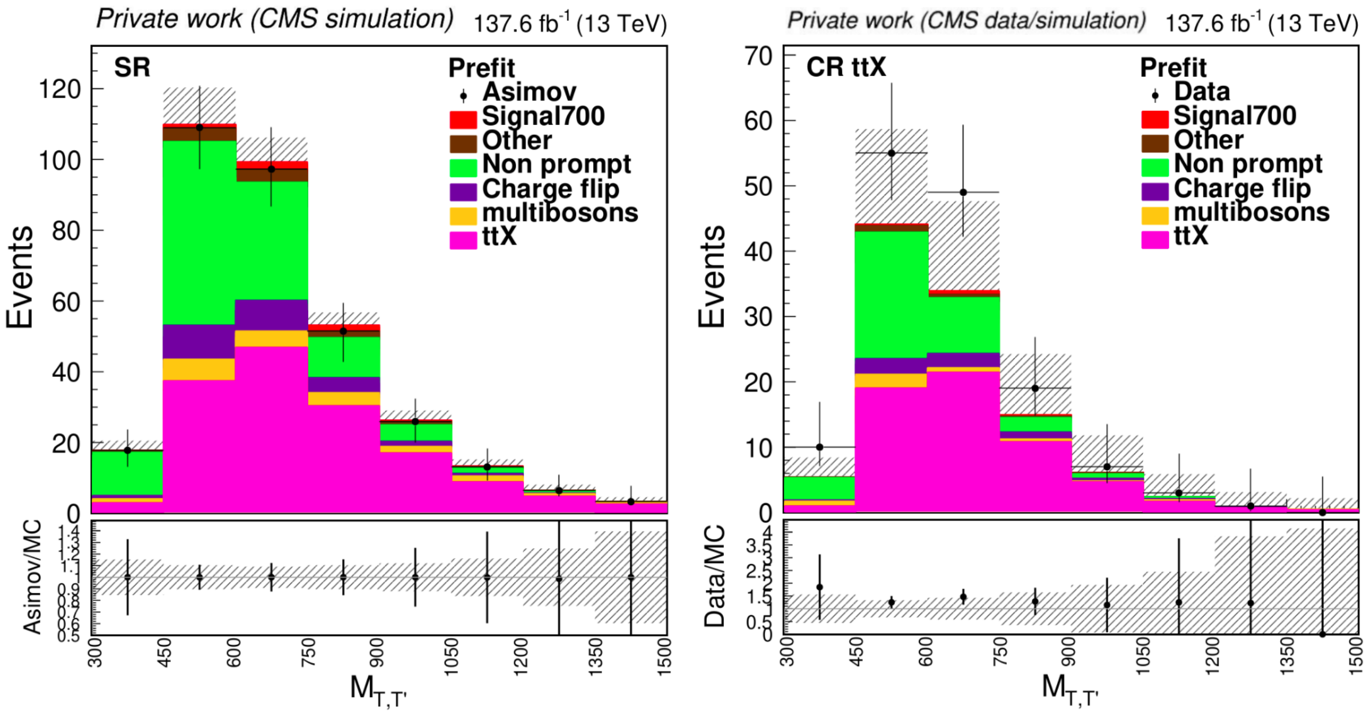


Figure 5.16: Distribution of $M_{T,T'}$ in the SR (left) and the ttX CR (right) before the background-only fit, comparing Run 2 simulation and data with $m_{T'} = 700$ GeV.

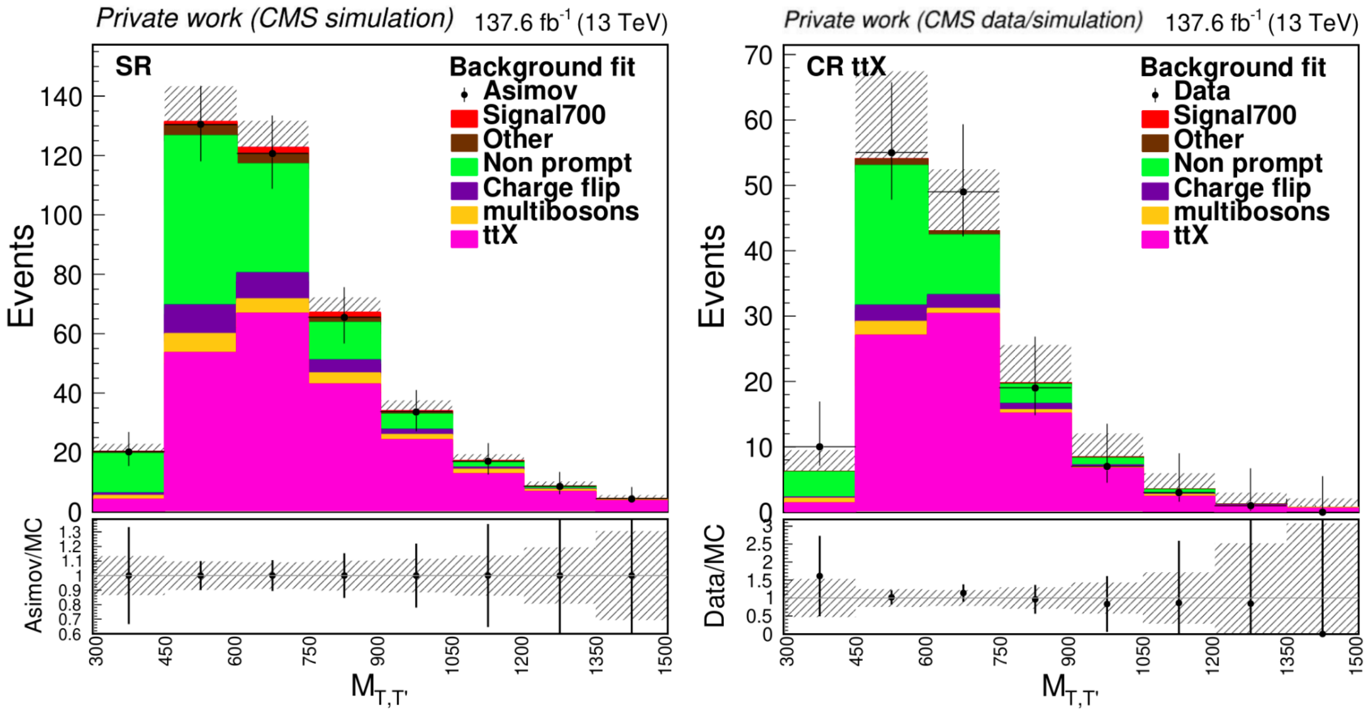


Figure 5.17: Distribution of $M_{T,T'}$ in the SR (left) and the ttX CR (right) after the background-only fit, comparing Run 2 simulation and data with $m_{T'} = 700$ GeV.

The limits extracted are upper limits on r at 95% CL. The results, as limits on the differential cross section of the T' → Ht production, are presented in Figure 5.18. They are calculated by multiplying the upper limits on r with the theoretical signal cross sections presented in Table 4.2.

The expected limits on the cross section are above the theoretical predictions, a result commonly observed in VLQs searches. It means that the present data does not show sufficient sensitivity to exclude the VLQs with the theoretical cross sections assumed. The limits in the SS dilepton final state are approximately 2 to 5 times less sensitive than those in the all-hadronic resolved, diphoton, and semileptonic final states ($\sigma \simeq 400$, 150 and 200 pb for $m_{T'} = 700$ GeV respectively vs 800 pb in the SS dilepton final state). This is confirmed by the expected significance and p-value, calculated for each mass point value of the T' using the 'Significance' method with the combine tool [322], as presented in Table 5.5. The significance does not exceed 0.3 σ, as expected due to the presence of two neutrinos in the final state, which complicates the total reconstruction of the T'. Eventually, comparison with real data will give additional information on the excess observed in the 2016 CMS all-hadronic resolved analysis.

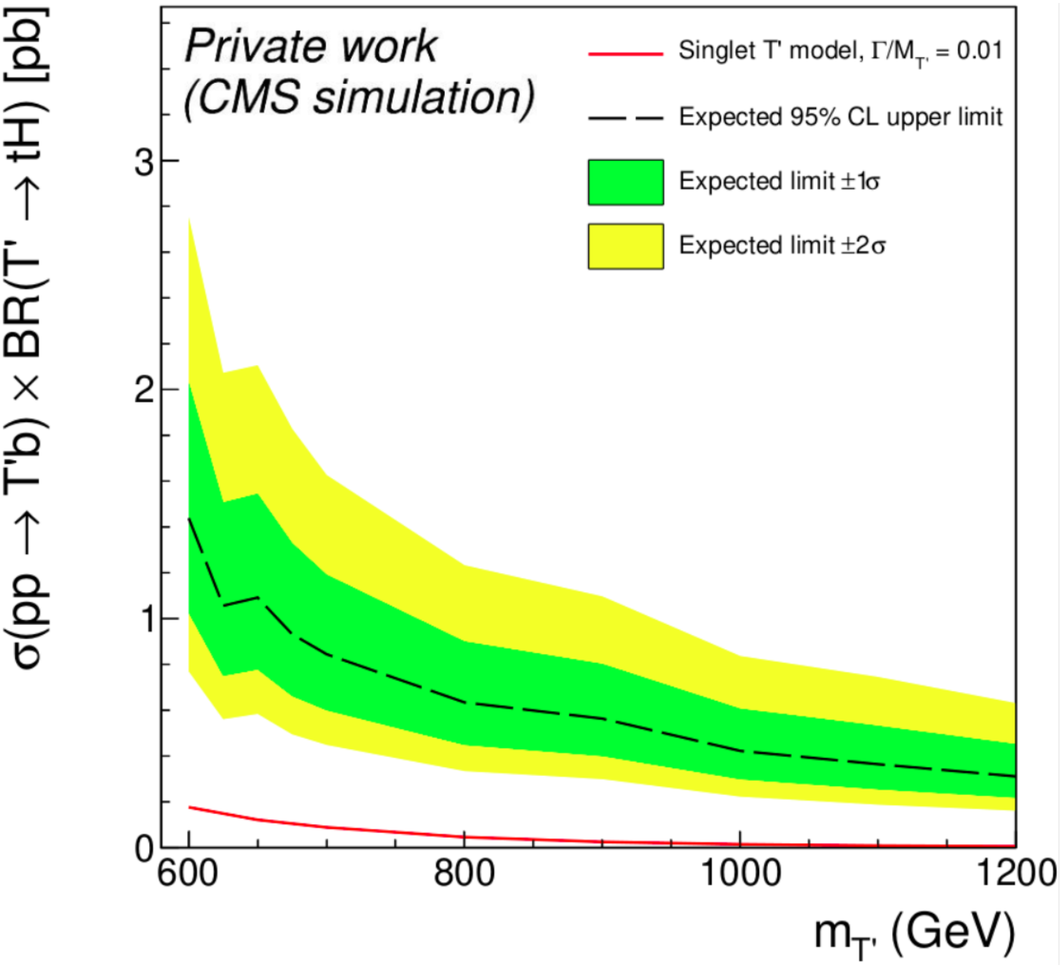


Figure 5.18: Expected exclusion limits at 95% CL of the cross section multiplied by the BR of the T' → Ht decay, computed using Run 2 simulation as a function of the nominal mass of the T'.

$m_{T'}$ (GeV)	Significance (σ)	P-value
600	0.243	0.404
625	0.284	0.388
650	0.221	0.412
675	0.228	0.410
700	0.211	0.417
800	0.149	0.441
900	0.091	0.464
1000	0.070	0.472
1100	0.052	0.479
1200	0.037	0.485

Table 5.5: Expected significance and p-value per mass point value of the T', computed using Run 2 simulation.

5.3.3 Beyond this analysis

Possible improvements of the analysis

Two methods could significantly enhance the sensitivity of the analysis within a short timeframe.

- A better inclusion of the neutrinos in the reconstruction of the final state: as previously mentioned, the neutrinos cannot be directly identified by the CMS detector, and their energy is accounted for in the MET. To directly use the neutrinos in the reconstruction of the T' and increase the sensitivity, several approaches can be made, such as using the M_{T2} variable as the discriminant observable [323]. This method involves calculating the transverse mass of both the top quark and the Higgs boson, and using the relation between the p_T of the two neutrinos through the MET to establish an upper limit, known as M_{T2}, on the mass of the particle of interest. A similar procedure was investigated during this thesis by using the W bosons, the Higgs boson, and the top quark nominal masses as additional constraints, but could not be fully realized due to time constraints.
- The implementation of machine learning: the application of various neural networks models, as discussed in Chapter 3, can improve the discrimination between the signal and the background, and particularly for the $t\bar{t}X$ processes. Machine learning techniques have become standard in high-energy physics research, often replacing traditional cut-based analysis strategies, as it is the case for the all-hadronic resolved analysis for the Run 3 period. However, due to time constraints, machine learning has not been integrated into this analysis.

Possible extensions of the analysis

The analysis presented here is based on the simplest model for the production and decay modes of the VLQs. To align experimental results more closely with theoretical predictions, the model can be extended by incorporating the following considerations.

- The production of the VLQ T' with a finite width and the inclusion of higher-order corrections: we saw in Chapter 1 that most of the VLQs searches were conducted in the NWA, with simulated samples generated at LO. To better depict theoretical models which approximate experimental expectations, it would be necessary to account for both a finite width and the inclusion of higher-order corrections during the VLQs production.

A preliminary model combining these aspects has been proposed, and will be the first step toward a more comprehensive theoretical framework [147].

- The extension of the T' decay into the tZ and ta channels: a corresponds to a neutral boson with a mass distinct from that of the Higgs boson. The analysis only considers the decay of the T' into a Higgs boson accompanied with a top quark. However, to ensure a comprehensive comparison with the all-hadronic resolved analysis, the tZ channel must be evaluated. Furthermore, recent observations have suggested the presence of a potential SM-like Higgs boson at a nominal mass between 70 GeV and 110 GeV [324]. This excess could expand our model to the T' decay into the ta channel.

- The extension of the VLQ T' nominal mass range: we saw in Chapter 1 that some analyses consider the nominal mass of the VLQs up to 3000 GeV. In our case, the decay products exhibit a large transverse momentum for nominal mass values of the T' exceeding 1000 GeV. Consequently, the two jets originating from the Higgs boson decay become so energetic that they could be merged into a single reconstructed AK8 jet. The event selection requires then 2 AK8 jets, 1 AK4 b-jet, two SS leptons accompanied with MET. This phenomenon contributes to the low expected significance for high T' mass values in section 5.3.2, where only AK4 jets were considered. Therefore, a revised analysis strategy could be developed for nominal mass values of the T' exceeding 1000 GeV, known as the 'boosted' regime.

- The extension of the experimental results to additional theoretical models: the analysis focuses on identifying a resonance decaying into a Higgs boson accompanied with a top quark. But this resonance is not strictly limited to a VLQ. Other hypothetical particles, such as the Vector-Like Leptons (VLLs) or the Heavy Neutral Leptons (HNLs), may also address the hierarchy problem in the Higgs sector through their interaction with the Higgs boson [325, 326].

In addition to this analysis, other final states are currently being investigated within the CMS collaboration as cross-check analyses of the 2016 excess. In particular, the OS dilepton final state in the resolved regime is investigated. The T' decays into a leptonic top quark accompanied with a Higgs boson, which subsequently decays into two leptonic W bosons. The reconstruction of the T' is facilitated by the spatial proximity of the neutrinos, allowing the MET to be directly used (with an additional hypothesis on the angle between the two neutrinos) in the definition of the invariant mass of the T' as the discriminant observable. However, the $t\bar{t}$ and DY processes dominate this final state due to their large cross sections, leading to an expected lower sensitivity than that in the SS dilepton final state. Both the SS and OS analyses may contribute to a combined scientific publication on the decay of the Vector-Like Quark T' in the tH channel in dilepton final states for the Run 2 period, in addition to the results in the all-hadronic final state. It is noteworthy that the trilepton final state was initially considered as an extension of the SS dilepton final state. The T' decays into a leptonic top quark accompanied with a Higgs boson, which subsequently decays into two leptonic W bosons. But it has not been further explored due to the presence of an additional neutrino in the final state, and the reduced theoretical cross section associated with the lower BR of the leptonic W boson decay compared to the hadronic W boson decay.

The analyses discussed here are centered on the Run 2 period. Nevertheless, investigations into VLQs have already started for the Run 3 period. In this context, the search for a single-produced VLQ T' in the tH channel in the SS dilepton final state could be continued as an extension of this thesis during the Run 3 period.

Conclusion

This thesis presents the search for the decay of the single-produced Vector-Like Quark T' into a Higgs boson accompanied with a top quark in the same sign dilepton final state for the Run 2 period, along with the development of the High Level Trigger path for the same decay in the all-hadronic resolved final state for the Run 3 period. The search for the Vector-Like Quark in the same sign dilepton final is a brand new analysis never before conducted in either the CMS or ATLAS experiments. A dedicated analysis strategy has been designed, and is presented throughout this manuscript.

The first step involved the identification of the various reconstructed particles targeting the final state. Selection criteria were studied, evaluated, and validated for the electrons, the muons, the jets, the b-jets, and the missing transverse energy. These objects benefit from the excellent resolution of the CMS detector, where each sub-detector is dedicated to the detection and reconstruction of a specific particle. Additionally, the corrections associated with every object, as provided by the CMS collaboration, were implemented and validated.

The second step involved defining the selection criteria in the signal region to identify the signal signature. Following the initial preselection at the trigger level, a cut-based approach was developed to achieve the discrimination between the signal and the various background processes. Every cut in the selection process was chosen based on the expected discrimination between the signal and background behaviors, and was refined to reduce specific background processes one at a time. This cut-based selection was subsequently optimized to preserve the background shape after the full selection for the considered nominal mass value of the Vector-Like Quark T', while ensuring a sufficient number of signal events. The last aspect of this step was the determination of the discriminant observable in the signal region, defined as the transverse mass of the Vector-Like Quark T'. Due to the presence of neutrinos in the final state which are not fully reconstructed within the CMS detector, several hypotheses were proposed, evaluated, and validated using particle-level information from simulated samples to define the transverse mass which provides the best discrimination between the signal and the background. These studies resulted in a reduction of the total background contribution by more than 99.9%, with the signal-over-background ratio reaching the 2% for mass point values of the Vector-Like Quark T' below 800 GeV.

The third step focused on the background estimation. Each background process which is not negligible in the signal region is estimated in a dedicated region, where it predominates. The background processes were regrouped in categories where a jet is misidentified as a lepton, and where the charge of a lepton is misidentified, treating these as independent background processes. These processes were evaluated in measurement regions, where they are considered almost pure. Thus, the data distribution replaces these background contributions in both the signal and control regions. This data-driven approach facilitates the overall analysis strategy, as only the signal region and the control regions will be used for estimating the signal strength. In total, four regions are defined in addition to the signal region: the ttX control region, the tt2l measurement region, the tt1l measurement region of the signal region, and the tt1l measurement region of the ttX control region.

The fourth step implemented the uncertainties in the analysis framework. The sources of these uncertainties were categorized into three types: the experimental systematic uncertainties, bringing together the uncertainties relative to the CMS detector miscalibrations, the theoretical systematic uncertainties, and the statistical uncertainties. Every source of uncertainty was correlated as appropriate between the regions, the years, and the processes. Their impact was evaluated with the combined fit involving the signal region and the ttX control region. The uncertainties with the most significant contributions were examined, and validated before the last step of the analysis.

The last step presented the results. The signal strength was first evaluated for each mass point value of the Vector-Like Quark T'. The expected exclusion limits were also presented with a 95% confidence level as a function of the nominal mass for the Vector-Like Quark T' production. This analysis demonstrates an expected sensitivity comparable to that of other final states, with a differential cross section ranging from 1400 to 400 pb for mass values of the Vector-Like Quark T' ranging from 600 to 1200 GeV. A comparison with real data will provide an important information regarding the potential presence of an excess for the decay of the Vector-Like Quark T' in the tH mode.

Several enhancements to the analysis strategy, such as the modification of the discriminant observable by including the missing transverse energy or the implementation of machine learning techniques, could be considered to improve the sensitivity in the same sign dilepton final state. Extending the theoretical framework could yield predictions which align more closely with experimental observations as well. Other final states are currently investigated within the CMS collaboration for the Run 2 period, and particularly the same Vector-Like Quark T' decay in the opposite sign dilepton final state. This work may contribute to a combined scientific publication on the decay of the Vector-Like Quark T' in the tH channel in dilepton and all-hadronic final states. Moreover, the search for Vector-Like Quarks continues for the Run 3 period, with further possible exploration of the same sign dilepton final state.

In addition to the analysis presented, a new High Level Trigger path has been designed for the Run 3 period, specifically targeting the Vector-Like Quark T' decay in the tH and tZ channels in the all-hadronic resolved final state. This new analysis extends the previous Run 2 analysis by implementing machine learning techniques. Various studies have been conducted to evaluate the jet reconstruction and b-jet identification algorithms with the highest efficiencies and pure rates, at both the online and offline levels of the analysis. Although initial challenges encountered during the data-taking periods of 2022 and 2023, the High Level Trigger path has been refined throughout this thesis to achieve excellent performances for the rest of the Run 3 period, with an overall efficiency exceeding 90%. In conclusion, the conditions are favorable for identifying the Vector-Like Quarks within the CMS collaboration, both presently and in the future.

Bibliography

[1] Galileo Galilei. *Dialogue concerning the two chief world systems, Ptolemaic and Copernican*. Univ of California Press, 2023.

[2] V I Yukalov. “Theory of cold atoms: Bose–Einstein statistics”. In: *Laser Physics* 26.6 (May 2016), p. 062001. DOI: 10.1088/1054-660X/26/6/062001. URL: https://dx.doi.org/10.1088/1054-660X/26/6/062001.

[3] Albert Einstein. “Über einem die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt”. In: *Annalen der physik* 4 (1905). DOI: https://doi.org/10.1002/andp.19053220607.

[4] Paul Adrien Maurice Dirac. “The quantum theory of the emission and absorption of radiation”. In: *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* 114.767 (1927), pp. 243–265. DOI: https://doi.org/10.1098/rspa.1927.0039.

[5] J. D. Jackson and L. B. Okun. “Historical roots of gauge invariance”. In: *Rev. Mod. Phys.* 73 (3 Sept. 2001), pp. 663–680. DOI: 10.1103/RevModPhys.73.663. URL: https://link.aps.org/doi/10.1103/RevModPhys.73.663.

[6] Lochlainn O’Raifeartaigh and Norbert Straumann. “Gauge theory: Historical origins and some modern developments”. In: *Rev. Mod. Phys.* 72 (1 Jan. 2000), pp. 1–23. DOI: 10.1103/RevModPhys.72.1. URL: https://link.aps.org/doi/10.1103/RevModPhys.72.1.

[7] W. Pauli. “Relativistic Field Theories of Elementary Particles”. In: *Rev. Mod. Phys.* 13 (3 July 1941), pp. 203–232. DOI: 10.1103/RevModPhys.13.203. URL: https://link.aps.org/doi/10.1103/RevModPhys.13.203.

[8] Particle Data Group et al. “Review of particle physics”. In: *Progress of theoretical and experimental physics* 2022.8 (2022), p. 083C01. DOI: https://doi.org/10.1093/ptep/ptac097.

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

[10] Jean Iliopoulos. “Introduction to the Standard Model of the Electro-Weak Interactions”. In: (2013). URL: https://cds.cern.ch/record/1551844.

[11] Murray Gell-Mann. “Symmetries of Baryons and Mesons”. In: *Phys. Rev.* 125 (), pp. 1067–1084. DOI: https://doi.org/10.1103/PhysRev.125.1067.

[12] O. W. Greenberg. “Spin and Unitary-Spin Independence in a Paraquark Model of Baryons and Mesons”. In: *Phys. Rev. Lett.* 13 (20 Nov. 1964), pp. 598–602. DOI: 10.1103/PhysRevLett.13.598. URL: https://link.aps.org/doi/10.1103/PhysRevLett.13.598.

[13] M. Y. Han and Y. Nambu. “Three-Triplet Model with Double SU(3) Symmetry”. In: *Phys. Rev.* 139 (4B Aug. 1965), B1006–B1010. DOI: `10.1103/PhysRev.139.B1006`. URL: `https://link.aps.org/doi/10.1103/PhysRev.139.B1006`.

[14] Andrei V Smilga. *Lectures on quantum chromodynamics*. World Scientific, 2001.

[15] David Griffiths. *Introduction to elementary particles*. John Wiley & Sons, 2020.

[16] J. C. Ward. “An Identity in Quantum Electrodynamics”. In: *Phys. Rev.* 78 (2 Apr. 1950), pp. 182–182. DOI: `10.1103/PhysRev.78.182`. URL: `https://link.aps.org/doi/10.1103/PhysRev.78.182`.

[17] Yasushi Takahashi. “On the generalized Ward identity”. In: *Il Nuovo Cimento (1955-1965)* 6.2 (1957), pp. 371–375. DOI: `https://doi.org/10.1007/BF02832514`.

[18] X. Fan et al. “Measurement of the Electron Magnetic Moment”. In: *Phys. Rev. Lett.* 130 (7 Feb. 2023), p. 071801. DOI: `10.1103/PhysRevLett.130.071801`. URL: `https://link.aps.org/doi/10.1103/PhysRevLett.130.071801`.

[19] David J. Gross and Frank Wilczek. “Ultraviolet Behavior of Non-Abelian Gauge Theories”. In: *Phys. Rev. Lett.* 30 (26 June 1973), pp. 1343–1346. DOI: `10.1103/PhysRevLett.30.1343`. URL: `https://link.aps.org/doi/10.1103/PhysRevLett.30.1343`.

[20] H. David Politzer. “Reliable Perturbative Results for Strong Interactions?” In: *Phys. Rev. Lett.* 30 (26 June 1973), pp. 1346–1349. DOI: `10.1103/PhysRevLett.30.1346`. URL: `https://link.aps.org/doi/10.1103/PhysRevLett.30.1346`.

[21] J. Bardeen, L. N. Cooper, and J. R. Schrieffer. “Theory of Superconductivity”. In: *Phys. Rev.* 108 (5 Dec. 1957), pp. 1175–1204. DOI: `10.1103/PhysRev.108.1175`. URL: `https://link.aps.org/doi/10.1103/PhysRev.108.1175`.

[22] Vladimir A Miransky. *Dynamical symmetry breaking in quantum field theories*. World Scientific, 1994.

[23] P. W. Anderson. “Plasmons, Gauge Invariance, and Mass”. In: *Phys. Rev.* 130 (1 Apr. 1963), pp. 439–442. DOI: `10.1103/PhysRev.130.439`. URL: `https://link.aps.org/doi/10.1103/PhysRev.130.439`.

[24] Y. Nambu and G. Jona-Lasinio. “Dynamical Model of Elementary Particles Based on an Analogy with Superconductivity. I”. In: *Phys. Rev.* 122 (1 Apr. 1961), pp. 345–358. DOI: `10.1103/PhysRev.122.345`. URL: `https://link.aps.org/doi/10.1103/PhysRev.122.345`.

[25] Y. Nambu and G. Jona-Lasinio. “Dynamical Model of Elementary Particles Based on an Analogy with Superconductivity. II”. In: *Phys. Rev.* 124 (1 Oct. 1961), pp. 246–254. DOI: `10.1103/PhysRev.124.246`. URL: `https://link.aps.org/doi/10.1103/PhysRev.124.246`.

[26] Jeffrey Goldstone. “Field theories with «Superconductor» solutions”. In: *Il Nuovo Cimento (1955-1965)* 19 (1961), pp. 154–164. DOI: `https://doi.org/10.1007/BF02812722`.

[27] Peter W. Higgs. “Broken Symmetries and the Masses of Gauge Bosons”. In: *Phys. Rev. Lett.* 13 (16 Oct. 1964), pp. 508–509. DOI: `10.1103/PhysRevLett.13.508`. URL: `https://link.aps.org/doi/10.1103/PhysRevLett.13.508`.

[28] F. Englert and R. Brout. “Broken Symmetry and the Mass of Gauge Vector Mesons”. In: *Phys. Rev. Lett.* 13 (9 Aug. 1964), pp. 321–323. DOI: `10.1103/PhysRevLett.13.321`. URL: `https://link.aps.org/doi/10.1103/PhysRevLett.13.321`.

[29] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. “Global Conservation Laws and Massless Particles”. In: *Phys. Rev. Lett.* 13 (20 Nov. 1964), pp. 585–587. DOI: 10.1103/PhysRevLett.13.585. URL: https://link.aps.org/doi/10.1103/PhysRevLett.13.585.

[30] Steven Weinberg. “Physical Processes in a Convergent Theory of the Weak and Electromagnetic Interactions”. In: *Phys. Rev. Lett.* 27 (24 Dec. 1971), pp. 1688–1691. DOI: 10.1103/PhysRevLett.27.1688. URL: https://link.aps.org/doi/10.1103/PhysRevLett.27.1688.

[31] Özer Özdal. “The Higgs boson and right-handed neutrinos in supersymmetric models”. PhD thesis. July 2016. DOI: 10.13140/RG.2.2.18314.52165.

[32] Sheldon L. Glashow. “Partial-symmetries of weak interactions”. In: *Nuclear Physics* 22.4 (1961), pp. 579–588. ISSN: 0029-5582. DOI: https://doi.org/10.1016/0029-5582(61)90469-2. URL: https://www.sciencedirect.com/science/article/pii/0029558261904692.

[33] A. Salam and J.C. Ward. “Electromagnetic and weak interactions”. In: *Physics Letters* 13.2 (1964), pp. 168–171. ISSN: 0031-9163. DOI: https://doi.org/10.1016/0031-9163(64)90711-5. URL: https://www.sciencedirect.com/science/article/pii/0031916364907115.

[34] Steven Weinberg. “A Model of Leptons”. In: *Phys. Rev. Lett.* 19 (21 Nov. 1967), pp. 1264–1266. DOI: 10.1103/PhysRevLett.19.1264. URL: https://link.aps.org/doi/10.1103/PhysRevLett.19.1264.

[35] Michela D’Onofrio and Kari Rummukainen. “Standard model cross-over on the lattice”. In: *Phys. Rev. D* 93 (2 Jan. 2016), p. 025003. DOI: 10.1103/PhysRevD.93.025003. URL: https://link.aps.org/doi/10.1103/PhysRevD.93.025003.

[36] Enrico Fermi. “Sulla quantizzazione del gas perfetto monoatomico”. In: *Rendiconti Lincei* 145 (1926). URL: https://www.scirp.org/reference/referencespapers?referenceid=1714963.

[37] Paul Adrien Maurice Dirac. “On the theory of quantum mechanics”. In: *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* 112.762 (1926), pp. 661–677. DOI: https://doi.org/10.1098/rspa.1926.0133.

[38] Paul Adrien Maurice Dirac. “Quantised singularities in the electromagnetic field”. In: *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* 133.821 (1931), pp. 60–72. DOI: https://doi.org/10.1098/rspa.1931.0130.

[39] C. S. Wu et al. “Experimental Test of Parity Conservation in Beta Decay”. In: *Phys. Rev.* 105 (4 Feb. 1957), pp. 1413–1415. DOI: 10.1103/PhysRev.105.1413. URL: https://link.aps.org/doi/10.1103/PhysRev.105.1413.

[40] R. P. Feynman and M. Gell-Mann. “Theory of the Fermi Interaction”. In: *Phys. Rev.* 109 (1 Jan. 1958), pp. 193–198. DOI: 10.1103/PhysRev.109.193. URL: https://link.aps.org/doi/10.1103/PhysRev.109.193.

[41] Sidney A Bludman. “On the universal Fermi interaction”. In: *Il Nuovo Cimento (1955-1965)* 9 (1958), pp. 433–445. DOI: https://doi.org/10.1007/BF02725099.

[42] Y Fukuda et al. “Measurements of the Solar Neutrino Flux from Super-Kamiokande’s First 300 Days”. In: *Phys. Rev. Lett.* 81 (6 Aug. 1998), pp. 1158–1162. DOI: 10.1103/PhysRevLett.81.1158. URL: https://link.aps.org/doi/10.1103/PhysRevLett.81.1158.

[43] Y Fukuda et al. "Measurement of the Flux and Zenith-Angle Distribution of Upward Throughgoing Muons by Super-Kamiokande". In: *Phys. Rev. Lett.* 82 (13 Mar. 1999), pp. 2644–2648. DOI: `10.1103/PhysRevLett.82.2644`. URL: `https://link.aps.org/doi/10.1103/PhysRevLett.82.2644`.

[44] N Agafonova et al. "Observation of a first ν_τ candidate event in the OPERA experiment in the CNGS beam". In: *Physics Letters B* 691.3 (2010), pp. 138–145. ISSN: 0370-2693. DOI: `https://doi.org/10.1016/j.physletb.2010.06.022`. URL: `https://www.sciencedirect.com/science/article/pii/S0370269310007537`.

[45] Kou Abe et al. "Evidence of electron neutrino appearance in a muon neutrino beam". In: *Phys. Rev. D* 88 (3 Aug. 2013), p. 032002. DOI: `10.1103/PhysRevD.88.032002`. URL: `https://link.aps.org/doi/10.1103/PhysRevD.88.032002`.

[46] WY Pauchy Hwang and Ta-You Wu. *Relativistic Quantum Mechanics and Quantum Fields: for the 21st Century*. World Scientific, 2018.

[47] Jeff Greensite. *An introduction to the confinement problem*. Vol. 821. Springer, 2011.

[48] Bo Andersson. *The lund model*. Cambridge University Press, 1998.

[49] Alexandre Obertelli and Hiroyuki Sagawa. "Nuclear Physics and Standard Model of Elementary Particles". In: *Modern Nuclear Physics: From Fundamentals to Frontiers*. Singapore: Springer Singapore, 2021, pp. 685–727. ISBN: 978-981-16-2289-2. DOI: `10.1007/978-981-16-2289-2_10`. URL: `https://doi.org/10.1007/978-981-16-2289-2_10`.

[50] V Aulchenko et al. "Observation of a resonance-like structure in the $\pi\pm\psi$'mass distribution in exclusive B $\rightarrow$ K$\pi\pm\psi$'decays". In: *23rd International Symposium on Lepton-Photon Interactions at High Energy, LP 2007*. 2007. DOI: `10.1103/PhysRevLett.100.142001`.

[51] Roel Aaij et al. "Observation of the Resonant Character of the $Z(4430)^-$ State". In: *Phys. Rev. Lett.* 112 (22 June 2014), p. 222002. DOI: `10.1103/PhysRevLett.112.222002`. URL: `https://link.aps.org/doi/10.1103/PhysRevLett.112.222002`.

[52] Roel Aaij et al. "Observation of $J/\psi p$ Resonances Consistent with Pentaquark States in $\Lambda_b^0 \rightarrow J/\psi K^- p$ Decays". In: *Phys. Rev. Lett.* 115 (7 Aug. 2015), p. 072001. DOI: `10.1103/PhysRevLett.115.072001`. URL: `https://link.aps.org/doi/10.1103/PhysRevLett.115.072001`.

[53] Richard P Feynman. "The behavior of hadron collisions at extreme energies". In: *Special Relativity and Quantum Theory: A Collection of Papers on the Poincaré Group* (1988), pp. 289–304. DOI: `https://doi.org/10.1007/978-94-009-3051-3_25`.

[54] Bogdan Povh et al. "Particles and nuclei". In: *An Introduction to the Physical Concepts* (1995). DOI: `https://doi.org/10.1007/978-3-662-46321-5`.

[55] J. H. Christenson et al. "Evidence for the 2π Decay of the K_2^0 Meson". In: *Phys. Rev. Lett.* 13 (4 July 1964), pp. 138–140. DOI: `10.1103/PhysRevLett.13.138`. URL: `https://link.aps.org/doi/10.1103/PhysRevLett.13.138`.

[56] Makoto Kobayashi and Toshihide Maskawa. "CP-Violation in the Renormalizable Theory of Weak Interaction". In: *Progress of Theoretical Physics* 49.2 (Feb. 1973), pp. 652–657. ISSN: 0033-068X. DOI: `10.1143/PTP.49.652`. URL: `https://doi.org/10.1143/PTP.49.652`.

[57] Nicola Cabibbo. "Unitary Symmetry and Leptonic Decays". In: *Phys. Rev. Lett.* 10 (12 June 1963), pp. 531–533. DOI: `10.1103/PhysRevLett.10.531`. URL: `https://link.aps.org/doi/10.1103/PhysRevLett.10.531`.

[58] S. L. Glashow, J. Iliopoulos, and L. Maiani. "Weak Interactions with Lepton-Hadron Symmetry". In: *Phys. Rev. D* 2 (7 Oct. 1970), pp. 1285–1292. DOI: `10.1103/PhysRevD.2.1285`. URL: `https://link.aps.org/doi/10.1103/PhysRevD.2.1285`.

[59] Claude Itzykson and Jean-Bernard Zuber. *Quantum field theory*. Courier Corporation, 2012.

[60] Michael E Peskin. *An introduction to quantum field theory*. CRC press, 2018.

[61] Chong-sa Lim, Toshiyuki Morii, and Shankar Nath Mukherjee. *The physics of the standard model and beyond*. World Scientific, 2004.

[62] Recai Erdem. "Counting the Observable Parameters of Yukawa Lagrangian". In: *Progress of Theoretical Physics* 93.5 (May 1995), pp. 961–967. ISSN: 0033-068X. DOI: `10.1143/ptp/93.5.961`. URL: `https://doi.org/10.1143/ptp/93.5.961`.

[63] F.J. Hasert et al. "Search for elastic muon-neutrino electron scattering". In: *Physics Letters B* 46.1 (1973), pp. 121–124. ISSN: 0370-2693. DOI: `https://doi.org/10.1016/0370-2693(73)90494-2`. URL: `https://www.sciencedirect.com/science/article/pii/0370269373904942`.

[64] G Arnison et al. "Experimental observation of isolated large transverse energy electrons with associated missing energy at s=540 GeV". In: *Physics Letters B* 122.1 (1983), pp. 103–116. ISSN: 0370-2693. DOI: `https://doi.org/10.1016/0370-2693(83)91177-2`. URL: `https://www.sciencedirect.com/science/article/pii/0370269383911772`.

[65] J. D. Bjorken. "The 1976 Nobel Prize in Physics". In: *Science* 194.4267 (1976), pp. 825–866. DOI: `10.1126/science.194.4267.825`. URL: `https://www.science.org/doi/abs/10.1126/science.194.4267.825`.

[66] Antonino Zichichi. "Foundations of Sequential Heavy Lepton Searches". In: *History of Original Ideas and Basic Discoveries in Particle Physics* (1996), pp. 227–275. DOI: `https://doi.org/10.1007/978-1-4613-1147-8_14`.

[67] Martin L Perl et al. "Evidence for Anomalous Lepton Production in $e^+ - e^-$ Annihilation". In: *Phys. Rev. Lett.* 35 (22 Dec. 1975), pp. 1489–1492. DOI: `10.1103/PhysRevLett.35.1489`. URL: `https://link.aps.org/doi/10.1103/PhysRevLett.35.1489`.

[68] S. W. Herb et al. "Observation of a Dimuon Resonance at 9.5 GeV in 400-GeV Proton-Nucleus Collisions". In: *Phys. Rev. Lett.* 39 (5 Aug. 1977), pp. 252–255. DOI: `10.1103/PhysRevLett.39.252`. URL: `https://link.aps.org/doi/10.1103/PhysRevLett.39.252`.

[69] Fumio Abe et al. "Observation of Top Quark Production in $\overline{p}p$ Collisions with the Collider Detector at Fermilab". In: *Phys. Rev. Lett.* 74 (14 Apr. 1995), pp. 2626–2631. DOI: `10.1103/PhysRevLett.74.2626`. URL: `https://link.aps.org/doi/10.1103/PhysRevLett.74.2626`.

[70] Shahriar Abachi et al. "Observation of the Top Quark". In: *Phys. Rev. Lett.* 74 (14 Apr. 1995), pp. 2632–2637. DOI: `10.1103/PhysRevLett.74.2632`. URL: `https://link.aps.org/doi/10.1103/PhysRevLett.74.2632`.

[71] The CMS Collaboration et al. "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC". In: *Physics Letters B* 716.1 (2012), pp. 30–61. ISSN: 0370-2693. DOI: `https://doi.org/10.1016/j.physletb.2012.08.021`. URL: `https://www.sciencedirect.com/science/article/pii/S0370269312008581`.

[72] Atlas Collaboration et al. “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC”. In: *Physics Letters B* 716.1 (2012), pp. 1–29. ISSN: 0370-2693. DOI: https://doi.org/10.1016/j.physletb.2012.08.020. URL: https://www.sciencedirect.com/science/article/pii/S037026931200857X.

[73] G. ’t Hooft and M. Veltman. “Regularization and renormalization of gauge fields”. In: *Nuclear Physics B* 44.1 (1972), pp. 189–213. ISSN: 0550-3213. DOI: https://doi.org/10.1016/0550-3213(72)90279-9. URL: https://www.sciencedirect.com/science/article/pii/0550321372902799.

[74] Nathaniel Craig. “Naturalness and Top-Down BSM Lecture 3”. In: *MITP Summer School* 2018 (2018).

[75] Abdus Salam. “Fundamental interaction”. In: (2020). DOI: https://doi.org/10.1036/1097-8542.275600.

[76] Nima Arkani–Hamed, Savas Dimopoulos, and Gia Dvali. “The hierarchy problem and new dimensions at a millimeter”. In: *Physics Letters B* 429.3 (1998), pp. 263–272. ISSN: 0370-2693. DOI: https://doi.org/10.1016/S0370-2693(98)00466-3. URL: https://www.sciencedirect.com/science/article/pii/S0370269398004663.

[77] John C Collins. *Renormalization: an introduction to renormalization, the renormalization group and the operator-product expansion*. Cambridge university press, 1984.

[78] G.’t Hooft. “Naturalness, Chiral Symmetry, and Spontaneous Chiral Symmetry Breaking”. In: *Recent Developments in Gauge Theories*. Ed. by G.’t Hooft et al. Boston, MA: Springer US, 1980, pp. 135–157. ISBN: 978-1-4684-7571-5. DOI: 10.1007/978-1-4684-7571-5_9. URL: https://doi.org/10.1007/978-1-4684-7571-5_9.

[79] Markus Fierz. “Über die relativistische Theorie kräftefreier Teilchen mit beliebigem Spin”. In: *Helvetica Physica Acta* 12 (Jan. 1939), pp. 3–37. URL: https://ui.adsabs.harvard.edu/abs/1939AcHPh..12....3F.

[80] W. Pauli. “The Connection Between Spin and Statistics”. In: *Phys. Rev.* 58 (8 Oct. 1940), pp. 716–722. DOI: 10.1103/PhysRev.58.716. URL: https://link.aps.org/doi/10.1103/PhysRev.58.716.

[81] P. Fayet and S. Ferrara. “Supersymmetry”. In: *Physics Reports* 32.5 (1977), pp. 249–334. ISSN: 0370-1573. DOI: https://doi.org/10.1016/0370-1573(77)90066-7. URL: https://www.sciencedirect.com/science/article/pii/0370157377900667.

[82] Stephen P Martin. “A supersymmetry primer”. In: *Perspectives on supersymmetry II*. World Scientific, 2010, pp. 1–153. DOI: https://doi.org/10.1142/9789814307505_0001.

[83] Savas Dimopoulos and Howard Georgi. “Softly broken supersymmetry and SU(5)”. In: *Nuclear Physics B* 193.1 (1981), pp. 150–162. ISSN: 0550-3213. DOI: https://doi.org/10.1016/0550-3213(81)90522-8. URL: https://www.sciencedirect.com/science/article/pii/0550321381905228.

[84] Csaba Csaki. “The minimal supersymmetric standard model”. In: *Modern Physics Letters A* 11.08 (1996), pp. 599–613. DOI: https://doi.org/10.1142/S021773239600062X.

[85] Howard Baer and Xerxes Tata. *Weak scale supersymmetry: From superfields to scattering events*. Cambridge University Press, 2007.

[86] Michael Dine. *Supersymmetry and string theory: Beyond the standard model*. Cambridge University Press, 2016.

[87] Atlas Collaboration et al. "SUSY August 2023 Summary Plot Update". In: (2023). URL: https://cds.cern.ch/record/2871728.

[88] Oliver Witzel. "Review on composite Higgs models". In: *arXiv preprint arXiv:1901.08216* (2019). DOI: https://doi.org/10.22323/1.334.0006.

[89] Howard Georgi and David B. Kaplan. "Composite Higgs and custodial SU(2)". In: *Physics Letters B* 145.3 (1984), pp. 216–220. ISSN: 0370-2693. DOI: https://doi.org/10.1016/0370-2693(84)90341-1. URL: https://www.sciencedirect.com/science/article/pii/0370269384903411.

[90] David B. Kaplan, Howard Georgi, and Savas Dimopoulos. "Composite Higgs scalars". In: *Physics Letters B* 136.3 (1984), pp. 187–190. ISSN: 0370-2693. DOI: https://doi.org/10.1016/0370-2693(84)91178-X. URL: https://www.sciencedirect.com/science/article/pii/037026938491178X.

[91] Ari Hietanen et al. "Fundamental composite Higgs dynamics on the lattice: SU (2) with two flavors". In: *Journal of High Energy Physics* 2014.7 (2014), pp. 1–15. DOI: https://doi.org/10.1007/JHEP07(2014)116.

[92] Rudy Arthur et al. "SU(2) gauge theory with two fundamental flavors: A minimal template for model building". In: *Phys. Rev. D* 94 (9 Nov. 2016), p. 094507. DOI: 10.1103/PhysRevD.94.094507. URL: https://link.aps.org/doi/10.1103/PhysRevD.94.094507.

[93] Avik Banerjee et al. "Constraining composite Higgs models using LHC data". In: *Journal of High Energy Physics* 2018.3 (2018), pp. 1–24. DOI: https://doi.org/10.1007/JHEP03(2018)062.

[94] Charanjit K Khosa, Veronica Sanz, et al. "On the Impact of the LHC Run 2 Data on General Composite Higgs Scenarios". In: *Advances in High Energy Physics* 2022 (2022). URL: https://www.hindawi.com/journals/ahep/2022/8970837.

[95] Martin Schmaltz and David Tucker-Smith. "Little higgs theories". In: *Annu. Rev. Nucl. Part. Sci.* 55 (2005), pp. 229–270. DOI: https://doi.org/10.1155/2022/8970837.

[96] Murray Gell-Mann, R. J. Oakes, and B. Renner. "Behavior of Current Divergences under $SU_3 \times SU_3$". In: *Phys. Rev.* 175 (5 Nov. 1968), pp. 2195–2199. DOI: 10.1103/PhysRev.175.2195. URL: https://link.aps.org/doi/10.1103/PhysRev.175.2195.

[97] Nima Arkani-Hamed et al. "The Littlest Higgs". In: *Journal of High Energy Physics* 2002.07 (Aug. 2002), p. 034. DOI: 10.1088/1126-6708/2002/07/034. URL: https://dx.doi.org/10.1088/1126-6708/2002/07/034.

[98] Martin Schmaltz. "The simplest little Higgs". In: *Journal of High Energy Physics* 2004.08 (Sept. 2004), p. 056. DOI: 10.1088/1126-6708/2004/08/056. URL: https://dx.doi.org/10.1088/1126-6708/2004/08/056.

[99] Th Kaluza. "On the Unification Problem in Physics". In: *Sitzungsber. Preuss. Akad. Wiss. Berlin (Math. Phys.)* 1921.arXiv: 1803.08616 (1921), pp. 966–972. DOI: https://doi.org/10.1142/S0218271818700017.

[100] Oskar Klein. "Quantentheorie und fünfdimensionale Relativitätstheorie". In: *Zeitschrift für Physik* 37.12 (1926), pp. 895–906. DOI: 10.34663/9783945561317-10.

[101] Nima Arkani-Hamed, Savas Dimopoulos, and Gia Dvali. "Phenomenology, astrophysics, and cosmology of theories with submillimeter dimensions and TeV scale quantum gravity". In: *Phys. Rev. D* 59 (8 Mar. 1999), p. 086004. DOI: 10.1103/PhysRevD.59.086004. URL: https://link.aps.org/doi/10.1103/PhysRevD.59.086004.

[102] Thomas Appelquist, Hsin-Chia Cheng, and Bogdan A. Dobrescu. "Bounds on universal extra dimensions". In: *Phys. Rev. D* 64 (3 June 2001), p. 035002. DOI: 10.1103/PhysRevD.64.035002. URL: https://link.aps.org/doi/10.1103/PhysRevD.64.035002.

[103] Lisa Randall and Raman Sundrum. "Large Mass Hierarchy from a Small Extra Dimension". In: *Phys. Rev. Lett.* 83 (17 Oct. 1999), pp. 3370–3373. DOI: 10.1103/PhysRevLett.83.3370. URL: https://link.aps.org/doi/10.1103/PhysRevLett.83.3370.

[104] Bin Guo et al. "Hierarchy problem and new warped extra dimension". In: *Phys. Rev. D* 98 (8 Oct. 2018), p. 085022. DOI: 10.1103/PhysRevD.98.085022. URL: https://link.aps.org/doi/10.1103/PhysRevD.98.085022.

[105] Marvin M Flores et al. "Updated LHC bounds on MUED after run 2". In: *International Journal of Modern Physics A* 38.01 (2023), p. 2350002. DOI: https://doi.org/10.1142/S0217751X23500021.

[106] João M. Alves et al. "Vector-like singlet quarks: A roadmap". In: *Physics Reports* 1057 (2024), pp. 1–69. ISSN: 0370-1573. DOI: https://doi.org/10.1016/j.physrep.2023.12.004. URL: https://www.sciencedirect.com/science/article/pii/S0370157323004374.

[107] Luca Panizzi. "Vector-like quarks: *tndpartners*". In: *Il nuovo cimento C* 37.2 (2014), pp. 69–79. DOI: 10.1393/ncc/i2014-11738-x.

[108] Gustavo Castelo Branco and MN Rebelo. "Vector-like Quarks". In: *arXiv preprint arXiv:2208.07235* (2022). DOI: https://doi.org/10.48550/arXiv.2208.07235.

[109] Otto Eberhardt et al. "Impact of a Higgs Boson at a Mass of 126 GeV on the Standard Model with Three and Four Fermion Generations". In: *Phys. Rev. Lett.* 109 (24 Dec. 2012), p. 241802. DOI: 10.1103/PhysRevLett.109.241802. URL: https://link.aps.org/doi/10.1103/PhysRevLett.109.241802.

[110] Tsutomu Yanagida. "Horizontal Symmetry and Masses of Neutrinos". In: *Progress of Theoretical Physics* 64.3 (Sept. 1980), pp. 1103–1105. ISSN: 0033-068X. DOI: 10.1143/PTP.64.1103. URL: https://doi.org/10.1143/PTP.64.1103.

[111] Ettore Majorana. "Teoria simmetrica dell'elettrone e del positrone". In: *Il Nuovo Cimento (1924-1942)* 14.4 (1937), pp. 171–184. DOI: https://doi.org/10.1007/BF02961314.

[112] B. Pontecorvo. "Inverse beta processes and nonconservation of lepton charge". In: *Zh. Eksp. Teor. Fiz.* 34 (1957), p. 247. URL: https://inspirehep.net/literature/42736.

[113] Ziro Maki, Masami Nakagawa, and Shoichi Sakata. "Remarks on the Unified Model of Elementary Particles". In: *Progress of Theoretical Physics* 28.5 (Nov. 1962), pp. 870–880. ISSN: 0033-068X. DOI: 10.1143/PTP.28.870. URL: https://doi.org/10.1143/PTP.28.870.

[114] F. del Aguila, M. Pérez-Victoria, and J. Santiago. "Effective description of quark mixing". In: *Physics Letters B* 492.1 (2000), pp. 98–106. ISSN: 0370-2693. DOI: https://doi.org/10.1016/S0370-2693(00)01071-6. URL: https://www.sciencedirect.com/science/article/pii/S0370269300010716.

[115] "Model-independent framework for searches of top partners". In: *Nuclear Physics B* 876.2 (2013), pp. 376–417. ISSN: 0550-3213. DOI: https://doi.org/10.1016/j.nuclphysb.2013.08.010. URL: https://www.sciencedirect.com/science/article/pii/S055032131300432X.

[116] Benjamin Fuks and Hua-Sheng Shao. "QCD next-to-leading-order predictions matched to parton showers for vector-like quark models". In: *The European Physical Journal C* 77 (2017), pp. 1–21. DOI: https://doi.org/10.1140/epjc/s10052-017-4686-z.

[117] J.A. Aguilar-Saavedra and B.M. Nobre. “Rare top decays t→cγ, t→cg and CKM unitarity”. In: *Physics Letters B* 553.3 (2003), pp. 251–260. ISSN: 0370-2693. DOI: https://doi.org/10.1016/S0370-2693(02)03230-6. URL: https://www.sciencedirect.com/science/article/pii/S0370269302032306.

[118] Juan Antonio Aguilar-Saavedra, Daniel E López-Fogliani, and C Muñoz. “Novel signatures for vector-like quarks”. In: *Journal of High Energy Physics* 2017.6 (2017), pp. 1–27. DOI: https://doi.org/10.1007/JHEP06(2017)095.

[119] Stephen P. Martin. “Extra vectorlike matter and the lightest Higgs scalar boson mass in low-energy supersymmetry”. In: *Phys. Rev. D* 81 (3 Feb. 2010), p. 035004. DOI: 10.1103/PhysRevD.81.035004. URL: https://link.aps.org/doi/10.1103/PhysRevD.81.035004.

[120] David B. Kaplan. “Flavor at ssc energies: A new mechanism for dynamically generated fermion masses”. In: *Nuclear Physics B* 365.2 (1991), pp. 259–278. ISSN: 0550-3213. DOI: https://doi.org/10.1016/S0550-3213(05)80021-5. URL: https://www.sciencedirect.com/science/article/pii/S0550321305800215.

[121] Kaustubh Agashe, Roberto Contino, and Alex Pomarol. “The minimal composite Higgs model”. In: *Nuclear Physics B* 719.1 (2005), pp. 165–187. ISSN: 0550-3213. DOI: https://doi.org/10.1016/j.nuclphysb.2005.04.035. URL: https://www.sciencedirect.com/science/article/pii/S0550321305003445.

[122] Nima Arkani-Hamed et al. “The Minimal Moose for a Little Higgs”. In: *Journal of High Energy Physics* 2002.08 (Aug. 2002), p. 021. DOI: 10.1088/1126-6708/2002/08/021. URL: https://dx.doi.org/10.1088/1126-6708/2002/08/021.

[123] Paolo Lodone. “Vector-like quarks in a “composite” Higgs model”. In: *Journal of High Energy Physics* 2008.12 (Dec. 2008), p. 029. DOI: 10.1088/1126-6708/2008/12/029. URL: https://dx.doi.org/10.1088/1126-6708/2008/12/029.

[124] Oleksii Matsedonskyi, Giuliano Panico, and Andrea Wulzer. “Light top partners for a light composite Higgs”. In: *Journal of High Energy Physics* 2013.1 (2013), pp. 1–40. DOI: https://doi.org/10.1007/JHEP01(2013)164.

[125] Andreas Crivellin et al. “Global fit of modified quark couplings to EW gauge bosons and vector-like quarks in light of the Cabibbo angle anomaly”. In: *Journal of High Energy Physics* 2023.3 (2023), pp. 1–40. DOI: https://doi.org/10.1007/JHEP03(2023)234.

[126] GC Branco et al. “Addressing the CKM unitarity problem with a vector-like up quark”. In: *Journal of High Energy Physics* 2021.7 (2021), pp. 1–24. DOI: https://doi.org/10.1007/JHEP07(2021)099.

[127] Francisco J Botella et al. “Decays of the heavy top and new insights on K in a one-VLQ minimal solution to the CKM unitarity problem”. In: *The European Physical Journal C* 82.4 (2022), p. 360. DOI: https://doi.org/10.1140/epjc/s10052-022-10299-9.

[128] G. ’t Hooft. “Symmetry Breaking through Bell-Jackiw Anomalies”. In: *Phys. Rev. Lett.* 37 (1 July 1976), pp. 8–11. DOI: 10.1103/PhysRevLett.37.8. URL: https://link.aps.org/doi/10.1103/PhysRevLett.37.8.

[129] R. D. Peccei and Helen R. Quinn. “CP Conservation in the Presence of Pseudoparticles”. In: *Phys. Rev. Lett.* 38 (25 June 1977), pp. 1440–1443. DOI: 10.1103/PhysRevLett.38.1440. URL: https://link.aps.org/doi/10.1103/PhysRevLett.38.1440.

[130] R. D. Peccei and Helen R. Quinn. “Constraints imposed by CP conservation in the presence of pseudoparticles”. In: *Phys. Rev. D* 16 (6 Sept. 1977), pp. 1791–1797. DOI: 10.1103/PhysRevD.16.1791. URL: https://link.aps.org/doi/10.1103/PhysRevD.16.1791.

[131] S. M. Barr. “Solving the Strong CP Problem without the Peccei-Quinn Symmetry”. In: *Phys. Rev. Lett.* 53 (4 July 1984), pp. 329–332. DOI: 10.1103/PhysRevLett.53.329. URL: https://link.aps.org/doi/10.1103/PhysRevLett.53.329.

[132] Luis Bento, Gustavo C. Branco, and Paulo A. Parada. “A minimal model with natural suppresion of strong CP violation”. In: *Physics Letters B* 267.1 (1991), pp. 95–99. ISSN: 0370-2693. DOI: https://doi.org/10.1016/0370-2693(91)90530-4. URL: https://www.sciencedirect.com/science/article/pii/0370269391905304.

[133] GC Branco, PA Parada, and MN Rebelo. “A Common origin for all CP violations”. In: (2003). DOI: https://doi.org/10.48550/arXiv.hep-ph/0307119. URL: https://cds.cern.ch/record/627762.

[134] S. Sultansoy and G. Unel. “The E6 inspired isosinglet quark and the Higgs boson”. In: *Physics Letters B* 669.1 (2008), pp. 39–45. ISSN: 0370-2693. DOI: https://doi.org/10.1016/j.physletb.2008.09.022. URL: https://www.sciencedirect.com/science/article/pii/S0370269308011453.

[135] Junhai Kang, Paul Langacker, and Brent D. Nelson. “Theory and phenomenology of exotic isosinglet quarks and squarks”. In: *Phys. Rev. D* 77 (3 Feb. 2008), p. 035003. DOI: 10.1103/PhysRevD.77.035003. URL: https://link.aps.org/doi/10.1103/PhysRevD.77.035003.

[136] Andy Buckley et al. “New sensitivity of current LHC measurements to vector-like quarks”. In: *SciPost Phys.* 9 (2020), p. 069. DOI: 10.21468/SciPostPhys.9.5.069. URL: https://scipost.org/10.21468/SciPostPhys.9.5.069.

[137] Giacomo Cacciapaglia et al. “Probing vector-like quark models with Higgs-boson pair production”. In: *Journal of High Energy Physics* 2017.7 (2017), pp. 1–26. DOI: https://doi.org/10.1007/JHEP07(2017)005.

[138] J. A. Aguilar-Saavedra et al. “Handbook of vectorlike quarks: Mixing and single production”. In: *Phys. Rev. D* 88 (9 Nov. 2013), p. 094010. DOI: 10.1103/PhysRevD.88.094010. URL: https://link.aps.org/doi/10.1103/PhysRevD.88.094010.

[139] F. del Aguila et al. “Vector-like fermion and standard Higgs production at hadron colliders”. In: *Nuclear Physics B* 334.1 (1990), pp. 1–23. ISSN: 0550-3213. DOI: https://doi.org/10.1016/0550-3213(90)90655-W. URL: https://www.sciencedirect.com/science/article/pii/055032139090655W.

[140] Giacomo Cacciapaglia et al. “Interplay of vector-like top partner multiplets in a realistic mixing set-up”. In: *Journal of High Energy Physics* 2015.9 (2015), pp. 1–50. DOI: https://doi.org/10.1007/JHEP09(2015)012.

[141] Giacomo Cacciapaglia et al. “The LHC potential of Vector-like quark doublets”. In: *Journal of High Energy Physics* 2018.11 (2018), pp. 1–28. DOI: https://doi.org/10.1007/JHEP11(2018)055.

[142] Patrick J Fox and Neal Weiner. “Light signals from a lighter Higgs”. In: *Journal of High Energy Physics* 2018.8 (2018), pp. 1–14. DOI: https://doi.org/10.1007/JHEP08(2018)025.

[143] Rachid Benbrik et al. “Signatures of vector-like top partners decaying into new neutral scalar or pseudoscalar bosons”. In: *Journal of High Energy Physics* 2020.5 (2020), pp. 1–48. DOI: https://doi.org/10.1007/JHEP05(2020)028.

[144] G. Breit and E. Wigner. “Capture of Slow Neutrons”. In: *Phys. Rev.* 49 (7 Apr. 1936), pp. 519–531. DOI: 10.1103/PhysRev.49.519. URL: https://link.aps.org/doi/10.1103/PhysRev.49.519.

[145] Rikkert Frederix et al. "The automation of next-to-leading order electroweak calculations". In: *Journal of High Energy Physics* 2018.7 (2018), pp. 1–121. DOI: https://doi.org/10.1007/JHEP07(2018)185.

[146] D.Y. Bardin et al. "Energy-dependent width effects in e+e—-annihilation near the Z-boson pole". In: *Physics Letters B* 206.3 (1988), pp. 539–542. ISSN: 0370-2693. DOI: https://doi.org/10.1016/0370-2693(88)91627-9. URL: https://www.sciencedirect.com/science/article/pii/0370269388916279.

[147] Aldo Deandrea et al. "Single production of vector-like quarks: the effects of large width, interference and NLO corrections". In: *Journal of High Energy Physics* 2021.8 (2021), pp. 1–63. DOI: https://doi.org/10.1007/JHEP08(2021)107.

[148] John M. Campbell et al. "NLO predictions for t-channel production of single top and fourth generation quarks at hadron colliders". In: *Journal of High Energy Physics* 2009.10 (Oct. 2009), p. 042. DOI: 10.1088/1126-6708/2009/10/042. URL: https://dx.doi.org/10.1088/1126-6708/2009/10/042.

[149] Giacomo Cacciapaglia et al. "Next-to-leading-order predictions for single vector-like quark production at the LHC". In: *Physics Letters B* 793 (2019), pp. 206–211. ISSN: 0370-2693. DOI: https://doi.org/10.1016/j.physletb.2019.04.056. URL: https://www.sciencedirect.com/science/article/pii/S0370269319302898.

[150] Albert M Sirunyan et al. "Search for bottom-type, vectorlike quark pair production in a fully hadronic final state in proton-proton collisions at $\sqrt{s} = 13$ TeV". In: *Phys. Rev. D* 102 (11 Dec. 2020), p. 112004. DOI: 10.1103/PhysRevD.102.112004. URL: https://link.aps.org/doi/10.1103/PhysRevD.102.112004.

[151] Xiao-Min Cui, Yu-Qi Li, and Yao-Bei Liu. "Search for pair production of the heavy vectorlike top partner in same-sign dilepton signature at the HL-LHC". In: *Phys. Rev. D* 106 (11 Dec. 2022), p. 115025. DOI: 10.1103/PhysRevD.106.115025. URL: https://link.aps.org/doi/10.1103/PhysRevD.106.115025.

[152] Arpon Paul, Sezen Sekmen, and Gokhan Unel. "Down type iso-singlet quarks at the HL-LHC and FCC-hh". In: *The European Physical Journal C* 81 (2021), pp. 1–14. DOI: https://doi.org/10.1140/epjc/s10052-021-08982-4.

[153] Xi-Yan Tian, Liu-Feng Du, and Yao-Bei Liu. "Search for single production of vector like top partners through th channel at the HE-LHC and FCC-hh". In: *The European Physical Journal C* 81 (2021), pp. 1–9. DOI: https://doi.org/10.1140/epjc/s10052-021-09385-1.

[154] Jin-Zhong Han et al. "Single production of vector like B quarks at the CLIC". In: *Phys. Rev. D* 105 (1 Jan. 2022), p. 015005. DOI: 10.1103/PhysRevD.105.015005. URL: https://link.aps.org/doi/10.1103/PhysRevD.105.015005.

[155] Albert M Sirunyan et al. "Search for pair production of vectorlike quarks in the fully hadronic final state". In: *Phys. Rev. D* 100 (7 Oct. 2019), p. 072001. DOI: 10.1103/PhysRevD.100.072001. URL: https://link.aps.org/doi/10.1103/PhysRevD.100.072001.

[156] M Aaboud et al. "Search for pair production of heavy vectorlike quarks decaying into hadronic final states in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector". In: *Phys. Rev. D* 98 (9 Nov. 2018), p. 092005. DOI: 10.1103/PhysRevD.98.092005. URL: https://link.aps.org/doi/10.1103/PhysRevD.98.092005.

[157] Georges Aad et al. "Search for pair-produced vector-like top and bottom partners in events with large missing transverse momentum in pp collisions with the ATLAS detector". In: *The European Physical Journal C* 83.8 (2023), p. 719. DOI: https://doi.org/10.1140/epjc/s10052-023-11790-7.

[158] The CMS Collaboration. "Search for vector-like quarks in events with two oppositely charged leptons and jets in proton-proton collisions at $\sqrt{s} = 13$ TeV". In: (2018). DOI: 10.1140/epjc/s10052-019-6855-8.

[159] Morad Aaboud et al. "Search for pair and single production of vector like quarks in final states with at least one Z boson decaying into a pair of electrons or muons in pp collision data collected with the ATLAS detector at $\sqrt{s} = 13$ TeV". In: *Phys. Rev. D* 98 (11 Dec. 2018), p. 112010. DOI: 10.1103/PhysRevD.98.112010. URL: https://link.aps.org/doi/10.1103/PhysRevD.98.112010.

[160] The ATLAS Collaboration. "Search for new phenomena in events with same-charge leptons and b-jets in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector". In: (2018). DOI: 10.1007/JHEP12(2018)039.

[161] The CMS Collaboration. "Search for pair production of vector-like quarks in leptonic final states in proton-proton collisions at $\sqrt{s} = 13$ TeV". In: *Journal of High Energy Physics* 2023.7 (July 2023), p. 20. ISSN: 1029-8479. DOI: 10.1007/JHEP07(2023)020. URL: https://doi.org/10.1007/JHEP07(2023)020.

[162] Tiago Vale. "Search for heavy fermions with LHC data". PhD thesis. Universidade do Minho, July 2021. URL: https://cds.cern.ch/record/2788832.

[163] G Aad et al. "Search for pair-production of vector-like quarks in pp collision events at $\sqrt{s} = 13$ TeV with at least one leptonically decaying Z boson and a third-generation quark with the ATLAS detector". In: *Physics Letters B* 843 (2023), p. 138019. ISSN: 0370-2693. DOI: https://doi.org/10.1016/j.physletb.2023.138019. URL: https://www.sciencedirect.com/science/article/pii/S0370269323003532.

[164] Albert M Sirunyan et al. "Search for electroweak production of a vector-like T quark using fully hadronic final states". In: *Journal of High Energy Physics* 2020.1 (2020), pp. 1–64. DOI: https://doi.org/10.1007/JHEP01(2020)036.

[165] The CMS Collaboration. "Search for production of a single vectorlike quark decaying to tH or tZ in the all-hadronic final state in pp collisions at $\sqrt{s} = 13$ TeV". In: *Phys. Rev. D* 110 (7 Oct. 2024), p. 072012. DOI: 10.1103/PhysRevD.110.072012. URL: https://link.aps.org/doi/10.1103/PhysRevD.110.072012.

[166] G Aad et al. "Search for single production of a vectorlike T quark decaying into a Higgs boson and top quark with fully hadronic final states using the ATLAS detector". In: *Phys. Rev. D* 105 (9 May 2022), p. 092012. DOI: 10.1103/PhysRevD.105.092012. URL: https://link.aps.org/doi/10.1103/PhysRevD.105.092012.

[167] Albert M Sirunyan et al. "Search for single production of vector-like quarks decaying to a b quark and a Higgs boson". In: *Journal of High Energy Physics* 2018.6 (June 2018), p. 31. ISSN: 1029-8479. DOI: 10.1007/JHEP06(2018)031. URL: https://doi.org/10.1007/JHEP06(2018)031.

[168] The ATLAS Collaboration. "Search for single vector-like B quark production and decay via $B \to bH(b\bar{b})$ in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector". In: *Journal of High Energy Physics* 2023.11 (Nov. 2023), p. 168. ISSN: 1029-8479. DOI: 10.1007/JHEP11(2023)168. URL: https://doi.org/10.1007/JHEP11(2023)168.

[169] Armen Tumasyan et al. "Search for a vector-like quark T' → tH via the diphoton decay mode of the Higgs boson in proton-proton collisions at $\sqrt{s} = 13$ TeV". In: *Journal of High Energy Physics* 2023.9 (2023), pp. 1–40. DOI: https://doi.org/10.1007/JHEP09(2023)057.

[170] The ATLAS Collaboration. "Search for single production of vector-like quarks decaying into Wb in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector". In: *Journal of High Energy Physics* 2019.5 (May 2019), p. 164. DOI: 10.1007/JHEP05(2019)164. URL: https://doi.org/10.1007/JHEP05(2019)164.

[171] The ATLAS Collaboration. "Search for single production of vector-like T quarks decaying into Ht or Zt in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector". In: *Journal of High Energy Physics* 2023.8 (Aug. 2023), p. 153. ISSN: 1029-8479. DOI: 10.1007/JHEP08(2023)153. URL: https://doi.org/10.1007/JHEP08(2023)153.

[172] The CMS Collaboration. "Search for single production of vector-like quarks decaying to a top quark and a W boson in proton-proton collisions at $\sqrt{s} = 13$ TeV". In: *The European Physical Journal C* 79.2 (Jan. 2019), p. 90. ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-019-6556-3. URL: https://doi.org/10.1140/epjc/s10052-019-6556-3.

[173] The CMS Collaboration. "Search for single production of a vector-like T quark decaying to a top quark and a Z boson in the final state with jets and missing transverse momentum at $\sqrt{s} = 13$ TeV". In: *Journal of High Energy Physics* 2022.5 (May 2022), p. 93. DOI: 10.1007/JHEP05(2022)093. URL: https://doi.org/10.1007/JHEP05(2022)093.

[174] The ATLAS Collaboraton. "Search for large missing transverse momentum in association with one top-quark in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector". In: *Journal of High Energy Physics* 2019.5 (May 2019), p. 41. ISSN: 1029-8479. DOI: 10.1007/JHEP05(2019)041. URL: https://doi.org/10.1007/JHEP05(2019)041.

[175] The ATLAS Collaboration. "Search for singly produced vector-like top partners in multilepton final states with 139 fb^{-1} of pp collision data at $\sqrt{s} = 13$ TeV with the ATLAS detector". In: (2023). URL: https://arxiv.org/abs/2307.07584.

[176] CERN. "CERN Annual report 2022". In: (2023). URL: https://cds.cern.ch/record/2857560/.

[177] Nicolas Chanon. "Observation des photons directs dans les premières données et préparation à la recherche du boson de Higgs dans l'expérience CMS au LHC (CERN)". PhD thesis. Oct. 2010. URL: https://theses.hal.science/tel-00598989.

[178] Dylan Apparu. "Recherche de nouvelles particules au long temps de vie dans l'expérience CMS". PhD thesis. Université de Strasbourg, June 2023. URL: https://theses.hal.science/tel-04457227.

[179] Polina Simkina. "Event reconstruction and analysis in CMS using artificial intelligence". PhD thesis. Université Paris-Saclay, Sept. 2023. URL: https://theses.hal.science/tel-04412128.

[180] Georges Charpak. "Nobel Physics Prize 1992". In: *CERN Courier* (1992), p. 1. URL: https://www.osti.gov/etdeweb/servlets/purl/22421208.

[181] D Decamp et al. "Determination of the number of light neutrino species". In: *Physics Letters B* 292.3 (1992), pp. 463–471. ISSN: 0370-2693. DOI: https://doi.org/10.1016/0370-2693(92)91204-M. URL: https://www.sciencedirect.com/science/article/pii/037026939291204M.

[182] ALEPH collaboration et al. "Precision electroweak measurements on the Z resonance". In: *Physics Reports* 427.5 (2006), pp. 257–454. ISSN: 0370-1573. DOI: https://doi.org/10.1016/j.physrep.2005.12.006. URL: https://www.sciencedirect.com/science/article/pii/S0370157305005119.

[183] O Brüning et al. "LHC Design Report (The LHC Main Ring vol 1) CERN-2004-003". In: (2004). URL: 10.5170/CERN-2004-003-V-1.

[184] G. Baur et al. "Production of antihydrogen". In: *Physics Letters B* 368.3 (1996), pp. 251–258. ISSN: 0370-2693. DOI: https://doi.org/10.1016/0370-2693(96)00005-6. URL: https://www.sciencedirect.com/science/article/pii/0370269396000056.

[185] "A new measurement of direct CP violation in two pion decays of the neutral kaon". In: *Physics Letters B* 465.1 (1999), pp. 335–348. ISSN: 0370-2693. DOI: https://doi.org/10.1016/S0370-2693(99)01030-8. URL: https://www.sciencedirect.com/science/article/pii/S0370269399010308.

[186] Lyndon Evans and Philip Bryant. "LHC Machine". In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08001. DOI: 10.1088/1748-0221/3/08/S08001. URL: https://dx.doi.org/10.1088/1748-0221/3/08/S08001.

[187] ATLAS Inner Detector. *Technical design report*. Tech. rep. CERN-LHCC-97-016, 1997, pp. 98–13. URL: https://cds.cern.ch/record/331063.

[188] Ewa Lopienska. "The CERN accelerator complex, layout in 2022. Complexe des accélérateurs du CERN en janvier 2022". In: (2022). URL: https://cds.cern.ch/record/2800984.

[189] The CMS Collaboration et al. *Physics Technical Design Report, Detector Performance and Software*. Tech. rep. CERN-LHCC-2006-001, 2006. URL: https://cds.cern.ch/record/922757.

[190] R Bailey and Paul Collier. *Standard filling schemes for various LHC operation modes*. Tech. rep. CERN-LHC-Project-Note-323, 2003. URL: https://cds.cern.ch/record/691782.

[191] L Bourhis, M Fontannaz, and J Ph Guillet. "Quark and gluon fragmentation functions into photons". In: *The European Physical Journal C-Particles and Fields* 2 (1998), pp. 529–537. DOI: https://doi.org/10.1007/s100529800708.

[192] The CMS Collaboration. "Public CMS Luminosity Information". In: (2024). URL: https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults#Summary_charts_of_luminosity.

[193] Burkhard Schmidt. "The High-Luminosity upgrade of the LHC: Physics and Technology Challenges for the Accelerator and the Experiments". In: *Journal of Physics: Conference Series* 706.2 (Apr. 2016), p. 022002. DOI: 10.1088/1742-6596/706/2/022002. URL: https://dx.doi.org/10.1088/1742-6596/706/2/022002.

[194] HiLumi Collaboration. "The HL-LHC project". In: (2024). URL: https://voisins.web.cern.ch/en/high-luminosity-lhc-hl-lhc.

[195] Roman Adolphi et al. "The CMS experiment at the CERN LHC". In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08004. DOI: 10.1088/1748-0221/3/08/S08004. URL: https://dx.doi.org/10.1088/1748-0221/3/08/S08004.

[196] Tai Sakuma. "Cutaway diagrams of CMS detector". In: *J. Phys.: Conf. Ser. 513 022032*. CMS-OUTREACH-2019-001. 2019. URL: https://cds.cern.ch/record/2665537.

[197] Izaak Neutelings. "CMS coordinate system". In: (2024). URL: https://tikz.net/axis3d_cms.

[198] Wolfgang Adam et al. “The CMS Phase-1 pixel detector upgrade”. In: *Journal of Instrumentation* 16.02 (Feb. 2021), P02027. DOI: `10.1088/1748-0221/16/02/P02027`. URL: `https://dx.doi.org/10.1088/1748-0221/16/02/P02027`.

[199] Laura Borrello et al. *Sensor design for the CMS silicon strip tracker*. Tech. rep. CERN-CMS-NOTE-2003-020, 2003. URL: `https://cds.cern.ch/record/687861`.

[200] The CMS Collaboration. “Description and performance of track and primary-vertex reconstruction with the CMS tracker”. In: *Journal of Instrumentation* 9.10 (Oct. 2014), P10009. DOI: `10.1088/1748-0221/9/10/P10009`. URL: `https://dx.doi.org/10.1088/1748-0221/9/10/P10009`.

[201] The CMS Collaboration et al. *The CMS electromagnetic calorimeter project: technical design report*. Tech. rep. 1997. URL: `https://cds.cern.ch/record/349375`.

[202] P Adzicet al. “Energy resolution of the barrel of the CMS Electromagnetic Calorimeter”. In: *Journal of Instrumentation* 2.04 (Apr. 2007), P04004. DOI: `10.1088/1748-0221/2/04/P04004`. URL: `https://dx.doi.org/10.1088/1748-0221/2/04/P04004`.

[203] M. Anfreville et al. “Laser monitoring system for the CMS lead tungstate crystal calorimeter”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 594.2 (2008), pp. 292–320. ISSN: 0168-9002. DOI: `https://doi.org/10.1016/j.nima.2008.01.104`. URL: `https://www.sciencedirect.com/science/article/pii/S0168900208001599`.

[204] Prasanna Kumar Siddireddy. *The CMS ECAL Trigger and DAQ system: electronics auto-recovery and monitoring*. Tech. rep. CMS-CR–2018-084, 2018. URL: `https://cds.cern.ch/record/2630618`.

[205] J Mans. *CMS technical design report for the phase 1 upgrade of the hadron calorimeter*. Tech. rep. CMS-TDR-010, 2012. URL: `https://cds.cern.ch/record/1481837`.

[206] The CMS Collaboration. “Performance of the CMS hadron calorimeter with cosmic ray muons and LHC beam data”. In: *Journal of Instrumentation* 5.03 (Mar. 2010), T03012. DOI: `10.1088/1748-0221/5/03/T03012`. URL: `https://dx.doi.org/10.1088/1748-0221/5/03/T03012`.

[207] G Acquistapace, CMS Collaboration, et al. *CMS, the magnet project: Technical design report*. Tech. rep. CERN-LHCC-97-010, 1997. URL: `https://cds.cern.ch/record/331056`.

[208] CMS collaboration et al. “Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV”. In: *Journal of Instrumentation* 13.06 (June 2018), P06015. DOI: `10.1088/1748-0221/13/06/P06015`. URL: `https://dx.doi.org/10.1088/1748-0221/13/06/P06015`.

[209] The CMS Collaboration. “The performance of the CMS muon detector in proton-proton collisions at $\sqrt{s} = 7$ TeV at the LHC”. In: *Journal of Instrumentation* 8.11 (Nov. 2013), P11002. DOI: `10.1088/1748-0221/8/11/P11002`. URL: `https://dx.doi.org/10.1088/1748-0221/8/11/P11002`.

[210] A Colaleo et al. *CMS technical design report for the muon endcap GEM upgrade*. Tech. rep. CERN-LHCC-2015-012, 2015. URL: `https://cds.cern.ch/record/2021453`.

[211] CMS collaboration et al. “Particle-flow reconstruction and global event description with the CMS detector”. In: *Journal of Instrumentation* 12.10 (Oct. 2017), P10003. DOI: `10.1088/1748-0221/12/10/P10003`. URL: `https://dx.doi.org/10.1088/1748-0221/12/10/P10003`.

[212] T. Speer et al. “Track reconstruction in the CMS tracker”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 559.1 (2006), pp. 143–147. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2005.11.207. URL: https://www.sciencedirect.com/science/article/pii/S0168900205022576.

[213] The CMS Collaboration. “CMS Tracking POG Performance Plots For 2017 with PhaseI pixel detector”. In: (2023). URL: https://twiki.cern.ch/twiki/bin/view/CMSPublic/TrackingPOGPerformance2017MC.

[214] Albert M Sirunyan et al. “Performance of the reconstruction and identification of high-momentum muons in proton-proton collisions at $\sqrt{s} = 13$ TeV”. In: *Journal of Instrumentation* 15.02 (Feb. 2020), P02027. DOI: 10.1088/1748-0221/15/02/P02027. URL: https://dx.doi.org/10.1088/1748-0221/15/02/P02027.

[215] CMS collaboration et al. “Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC”. In: *Journal of Instrumentation* 16.05 (May 2021), P05014. DOI: 10.1088/1748-0221/16/05/P05014. URL: https://dx.doi.org/10.1088/1748-0221/16/05/P05014.

[216] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. “The anti-kt jet clustering algorithm”. In: *Journal of High Energy Physics* 2008.04 (Apr. 2008), p. 063. DOI: 10.1088/1126-6708/2008/04/063. URL: https://dx.doi.org/10.1088/1126-6708/2008/04/063.

[217] The CMS Collaboration et al. “Jet energy scale and resolution measurement with Run 2 Legacy Data Collected by CMS at 13 TeV”. In: *CMS Detector Performance Summary CMS-DP-2021-033, CERN* (2021). URL: https://cds.cern.ch/record/2792322.

[218] Reuven Y Rubinstein and Dirk P Kroese. *Simulation and the Monte Carlo method.* John Wiley & Sons, 2016.

[219] Johan Alwall et al. “The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations”. In: *Journal of High Energy Physics* 2014.7 (2014), pp. 1–157. DOI: https://doi.org/10.1007/JHEP07(2014)079.

[220] Stefano Frixione, Paolo Nason, and Carlo Oleari. “Matching NLO QCD computations with parton shower simulations: the POWHEG method”. In: *Journal of High Energy Physics* 2007.11 (2007), p. 070. DOI: 10.1088/1126-6708/2007/11/070.

[221] Simone Alioli et al. “A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX”. In: *Journal of High Energy Physics* 2010.6 (2010), pp. 1–58. DOI: https://doi.org/10.1007/JHEP06(2010)043.

[222] J. Alwall et al. “A standard format for Les Houches Event Files”. In: *Computer Physics Communications* 176.4 (2007), pp. 300–304. ISSN: 0010-4655. DOI: https://doi.org/10.1016/j.cpc.2006.11.010. URL: https://www.sciencedirect.com/science/article/pii/S0010465506004164.

[223] Christian Bierlich et al. “A comprehensive guide to the physics and usage of PYTHIA 8.3”. In: *SciPost Phys. Codebases* (2022), p. 8. DOI: 10.21468/SciPostPhysCodeb.8. URL: https://scipost.org/10.21468/SciPostPhysCodeb.8.

[224] The CMS Collaboration. “The CMS Offline WorkBook”. In: (2023). URL: https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBook.

[225] Rene Brun and Fons Rademakers. “ROOT — An object oriented data analysis framework”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 389.1 (1997), pp. 81–86. ISSN: 0168-9002. DOI: `https://doi.org/10.1016/S0168-9002(97)00048-X`. URL: `https://www.sciencedirect.com/science/article/pii/S016890029700048X`.

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

[227] A. Fagot et al. “Ramp;D towards the CMS RPC Phase-2 upgrade”. In: *Journal of Instrumentation* 11.09 (Sept. 2016), p. C09017. DOI: `10.1088/1748-0221/11/09/C09017`. URL: `https://dx.doi.org/10.1088/1748-0221/11/09/C09017`.

[228] K Klein et al. *The phase-2 upgrade of the CMS tracker.* Tech. rep. CERN-LHCC-2017-009, 2021. URL: `https://cds.cern.ch/record/2272264`.

[229] The CMS Collaboration. *A MIP timing detector for the CMS phase-2 upgrade.* Tech. rep. CERN-LHCC-2019-003, 2019. URL: `https://cds.cern.ch/record/2667167`.

[230] Didier Contardo and Austin Ball. *The Phase-2 Upgrade of the CMS Barrel Calorimeters.* Tech. rep. CERN-LHCC-2017-011, 2017. URL: `https://cds.cern.ch/record/2283187`.

[231] Didier Contardo and Austin Ball. *The Phase-2 Upgrade of the CMS Endcap Calorimeter.* Tech. rep. CERN-LHCC-2017-023, 2017. URL: `https://cds.cern.ch/record/2293646`.

[232] Arabella Martelli. “The CMS HGCal detector for HL-LHC upgrade”. In: *arXiv preprint arXiv:1708.08234* (2017). DOI: `https://doi.org/10.48550/arXiv.1708.08234`.

[233] Didier Contardo and Austin Ball. *The Phase-2 Upgrade of the CMS Muon Detectors.* Tech. rep. CERN-LHCC-2017-012, 2017. URL: `https://cds.cern.ch/record/2283189`.

[234] The CMS Collaboration. “The CMS trigger system”. In: *Journal of Instrumentation* 12.01 (Jan. 2017), P01020. DOI: `10.1088/1748-0221/12/01/P01020`. URL: `https://dx.doi.org/10.1088/1748-0221/12/01/P01020`.

[235] Marino Missiroli. “ConfDB GUI”. In: (2022). URL: `https://twiki.cern.ch/twiki/bin/view/CMS/EvfConfDBGUI`.

[236] The CMS Collaboration. “Pileup mitigation at CMS in 13 TeV data”. In: *Journal of Instrumentation* 15.09 (Sept. 2020), P09018. DOI: `10.1088/1748-0221/15/09/P09018`. URL: `https://dx.doi.org/10.1088/1748-0221/15/09/P09018`.

[237] The CMS Collaboration. In: *Journal of Instrumentation* 12.10 (Oct. 2017), P10003. DOI: `10.1088/1748-0221/12/10/P10003`. URL: `https://dx.doi.org/10.1088/1748-0221/12/10/P10003`.

[238] Daniele Bertolini et al. “Pileup per particle identification”. In: *Journal of High Energy Physics* 2014.10 (2014), pp. 1–22. DOI: `https://doi.org/10.1007/JHEP10(2014)059`.

[239] The CMS Collaboration. “Pileup-per-particle identification: optimisation for Run 2 Legacy and beyond”. In: (2021). URL: `https://cds.cern.ch/record/2751563`.

[240] The CMS Collaboration. “Jet algorithms performance in 13 TeV data”. In: (2016). URL: `https://cds.cern.ch/record/2256875`.

[241] The CMS Collaboration. “Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV”. In: *Journal of Instrumentation* 13.05 (May 2018), P05011. DOI: `10.1088/1748-0221/13/05/P05011`. URL: `https://dx.doi.org/10.1088/1748-0221/13/05/P05011`.

[242] Daniel Guest et al. “Jet flavor classification in high-energy physics with deep neural networks”. In: *Phys. Rev. D* 94 (11 Dec. 2016), p. 112002. DOI: 10.1103/PhysRevD.94.112002. URL: https://link.aps.org/doi/10.1103/PhysRevD.94.112002.

[243] E. Bols et al. “Jet flavour classification using DeepJet”. In: *Journal of Instrumentation* 15.12 (Dec. 2020), P12012. DOI: 10.1088/1748-0221/15/12/P12012. URL: https://dx.doi.org/10.1088/1748-0221/15/12/P12012.

[244] Jieun Choi. “HLT Efficiency measurement in 5 (2 b-tagged) jet”. In: (2022). URL: https://indico.in2p3.fr/event/28472/contributions/116857/attachments/74640/107645/VHF_5bjetHLTEfficiency_Jieun_20221128.pdf.

[245] Huilin Qu and Loukas Gouskos. “Jet tagging via particle clouds”. In: *Phys. Rev. D* 101 (5 Mar. 2020), p. 056019. DOI: 10.1103/PhysRevD.101.056019. URL: https://link.aps.org/doi/10.1103/PhysRevD.101.056019.

[246] The CMS Collaboration. “Performance of the ParticleNet tagger on small and large-radius jets at High Level Trigger in Run 3”. In: (2023). URL: https://cds.cern.ch/record/2857440.

[247] Jieun Choi. “B2G Trigger Review”. In: (2024). URL: https://indico.cern.ch/event/1344493/contributions/5658953/attachments/2787339/4860037/B2GHLT_Review_25Jan2024.pdf.

[248] The CMS Collaboration. “Tau identification in CMS during LHC Run 2”. In: (2021). URL: https://cds.cern.ch/record/2797703.

[249] The CMS Collaboration. “How to work with files for Good Luminosity Sections in JSON format”. In: (2023). URL: https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideGoodLumiSectionsJSONFile.

[250] The CMS Collaboration. *Simulation of the Silicon Strip Tracker pre-amplifier in early 2016 data*. Tech. rep. CMS-DP-2020-045, 2020. URL: https://cds.cern.ch/record/2740688.

[251] Luca Panizzi and Alexandra Carvalho. “Notes on single B production”. In: (2017). URL: https://github.com/CrossSectionsLHC/TopPartners_SingleProduction/blob/master/parametrization_general.pdf.

[252] Luca Panizzi and Alexandra Carvalho. “Cross sections and Branching Ratios for the single T’ in the Narrow-Width Approximation”. In: (2017). URL: https://github.com/CrossSectionsLHC/TopPartners_SingleProduction/blob/master/SingleT/CrossSections_BR_SingleT_NWA.csv.

[253] Albert Sirunyan et al. “Extraction and validation of a new set of CMS pythia8 tunes from underlying-event measurements”. In: *European Physical Journal C: Particles and Fields* 80.1 (2020). DOI: https://doi.org/10.1140/epjc/s10052-019-7499-4.

[254] CMS Top PAG. *Run-2 UltraLegacy Datasets for Analysis*. 2023. URL: https://twiki.cern.ch/twiki/bin/view/CMS/TopTrigger.

[255] Rikkert Frederix and Stefano Frixione. “Merging meets matching in MC@NLO”. In: *Journal of High Energy Physics* 2012.12 (2012), pp. 1–41. DOI: https://doi.org/10.1007/JHEP12(2012)061.

[256] Anna Kulesza et al. “Associated top quark pair production with a heavy boson: differential cross sections at NLO+NNLL accuracy”. In: *The European Physical Journal C* 80.5 (2020), p. 428. DOI: https://doi.org/10.1140/epjc/s10052-020-7987-6.

[257] Frederix Rikkert and Tsinikos Ioannis. “On improving NLO merging for ttW production”. In: *Journal of High Energy Physics* 2021.11 (2021). DOI: https://doi.org/10.1007/JHEP11(2021)029.

[258] Johan Alwall et al. “Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions”. In: *The European Physical Journal C* 53 (2008), pp. 473–500. DOI: https://doi.org/10.1140/epjc/s10052-007-0490-5.

[259] Stefano Frixione, Giovanni Ridolfi, and Paolo Nason. “A positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction”. In: *Journal of High Energy Physics* 2007.09 (2007), p. 126. DOI: 10.1088/1126-6708/2007/09/126.

[260] Michał Czakon and Alexander Mitov. “Top++: A program for the calculation of the top-pair cross-section at hadron colliders”. In: *Computer Physics Communications* 185.11 (2014), pp. 2930–2938. ISSN: 0010-4655. DOI: https://doi.org/10.1016/j.cpc.2014.06.021. URL: https://www.sciencedirect.com/science/article/pii/S0010465514002264.

[261] Top LHC Working Group. “NNLO+NNLL top-quark-pair cross sections. ATLAS-CMS recommended predictions for top-quark-pair cross sections using the Top++v2.0 program (M. Czakon, A. Mitov, 2013)”. In: (2023). URL: https://twiki.cern.ch/twiki/bin/view/LHCPhysics/TtbarNNLO.

[262] Paolo Nason and Giulia Zanderighi. “W+ W-, WZ and ZZ production in the powheg-BOX-V2”. In: *The European Physical Journal C* 74.1 (2014), p. 2702. DOI: https://doi.org/10.1140/epjc/s10052-013-2702-5.

[263] T. Gehrmann et al. “W^+W^- Production at Hadron Colliders in Next to Next to Leading Order QCD”. In: *Phys. Rev. Lett.* 113 (21 Nov. 2014), p. 212001. DOI: 10.1103/PhysRevLett.113.212001. URL: https://link.aps.org/doi/10.1103/PhysRevLett.113.212001.

[264] S Bolognesi, Y Gao, AV Gritsan, et al. “On the spin and parity of a single-produced resonance at the LHC Preprint”. In: *arXiv preprint arXiv:1208.4018* (2012). DOI: https://doi.org/10.1103/PhysRevD.86.095031.

[265] Paolo Nason. “A new method for combining NLO QCD with shower Monte Carlo algorithms”. In: *Journal of High Energy Physics* 2004.11 (2004), p. 040. DOI: 10.1088/1126-6708/2004/11/040.

[266] Heribertus B Hartanto et al. “Higgs boson production in association with top quarks in the POWHEG BOX”. In: *Physical Review D* 91.9 (2015), p. 094003. DOI: 10.1103/PhysRevD.91.094003.

[267] Nikolaos Kidonakis and Nodoka Yamanaka. “Higher-order corrections for tW production at high-energy hadron colliders”. In: *JHEP* 05 (2021), p. 278. DOI: 10.1007/JHEP05(2021)278. arXiv: 2102.11300 [hep-ph].

[268] Top LHC Working Group. “Predictions at NNLO of single top-quark production cross-sections”. In: (2022). URL: https://twiki.cern.ch/twiki/bin/view/LHCPhysics/SingleTopNNLORef.

[269] CMS Top PAG. “Modelling of the single top-quark production in association with the Higgs boson at 13 TeV”. In: (2016). URL: https://twiki.cern.ch/twiki/bin/viewauth/CMS/SingleTopHiggsGeneration13TeV.

[270] Ryan Gavin et al. “FEWZ 2.0: a code for hadronic Z production at next-to-next-to-leading order”. In: *Computer Physics Communications* 182.11 (2011), pp. 2388–2403. DOI: https://doi.org/10.1016/j.cpc.2011.06.008.

[271] CMS Higgs PAG. “Summary table of samples produced for the 1 Billion campaign, with 25ns bunch-crossing”. In: (2024). URL: https://twiki.cern.ch/twiki/bin/viewauth/CMS/SummaryTable1G25ns.

[272] The CMS Collaboration. "ECAL Clustering". In: (2009). URL: https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideEcalRecoClustering.

[273] CMS EGamma POG. "Cut Based Electron ID for Run 2". In: (2023). URL: https://twiki.cern.ch/twiki/bin/view/CMS/CutBasedElectronIdentificationRun2.

[274] The CMS Collaboration. "Search for physics beyond the standard model in events with jets and two same-sign or at least three charged leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV". In: *Eur. Phys. J. C* 80 (2020), p. 752. DOI: https://doi.org/10.1140/epjc/s10052-020-8168-3.

[275] The CMS Collaboration. "Search for physics beyond the standard model in events with two leptons of same sign, missing transverse momentum, and jets in proton–proton collisions at $\sqrt{s} = 13$ TeV". In: *Eur. Phys. J. C* 77 (2017), p. 578. DOI: https://doi.org/10.1140/epjc/s10052-017-5079-z.

[276] The CMS Collaboration et al. "Search for production of four top quarks in final states with same-sign or multiple leptons in proton–proton collisions at $\sqrt{s} = 13$ TeV". In: (2020). DOI: https://doi.org/10.1140/epjc/s10052-019-7593-7.

[277] Edward Ryan Heller. "A search for supersymmetry with the CMS detector in the single-lepton final state using the sum of masses of large-radius jets". PhD thesis. University of California Santa Barbara, 2018. URL: https://escholarship.org/content/qt98x7b1rd/qt98x7b1rd_noSplash_9ca99841c75df59604dcf7a2939cb87f.pdf.

[278] CMS Muon POG. "Baseline muon selections for Run-II". In: (2022). URL: https://twiki.cern.ch/twiki/bin/view/CMS/SWGuideMuonIdRun2.

[279] CMS JetMET POG. "Jet Identification". In: (2024). URL: https://twiki.cern.ch/twiki/bin/view/CMS/JetID.

[280] CMS JetMET POG. "Recommended Jet Energy Corrections and Uncertainties For Data and MC". In: (2023). URL: https://twiki.cern.ch/twiki/bin/view/CMS/JECDataMC.

[281] CMS BTV POG. "BTV Scale Factors". In: (2023). URL: https://btv-wiki.docs.cern.ch/ScaleFactors.

[282] The CMS Collaboration. "Tag and Probe". In: (2014). URL: https://twiki.cern.ch/twiki/bin/view/CMSPublic/TagAndProbe.

[283] CMS EGamma POG. "Electron efficiencies and scale factors". In: (2020). URL: https://twiki.cern.ch/twiki/bin/view/CMS/EgammaIDRecipesRun2.

[284] CMS Muon POG. "Muon recommendations for 2016 Ultra Legacy data and Monte Carlo". In: (2023). URL: https://twiki.cern.ch/twiki/bin/view/CMS/MuonUL2016.

[285] CMS Muon POG. "Muon recommendations for 2017 Ultra Legacy data and Monte Carlo". In: (2023). URL: https://twiki.cern.ch/twiki/bin/view/CMS/MuonUL2017.

[286] CMS Muon POG. "Muon recommendations for 2018 Ultra Legacy data and Monte Carlo". In: (2023). URL: https://twiki.cern.ch/twiki/bin/view/CMS/MuonUL2018.

[287] Garvita Agarwal. "Jet Energy Scale and Resolution Measurements in CMS". In: *arXiv preprint arXiv:2301.02175* (2023). DOI: https://doi.org/10.22323/1.414.0652.

[288] CMS JetMET POG. "Recommended Jet Energy Corrections and Uncertainties For Data and MC". In: (2023). URL: https://twiki.cern.ch/twiki/bin/view/CMS/JECDataMC.

[289] A.M. Sirunyan et al. "Performance of missing transverse momentum reconstruction in proton-proton collisions at $\sqrt{s} = 13$ TeV using the CMS detector". In: *Journal of Instrumentation* 14.07 (July 2019), P07004. DOI: 10.1088/1748-0221/14/07/P07004. URL: https://dx.doi.org/10.1088/1748-0221/14/07/P07004.

[290] CMS BTV POG. “Methods to apply b-tagging efficiency scale factors”. In: (2022). URL: https://twiki.cern.ch/twiki/bin/view/CMS/BTagSFMethods.

[291] CMS BTV POG. “Approved b-tagging scale factors measurements for Run2 data”. In: (2024). URL: https://gitlab.cern.ch/cms-nanoAOD/jsonpog-integration/-/tree/master/POG/BTV.

[292] CMS JetMET POG. “Noise Filter Recommendations for Run II & Run III”. In: (2024). URL: https://twiki.cern.ch/twiki/bin/view/CMS/MissingETOptionalFiltersRun2.

[293] Giovanni Punzi. “Sensitivity of searches for new signals and its optimization”. In: *arXiv preprint physics/0308063* (2003). DOI: https://doi.org/10.48550/arXiv.physics/0308063.

[294] CMS Higgs PAG. “L1 Trigger Check for L=1e31 HLT trigger menu”. In: (2014). URL: https://twiki.cern.ch/twiki/bin/view/Sandbox/LeonardApanasevichL1TriggerCheck#Task.

[295] Barnet Woolf. “The log likelihood ratio test (the G-test)”. In: *Annals of human genetics* 21.4 (1957), pp. 397–409. DOI: https://doi.org/10.1111/j.1469-1809.1972.tb00293.x.

[296] “TH1 Class Reference”. In: (). URL: https://root.cern.ch/doc/master/classTH1.html#adb27b38cd934622c8702c402bb840b23.

[297] Atlas Collaboration et al. *Performance of mass-decorrelated jet substructure observables for hadronic two-body decay tagging in ATLAS*. Tech. rep. 2018. URL: https://cds.cern.ch/record/2630973.

[298] Nathan Lalloué. “Recherche d’un secteur sombre avec le détecteur ATLAS”. PhD thesis. Université Grenoble Alpes [2020-....], Oct. 2022. URL: https://theses.hal.science/tel-03966197.

[299] CMS collaboration et al. “Measurement of the Higgs boson production rate in association with top quarks in final states with electrons, muons, and hadronically decaying tau leptons at $\sqrt{s} = 13$ TeV”. In: *arXiv preprint arXiv:2011.03652* (2020). DOI: https://doi.org/10.1140/epjc/s10052-021-09014-x.

[300] Javier Andrés Brochero Cifuentes et al. “Measurement of the cross section of top quark-antiquark pair production in association with a W boson in proton-proton collisions at $\sqrt{s} = 13$ TeV”. In: (2023). DOI: 10.1007/JHEP07(2023)219. URL: https://doi.org/10.1007/JHEP07(2023)219.

[301] Aram Hayrapetyan et al. “Observation of four top quark production in proton-proton collisions at s= 13TeV”. In: *Physics Letters B* 847 (2023), p. 138290. DOI: https://doi.org/10.1016/j.physletb.2023.138290.

[302] CMS TSG POG. “Reweighting recipe to emulate Level 1 ECAL and Muon prefiring”. In: (2021). URL: https://twiki.cern.ch/twiki/bin/view/CMS/L1PrefiringWeightRecipe.

[303] CMS Top PAG. “The modeling of the top quark pT”. In: (2023). URL: https://twiki.cern.ch/twiki/bin/view/CMS/TopPtReweighting.

[304] The CMS Collaboration et al. “Search for flavor-changing neutral current interactions of the top quark mediated by a Higgs boson in proton-proton collisions at 13 TeV”. In: (2024). DOI: https://doi.org/10.48550/arXiv.2407.15172.

[305] CMS Lumi POG. “Luminosity recommendations for Run 2 analyses”. In: (2023). URL: https://twiki.cern.ch/twiki/bin/view/CMS/LumiRecommendationsRun2.

[306] CMS PVT POG. “Utilities for Accessing Pileup Information for Data”. In: (2024). URL: https://twiki.cern.ch/twiki/bin/view/CMS/PileupJSONFileforData.

[307] The CMS collaboration. “Determination of jet energy calibration and transverse momentum resolution in CMS”. In: *Journal of Instrumentation* 6.11 (Nov. 2011), P11002. DOI: `10.1088/1748-0221/6/11/P11002`. URL: `https://dx.doi.org/10.1088/1748-0221/6/11/P11002`.

[308] CMS JetMET POG. “Jet energy scale uncertainty sources”. In: (2022). URL: `https://twiki.cern.ch/twiki/bin/view/CMS/JECUncertaintySources`.

[309] Richard D Ball et al. “Parton distributions from high-precision collider data: NNPDF Collaboration”. In: *The European Physical Journal C* 77 (2017), pp. 1–75. DOI: `https://doi.org/10.1140/epjc/s10052-017-5199-5`.

[310] Jon Butterworth et al. “PDF4LHC recommendations for LHC Run II”. In: *Journal of Physics G: Nuclear and Particle Physics* 43.2 (Jan. 2016), p. 023001. DOI: `10.1088/0954-3899/43/2/023001`. URL: `https://dx.doi.org/10.1088/0954-3899/43/2/023001`.

[311] The CMS Collaboration et al. “Search for flavor-changing neutral current interactions of the top quark and the Higgs boson decaying to a bottom quark-antiquark pair at $\sqrt{s} = 13$ TeV”. In: *Journal of High Energy Physics* (2022). DOI: `https://doi.org/10.1007/JHEP02(2022)169`.

[312] Armen Tumasyan et al. “Measurement of the inclusive and differential WZ production cross sections, polarization angles, and triple gauge couplings in pp collisions at $\sqrt{s} = 13$ TeV”. In: *JHEP* 2207.CERN-EP-2021-163 (2021), p. 032. DOI: `https://doi.org/10.1007/JHEP07(2022)032`.

[313] Benedikt Biedermann, Ansgar Denner, and Mathieu Pellen. “Complete NLO corrections to W+ W+ scattering and its irreducible background at the LHC”. In: *Journal of High Energy Physics* 2017.10 (2017), pp. 1–24. DOI: `https://doi.org/10.1007/JHEP10(2017)124`.

[314] CMS collaboration et al. “The CMS statistical analysis and combination tool: COMBINE”. In: *arXiv preprint arXiv:2404.06614* (2024). DOI: `https://doi.org/10.48550/arXiv.2404.06614`.

[315] The CMS Collaboration. “The Combine framework”. In: (2024). URL: `https://cms-analysis.github.io/HiggsAnalysis-CombinedLimit/latest/`.

[316] Roger Barlow and Christine Beeston. “Fitting using finite Monte Carlo samples”. In: *Computer Physics Communications* 77.2 (1993), pp. 219–228. ISSN: 0010-4655. DOI: `https://doi.org/10.1016/0010-4655(93)90005-W`. URL: `https://www.sciencedirect.com/science/article/pii/001046559390005W`.

[317] J.S. Conway. “Incorporating nuisance parameters in likelihoods for multisource spectra”. In: *arXiv preprint arXiv:1103.0354* (2011). DOI: `https://doi.org/10.48550/arXiv.1103.0354`.

[318] CMS Higgs PAG. “Nuisance parameter impacts”. In: (2024). URL: `http://cms-analysis.github.io/HiggsAnalysis-CombinedLimit/latest/part3/nonstandard/#nuisance-parameter-impacts`.

[319] CMS Higgs PAG. “Likelihood Fits and Scans”. In: (2024). URL: `https://cms-analysis.github.io/HiggsAnalysis-CombinedLimit/part3/commonstatsmethods/?h=multidimfit#likelihood-fits-and-scans`.

[320] CMS Higgs PAG. “Asymptotic Frequentist Limits”. In: (2024). URL: `https://cms-analysis.github.io/HiggsAnalysis-CombinedLimit/latest/part3/commonstatsmethods/?h=asymptotic#asymptotic-frequentist-limits`.

[321] A.L. Read. "Presentation of search results: the CLs technique". In: *Journal of Physics G: Nuclear and Particle Physics* 28.10 (Sept. 2002), p. 2693. DOI: 10.1088/0954-3899/28/10/313. URL: https://dx.doi.org/10.1088/0954-3899/28/10/313.

[322] CMS Higgs PAG. "Asymptotic Significances". In: (2024). URL: http://cms-analysis.github.io/HiggsAnalysis-CombinedLimit/latest/part3/commonstatsmethods/?h=significance#asymptotic-significances.

[323] C.G Lester and D.J Summers. "Measuring masses of semi-invisibly decaying particle pairs produced at hadron colliders". In: *Physics Letters B* 463.1 (1999), pp. 99–103. ISSN: 0370-2693. DOI: https://doi.org/10.1016/S0370-2693(99)00945-4. URL: https://www.sciencedirect.com/science/article/pii/S0370269399009454.

[324] The CMS collaboration et al. "Search for a standard model-like Higgs boson in the mass range between 70 and 110 GeV in the diphoton final state in proton-proton collisions at $\sqrt{s} = 13$ TeV". In: (2024). DOI: https://doi.org/10.48550/arXiv.2405.18149.

[325] SN Gninenko, DS Gorbunov, and ME Shaposhnikov. "Search for GeV-Scale Sterile Neutrinos Responsible for Active Neutrino Oscillations and Baryon Asymmetry of the Universe". In: *Advances in High Energy Physics* 2012.1 (2012), p. 718259. DOI: https://doi.org/10.48550/arXiv.1301.5516.

[326] Adam Falkowski, David M Straub, and Avelino Vicente. "Vector-like leptons: Higgs decays and collider phenomenology". In: *Journal of High Energy Physics* 2014.5 (2014), pp. 1–18. DOI: https://doi.org/10.1007/JHEP05(2014)092.

[327] CMS EGamma POG. "Approved electron isolation and reconstruction scale factors measurements for Run2 data". In: (2023). URL: https://twiki.cern.ch/twiki/bin/view/CMS/EgammaUL2016To2018.

[328] CMS Muon POG. "Approved muon trigger, identification, isolation and reconstruction scale factors measurements for Run2 data". In: (2024). URL: https://gitlab.cern.ch/cms-muonPOG/muonefficiencies/-/tree/master/Run2/UL.

Appendix A

Trigger efficiencies and object corrections

A.1 Selection at the HLT level

The trigger efficiencies for both the single electron and single muon HLT trigger paths are presented from Figure A.1 to A.3 as a function of the p_T of the leading lepton, with 2016 preVFP, 2016 postVFP, and 2017 simulated and data samples respectively. The computation of trigger efficiencies is discussed in Chapter 4, and illustrated with 2018 simulated and data samples. The trigger MC efficiencies are calculated using the signal sample with a nominal mass of the T' $m_{T'} = 700$ GeV, while data efficiencies are calculated using the MET primary datasets provided by the CMS collaboration. Both efficiencies are calculated requiring 2 SS leptons in the offline selection as the first criterion of the preselection of the SS dilepton analysis, and the HLT_PFMET120_PFMHT120_IDTight and HLT_PFMET140_PFMHT140_IDTight paths define the reference HLT paths for 2016 and 2017 respectively.

A.2 Object corrections

A.2.1 Electron corrections

The electron SFs are presented from Figure A.4 to A.11 in the η-p_T plane of the electrons, with 2016 preVFP, 2016 postVFP, 2017, and 2018 samples. The computation of electron SFs is discussed in Chapter 4. The ID SFs are presented for both the loose and tight WPs, which are used in the definition of the fakeable and tight electrons respectively.

A.2.2 Muon corrections

The muon SFs are presented from Figure A.12 to A.23 in the η-p_T plane of the muons, with 2016 preVFP, 2016 postVFP, 2017, and 2018 samples. The computation of muon SFs is discussed in Chapter 4. The isolation SFs are presented for both the loose and tight WPs, which are used in the definition of the fakeable and tight muons respectively.

A.2.3 B-tagging corrections

The efficiencies ϵ_{b-tag} and ϵ_{light} are presented from Figure A.24 to A.26 in the p_T-η plane of the jets, with 2016 preVFP, 2016 postVFP, and 2017 simulated samples. The computation of these efficiencies are discussed in Chapter 4, and illustrated with 2018 simulated samples. They are calculated using the cut-based selection.

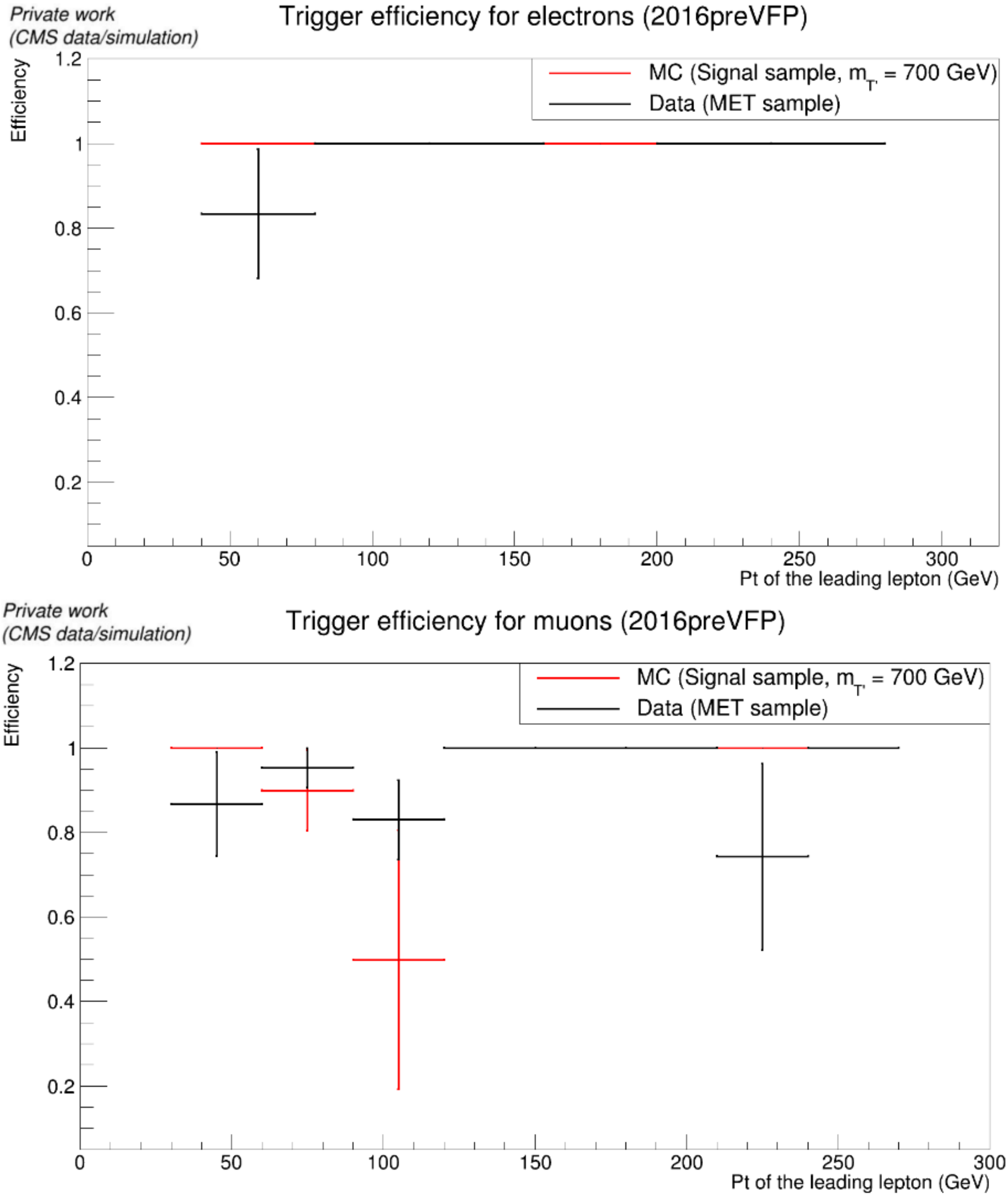


Figure A.1: Trigger efficiencies for both electron (top) and muon (bottom) HLT trigger paths, comparing 2016 preVFP simulation (red) and data (black) with $m_{T'} = 700$ GeV.

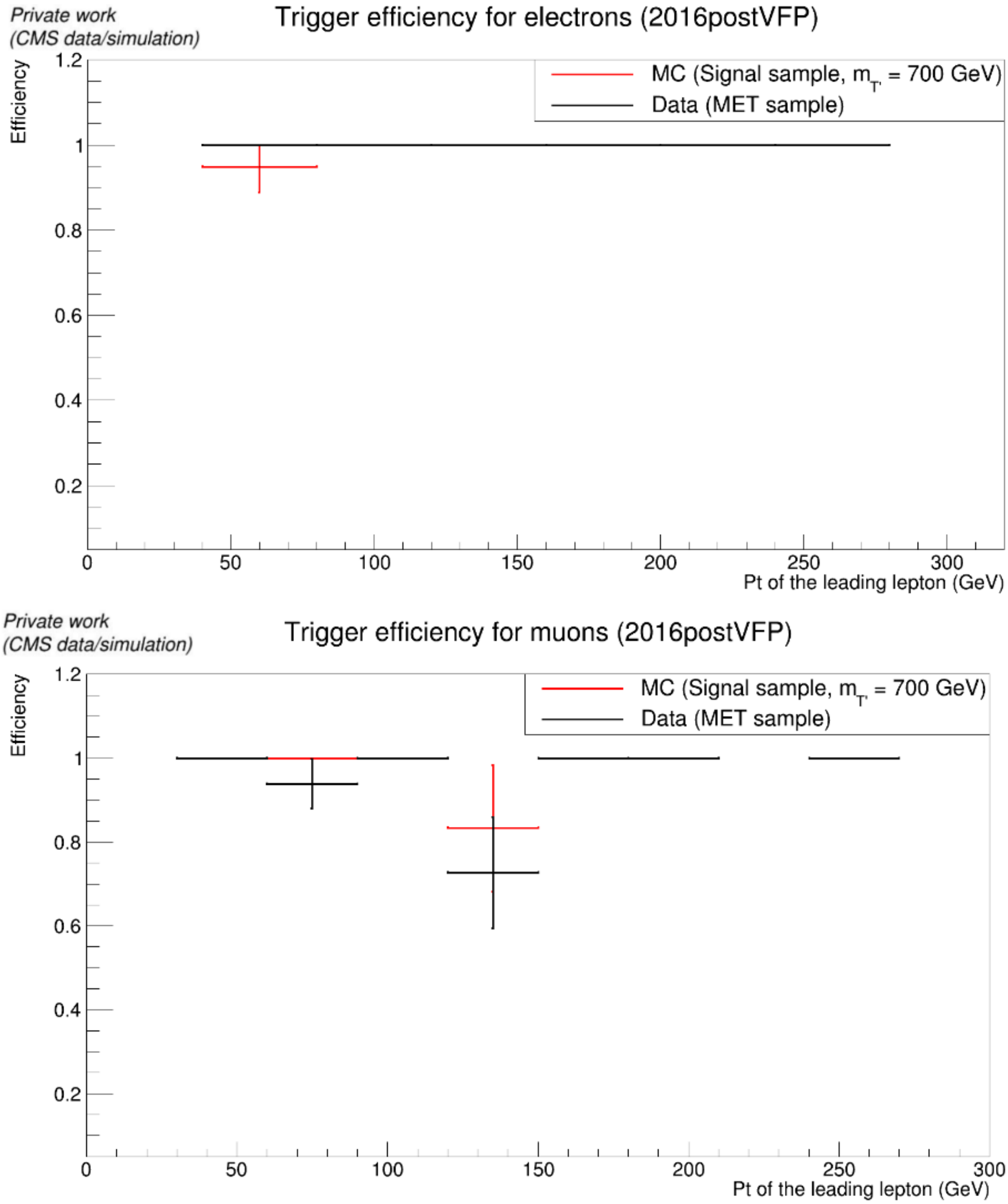


Figure A.2: Trigger efficiencies for both electron (top) and muon (bottom) HLT trigger paths, comparing 2016 postVFP simulation (red) and data (black) with $m_{T'} = 700$ GeV.

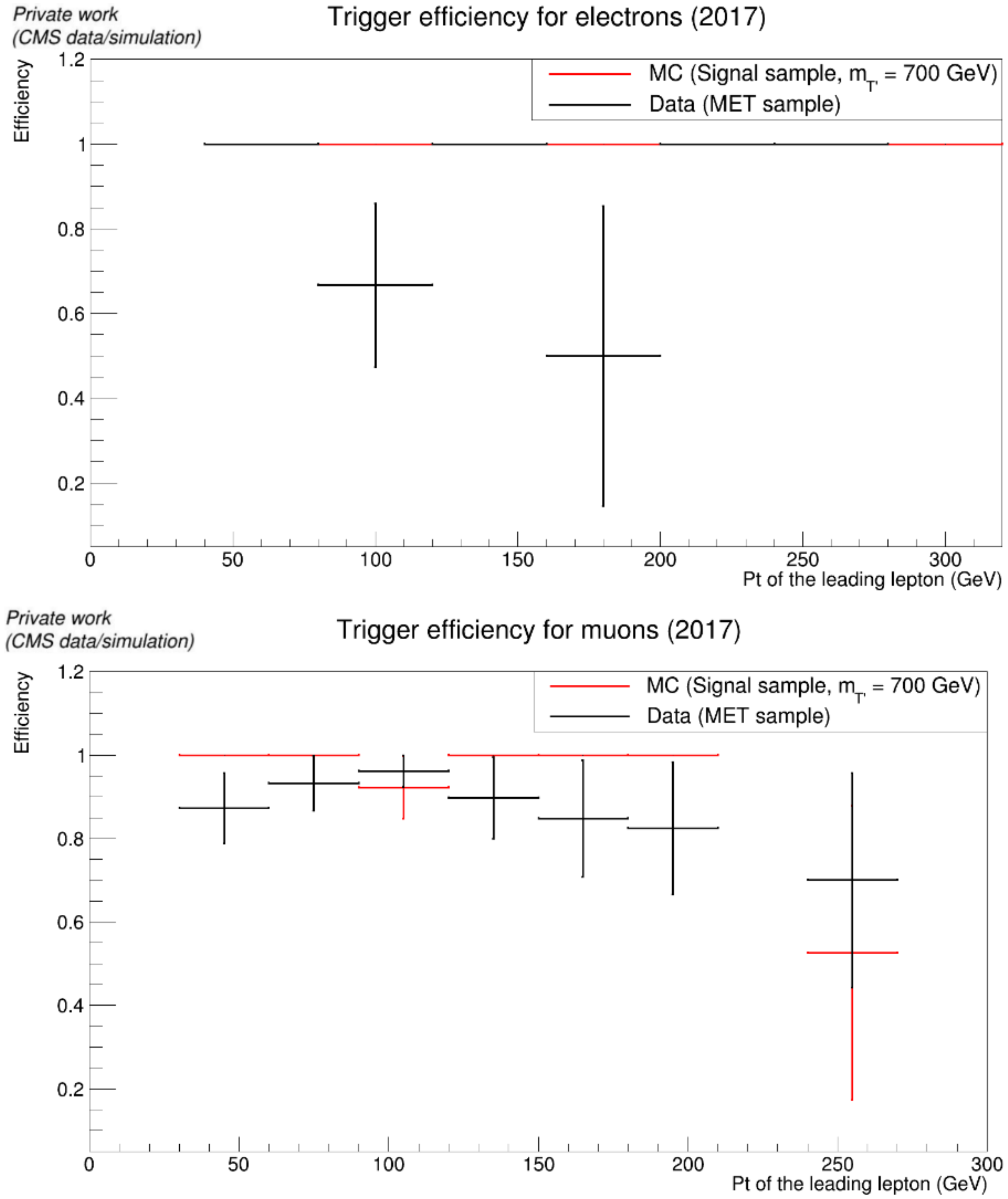
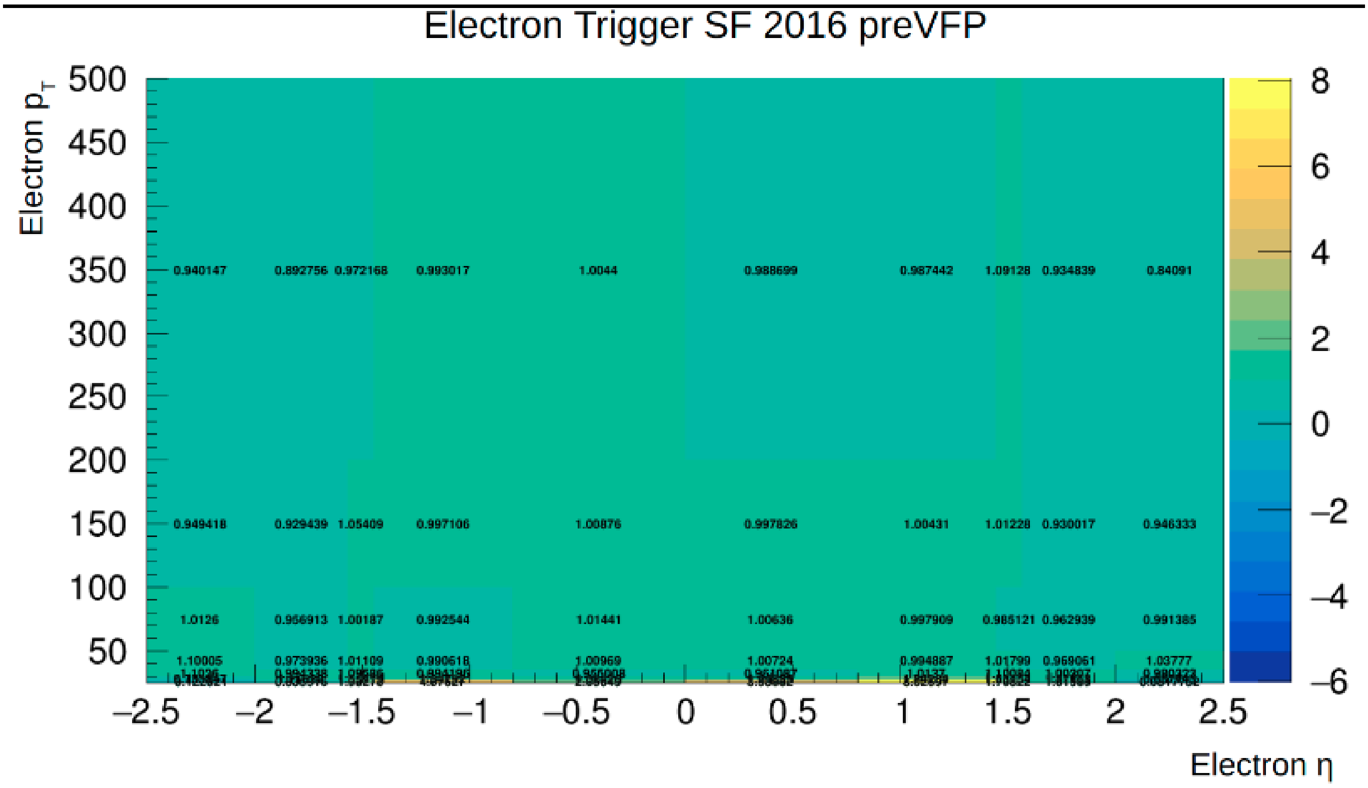


Figure A.3: Trigger efficiencies for both electron (top) and muon (bottom) HLT trigger paths, comparing 2017 simulation (red) and data (black) with $m_{T'} = 700$ GeV.



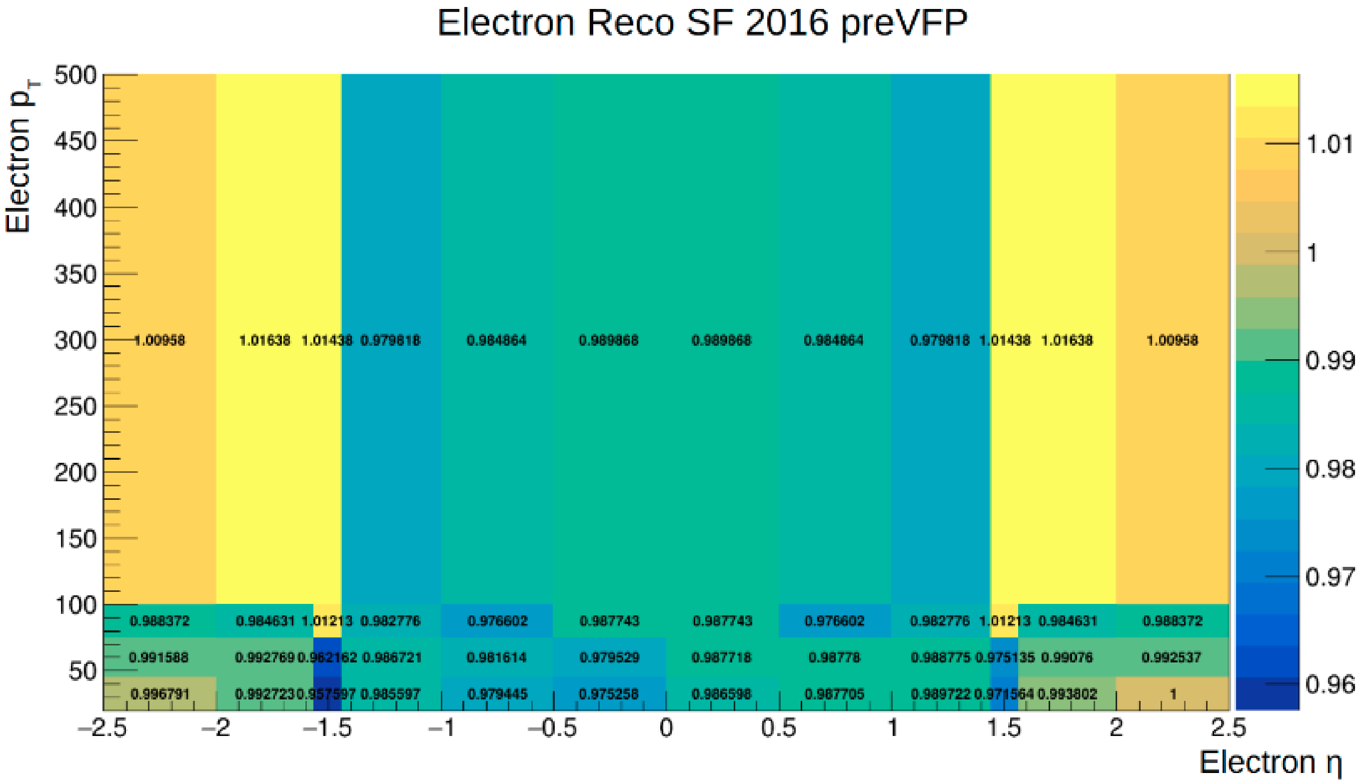


Figure A.4: Trigger (top) and reconstruction (bottom) electron SFs, computed using 2016 pre-VFP simulation and data [327].

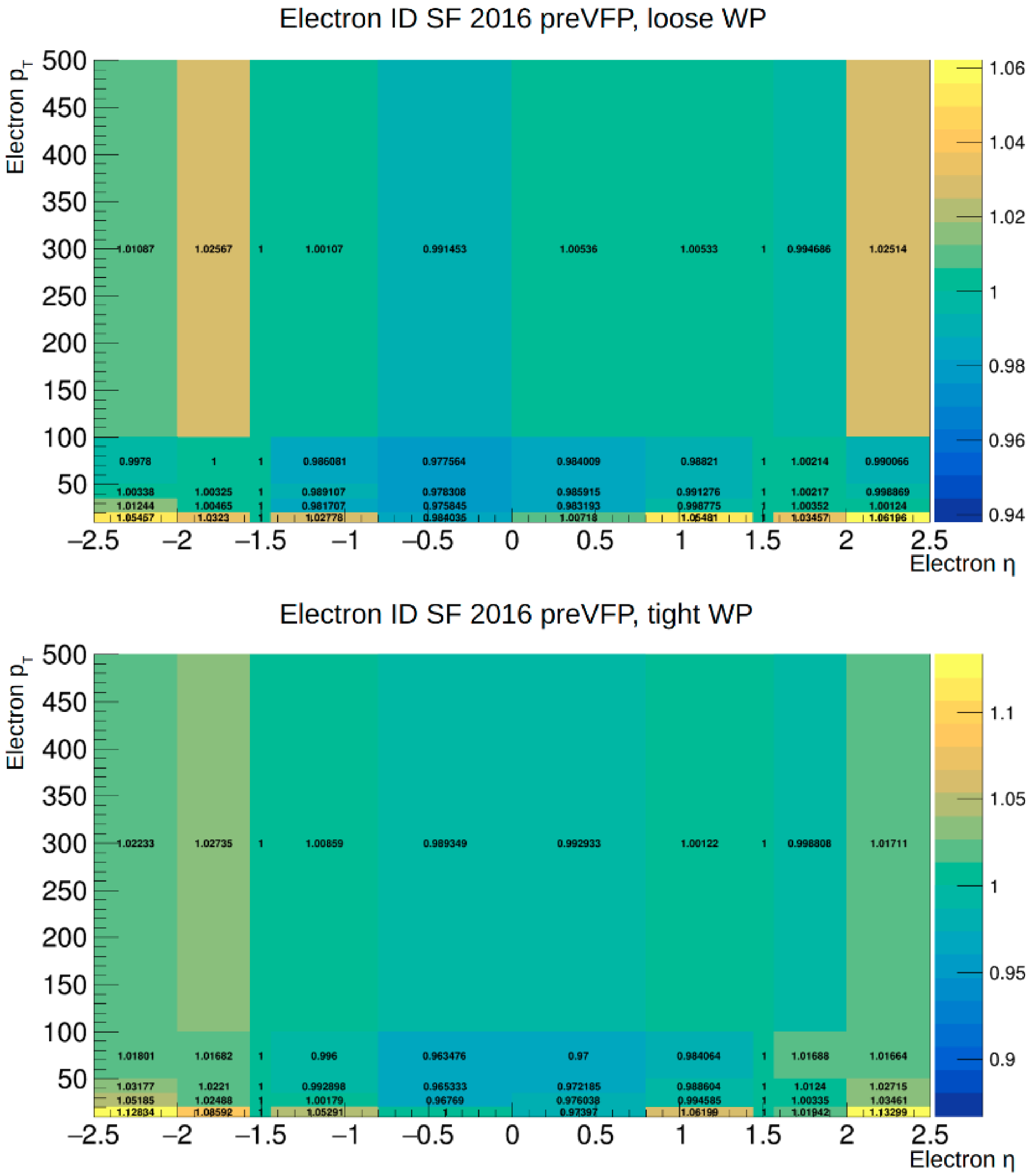


Figure A.5: Loose (top) and tight (bottom) ID electron SFs, computed using 2016 preVFP simulation and data [327].

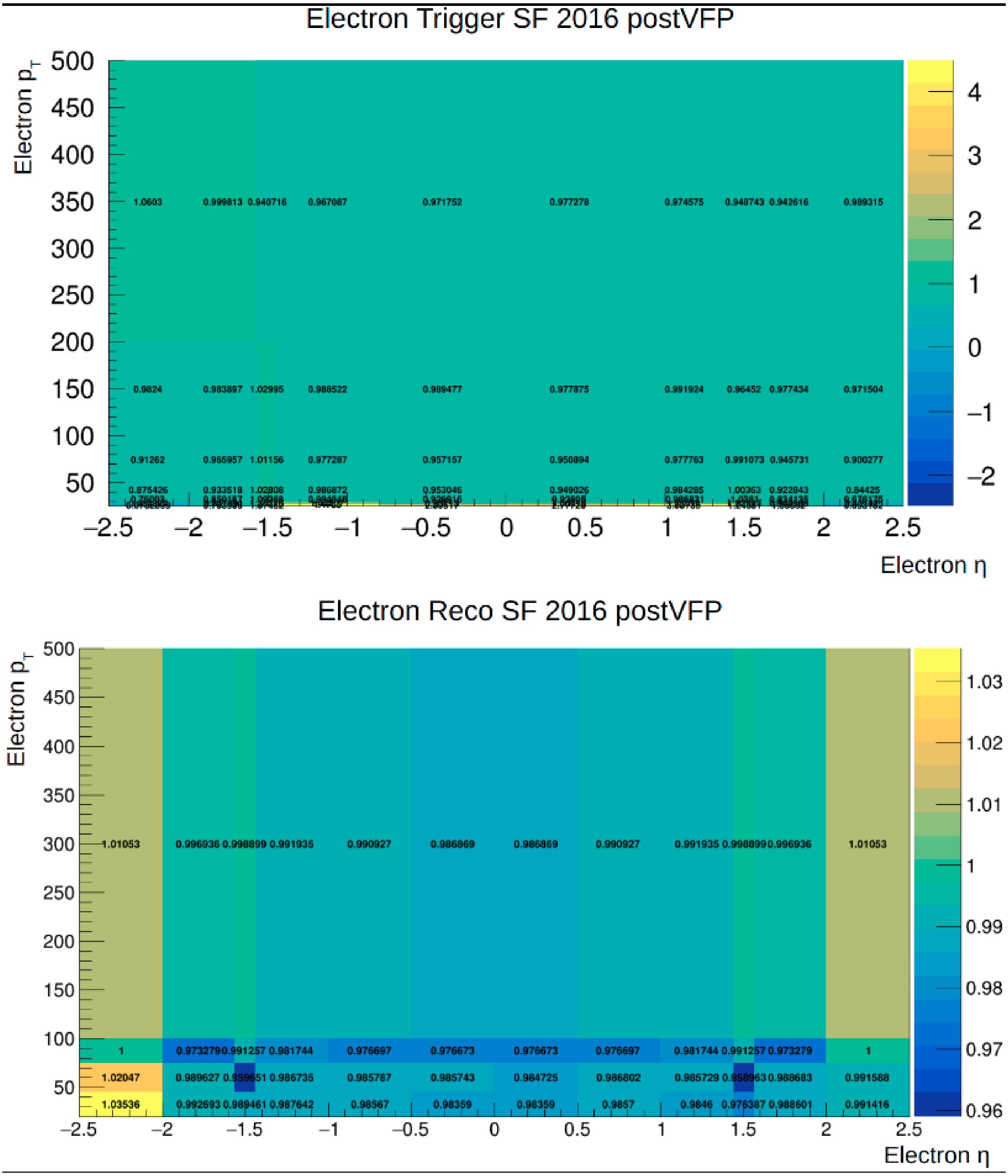


Figure A.6: Trigger (top) and reconstruction (bottom) electron SFs, computed using 2016 postVFP simulation and data [327].

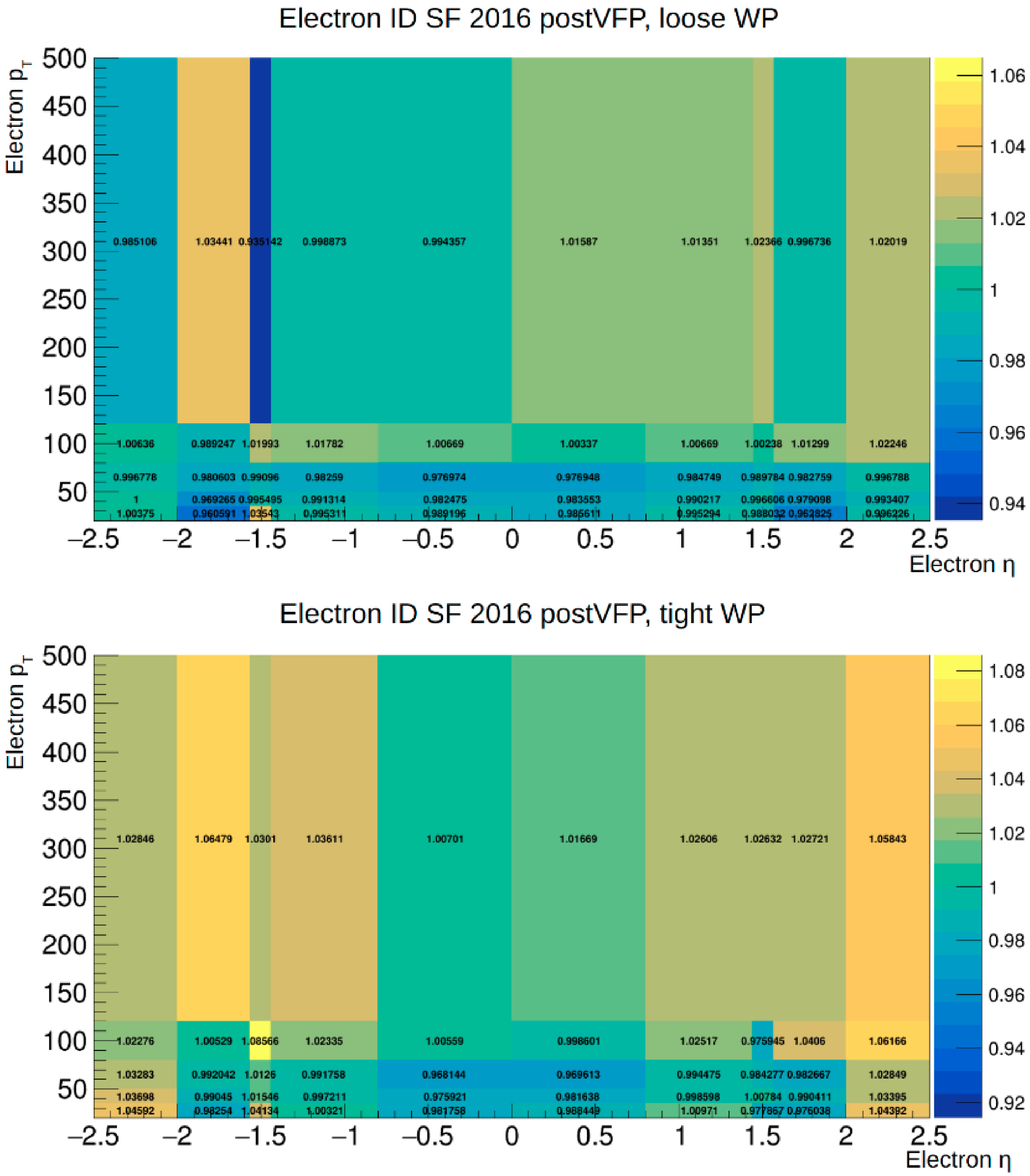


Figure A.7: Loose (top) and tight (bottom) ID electron SFs, computed using 2016 postVFP simulation and data [327].

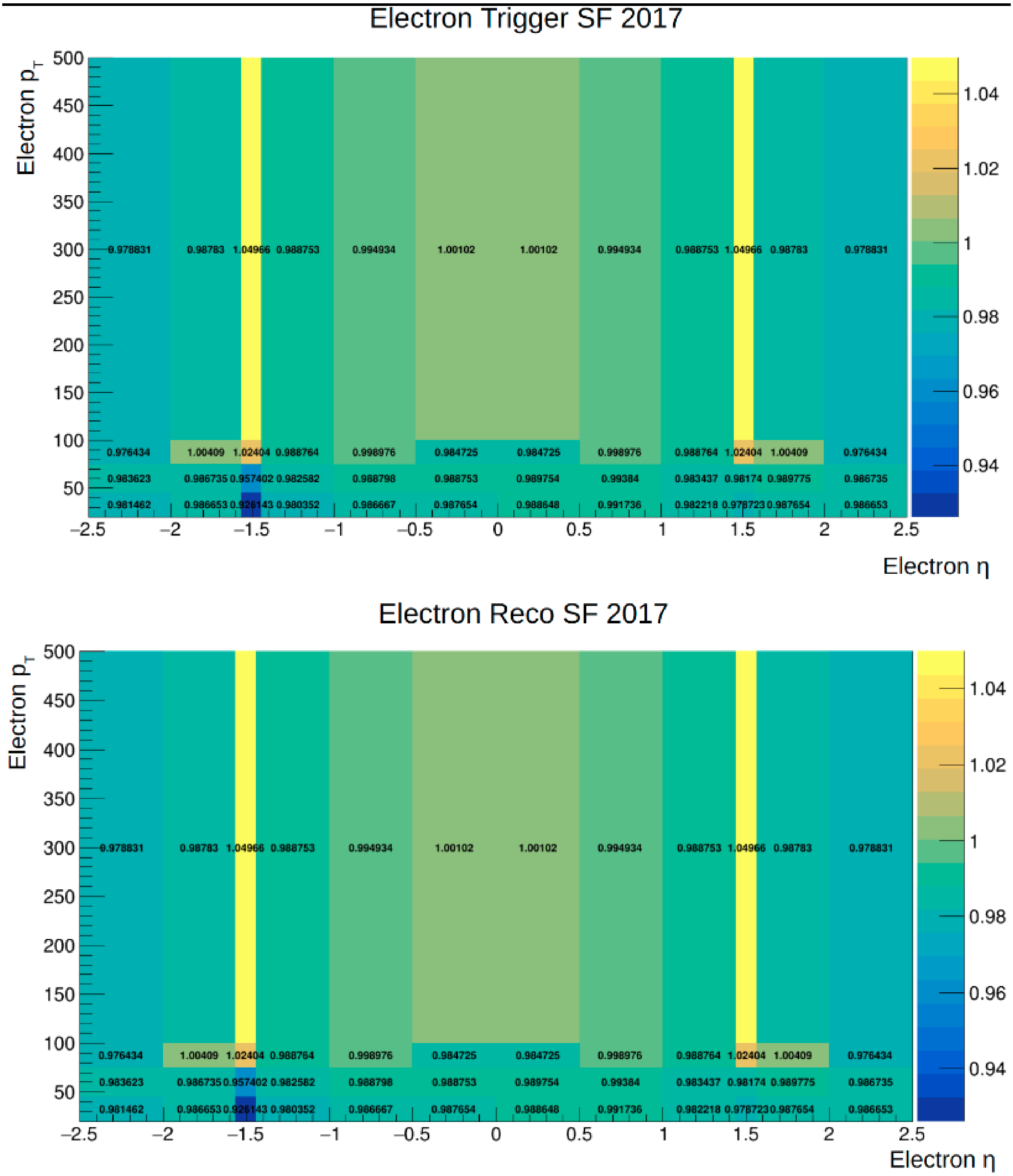


Figure A.8: Trigger (top) and reconstruction (bottom) electron SFs, computed using 2017 simulation and data [327].

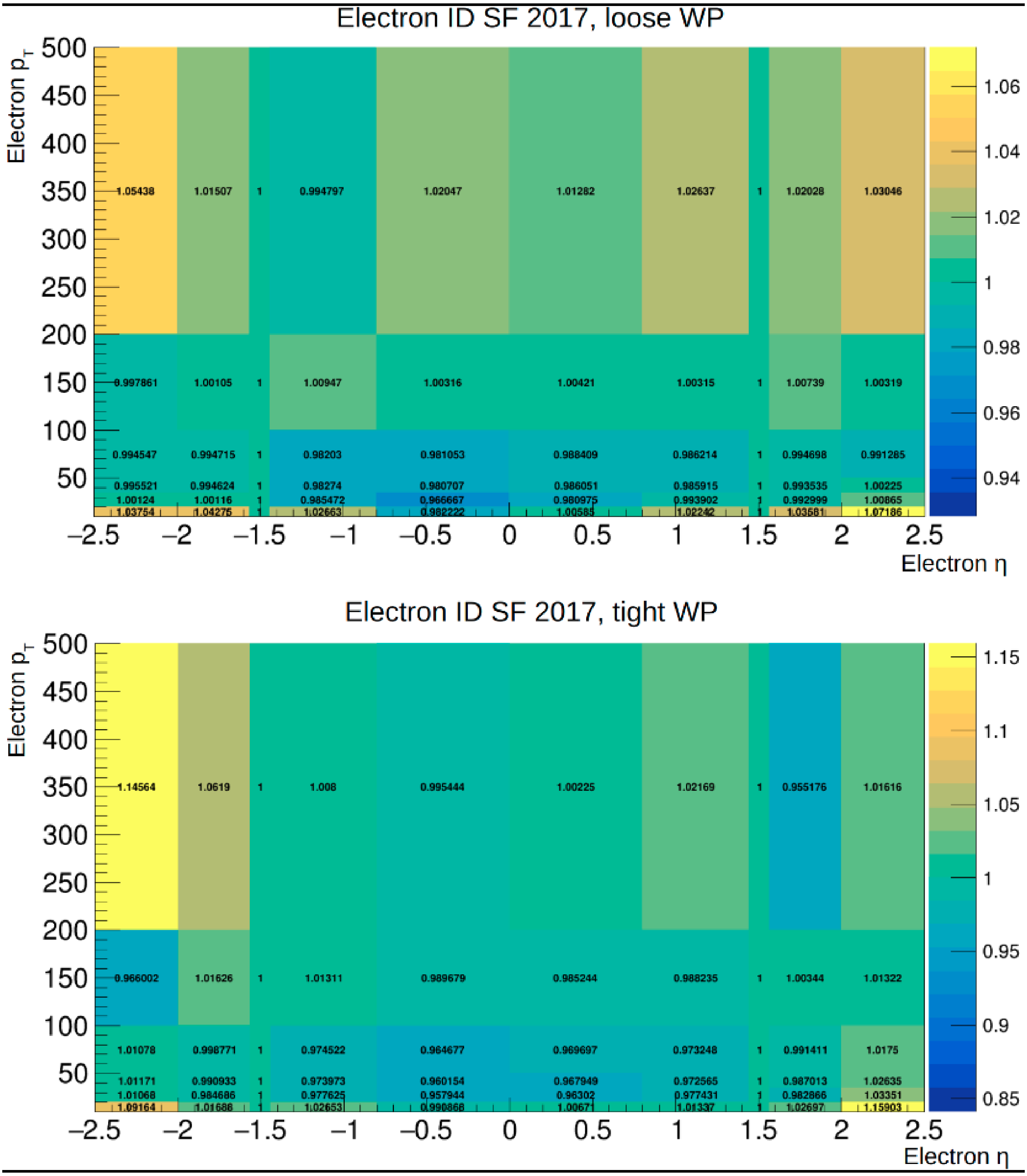


Figure A.9: Loose (top) and tight (bottom) ID electron SFs, computed using 2017 simulation and data [327].

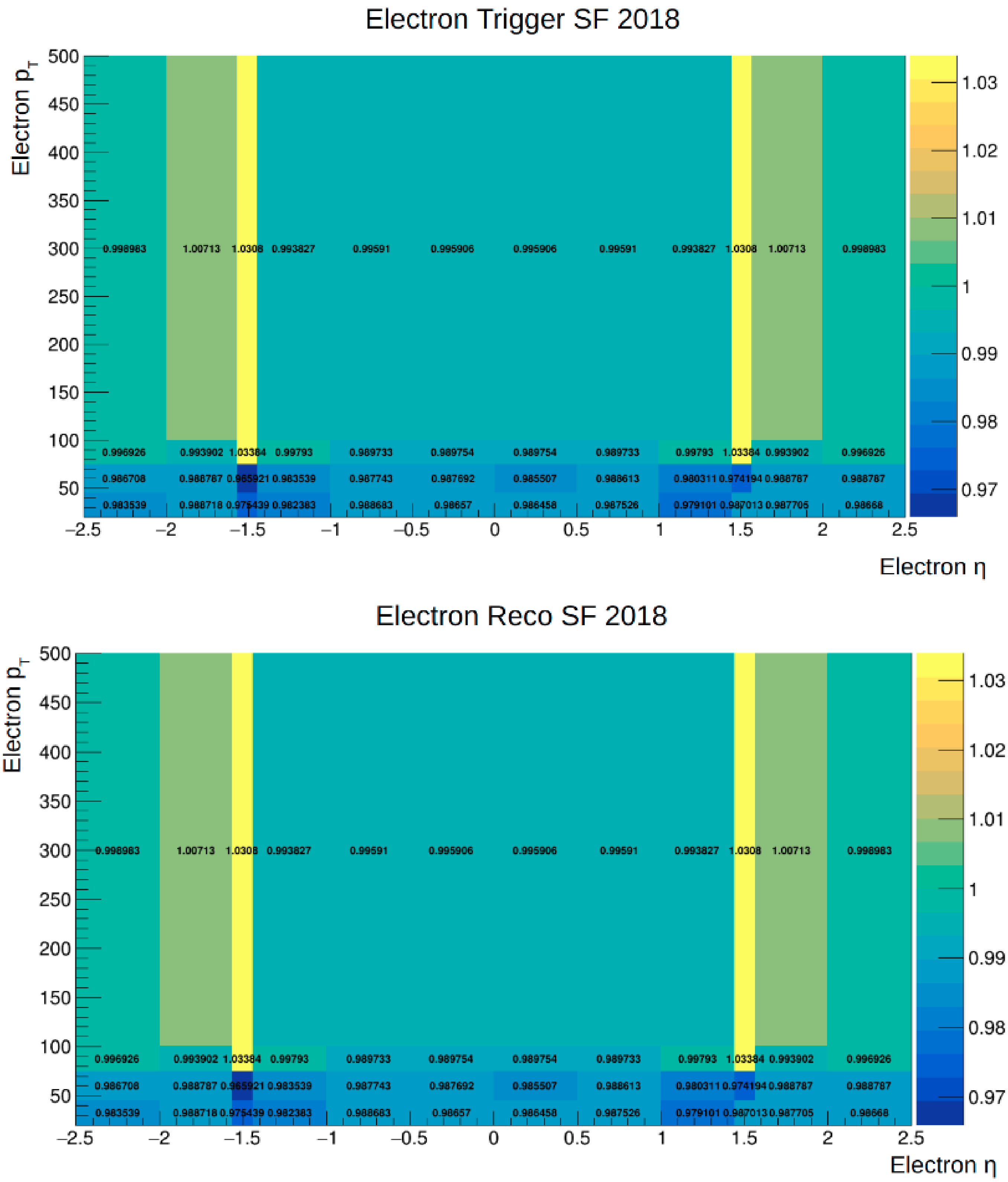


Figure A.10: Trigger (top) and reconstruction (bottom) electron SFs, computed using 2018 simulation and data [327].

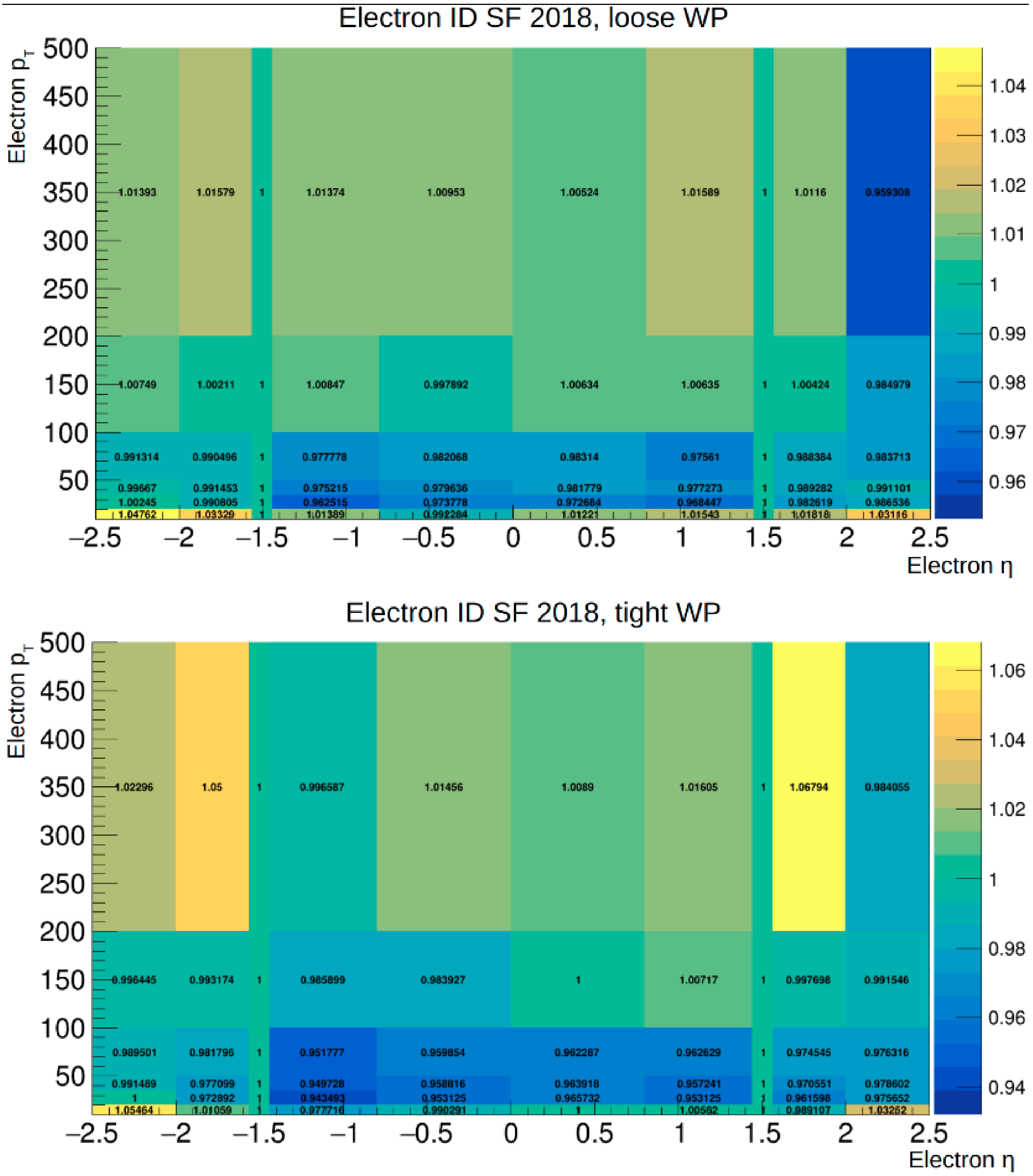


Figure A.11: Loose (top) and tight (bottom) ID electron SFs, computed using 2018 simulation and data [327].

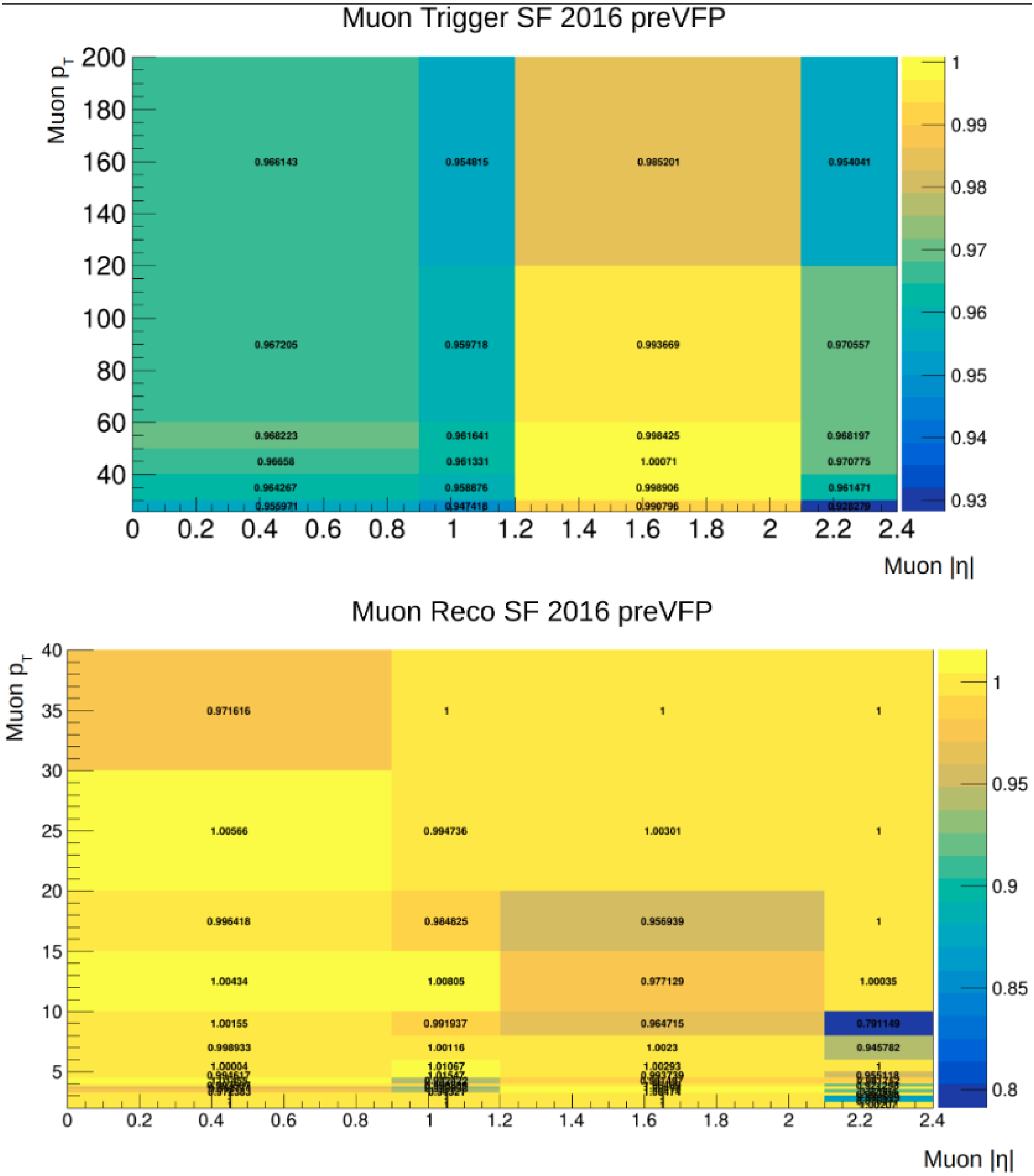


Figure A.12: Trigger (top) and reconstruction (bottom) muon SFs, computed using 2016 preVFP simulation and data [328].

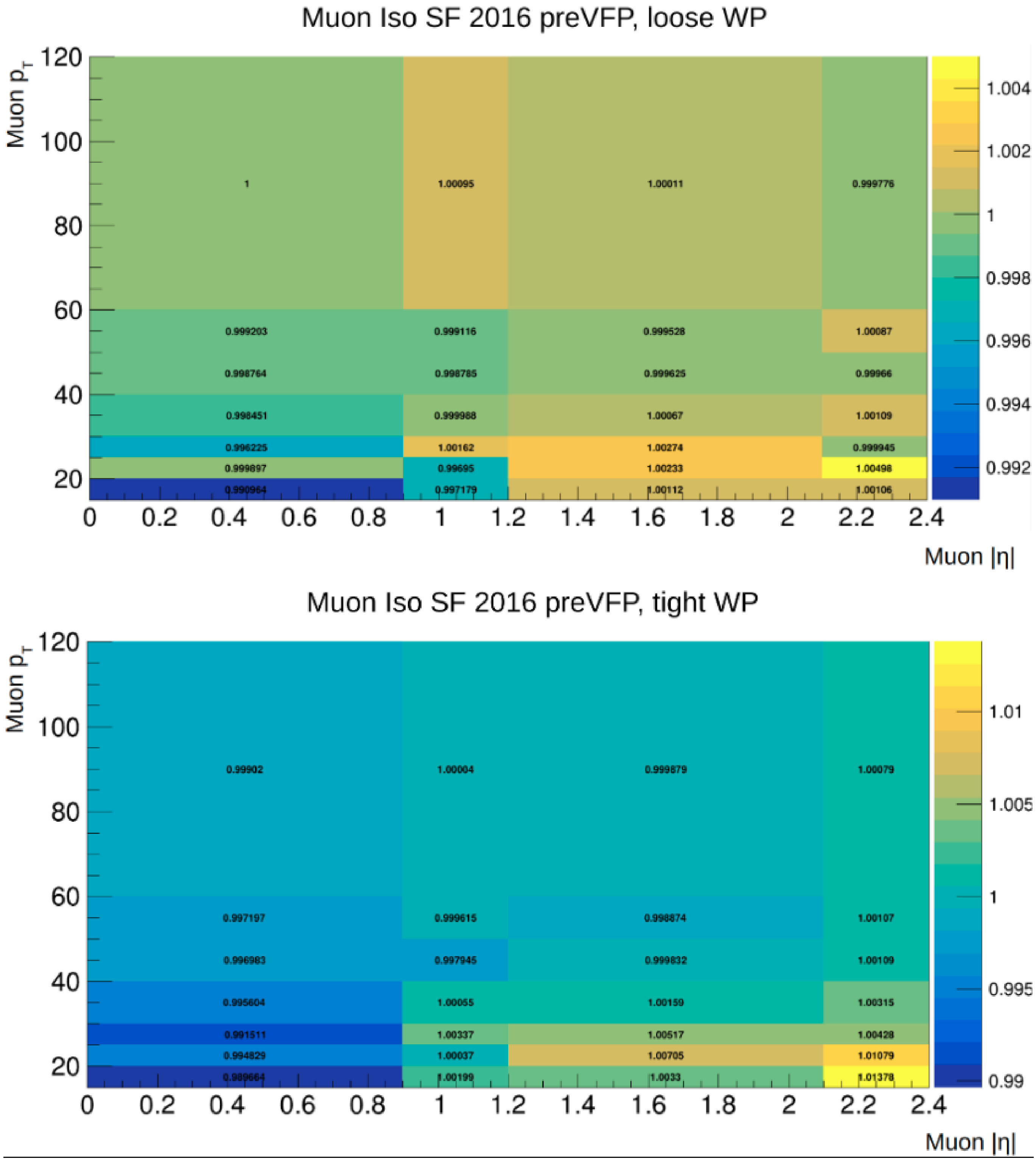


Figure A.13: Loose (top) and tight (bottom) isolation muon SFs, computed using 2016 preVFP simulation and data [328].

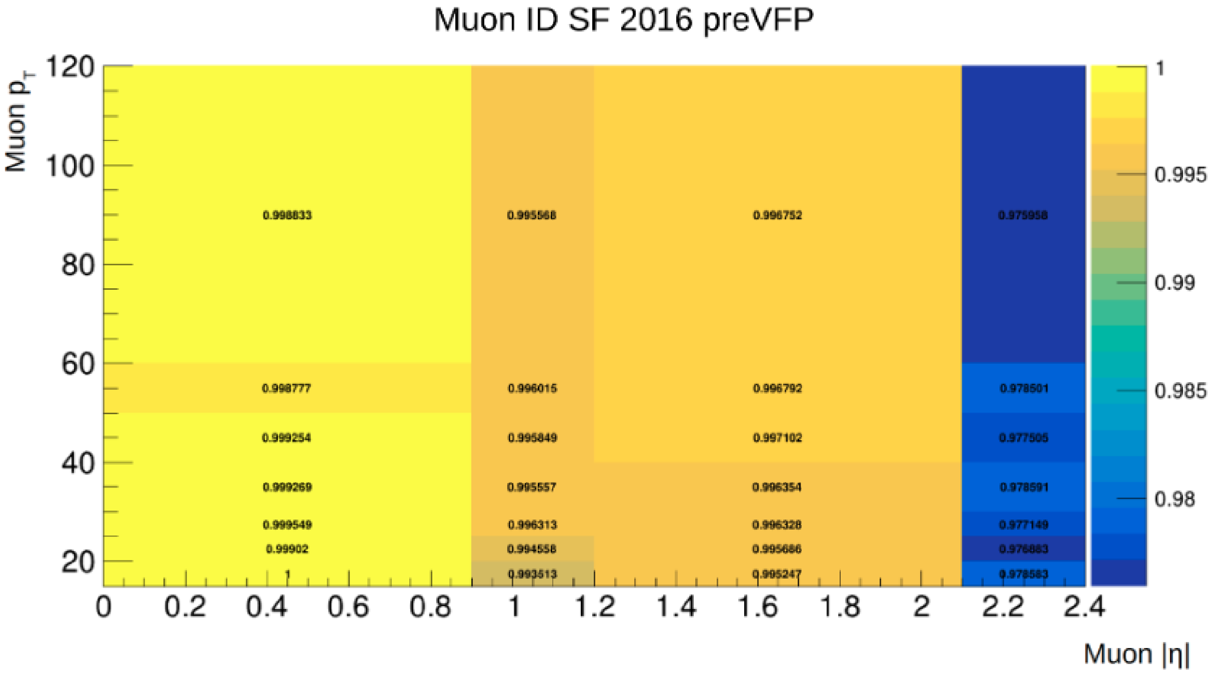


Figure A.14: ID muon SFs, computed using 2016 preVFP simulation and data [328].

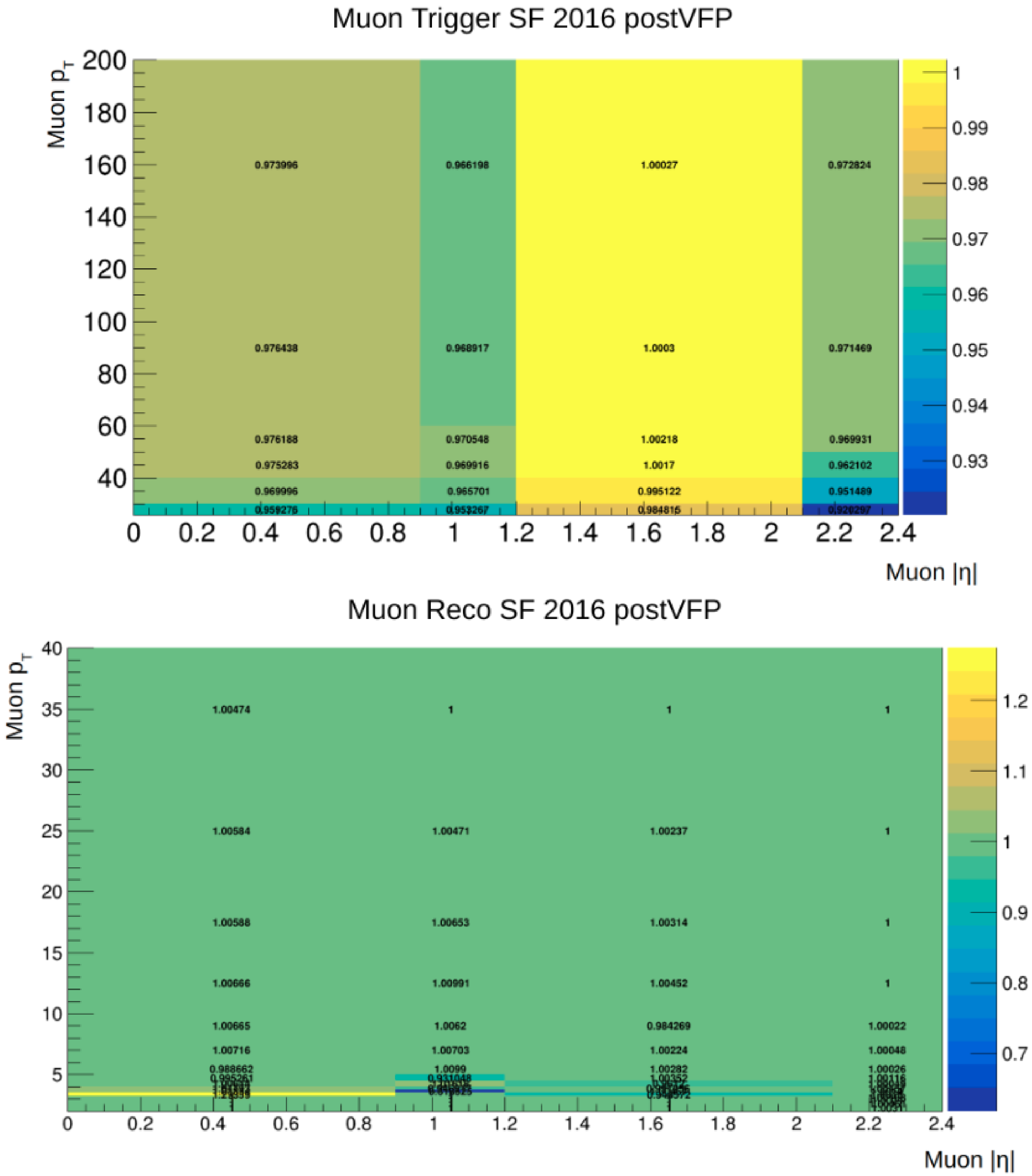


Figure A.15: Trigger (top) and reconstruction (bottom) muon SFs, computed using 2016 postVFP simulation and data [328].

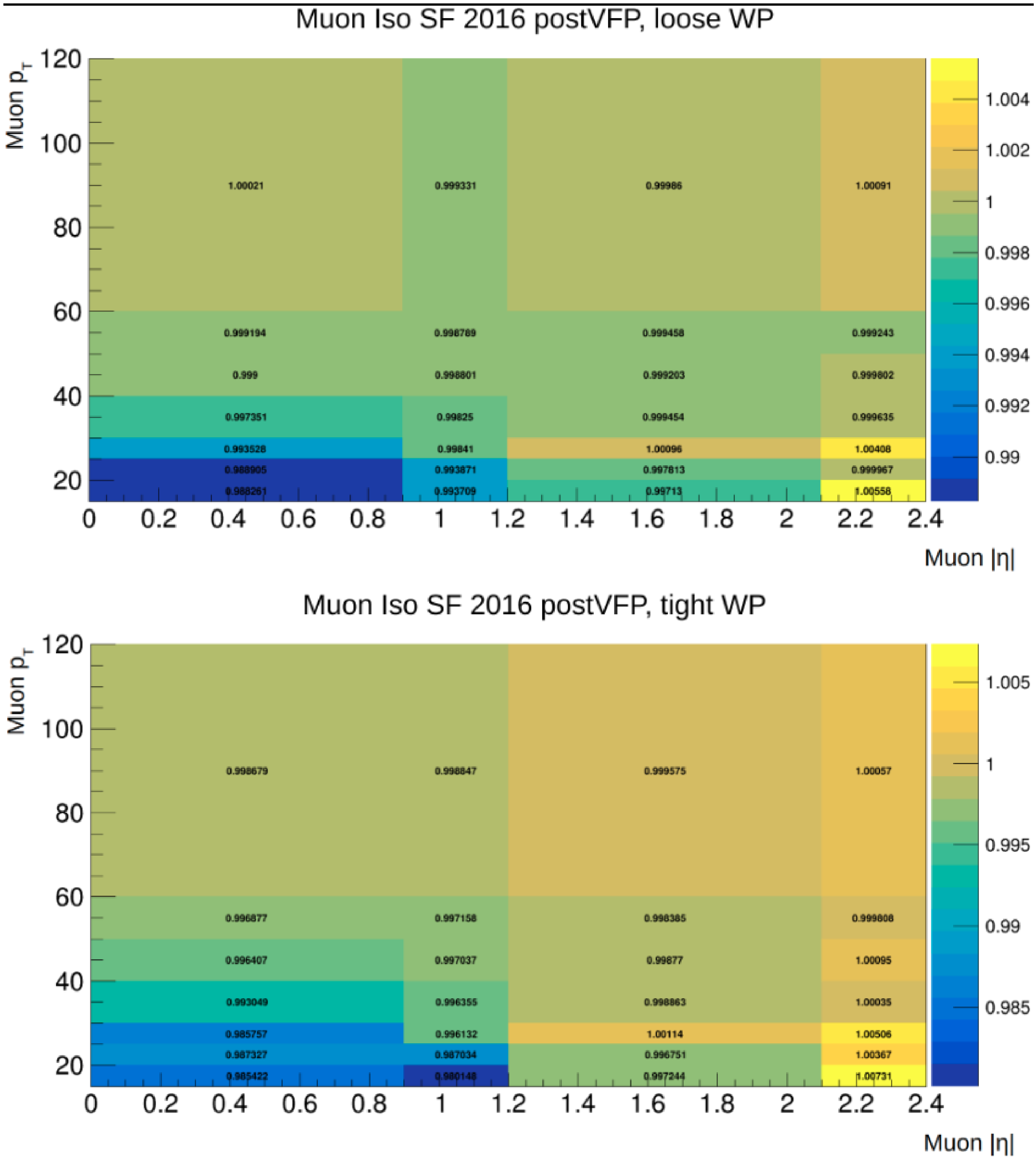


Figure A.16: Loose (top) and tight (bottom) isolation muon SFs, computed using 2016 postVFP simulation and data [328].

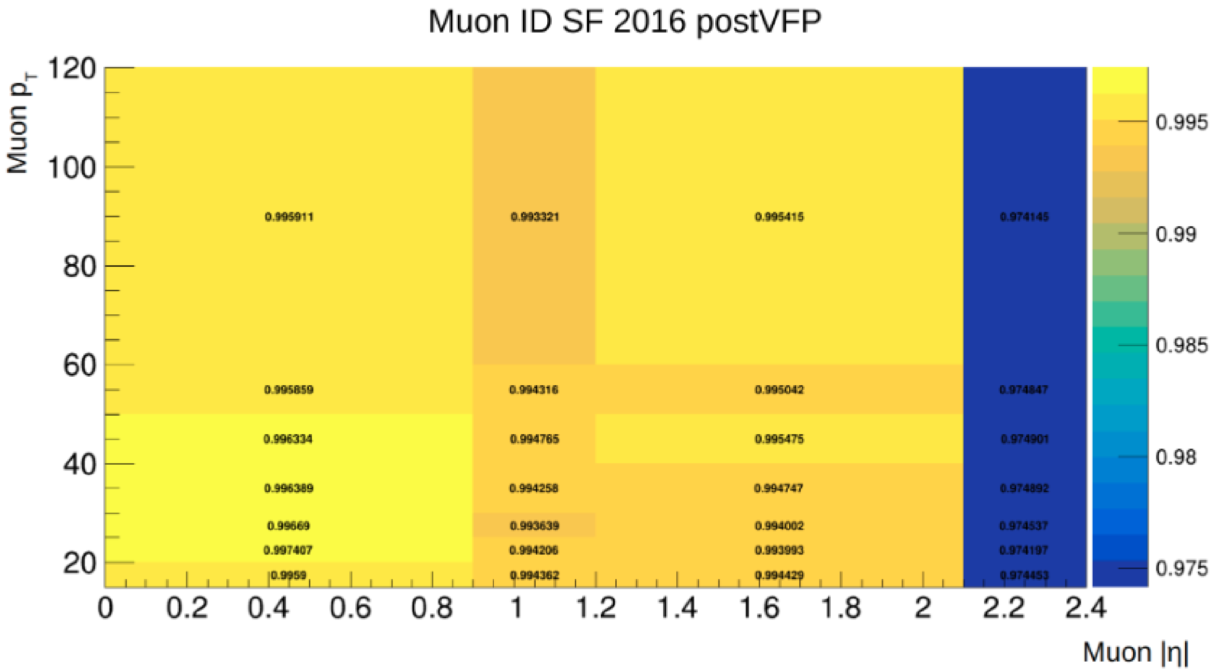


Figure A.17: ID muon SFs, computed using 2016 postVFP simulation and data [328].

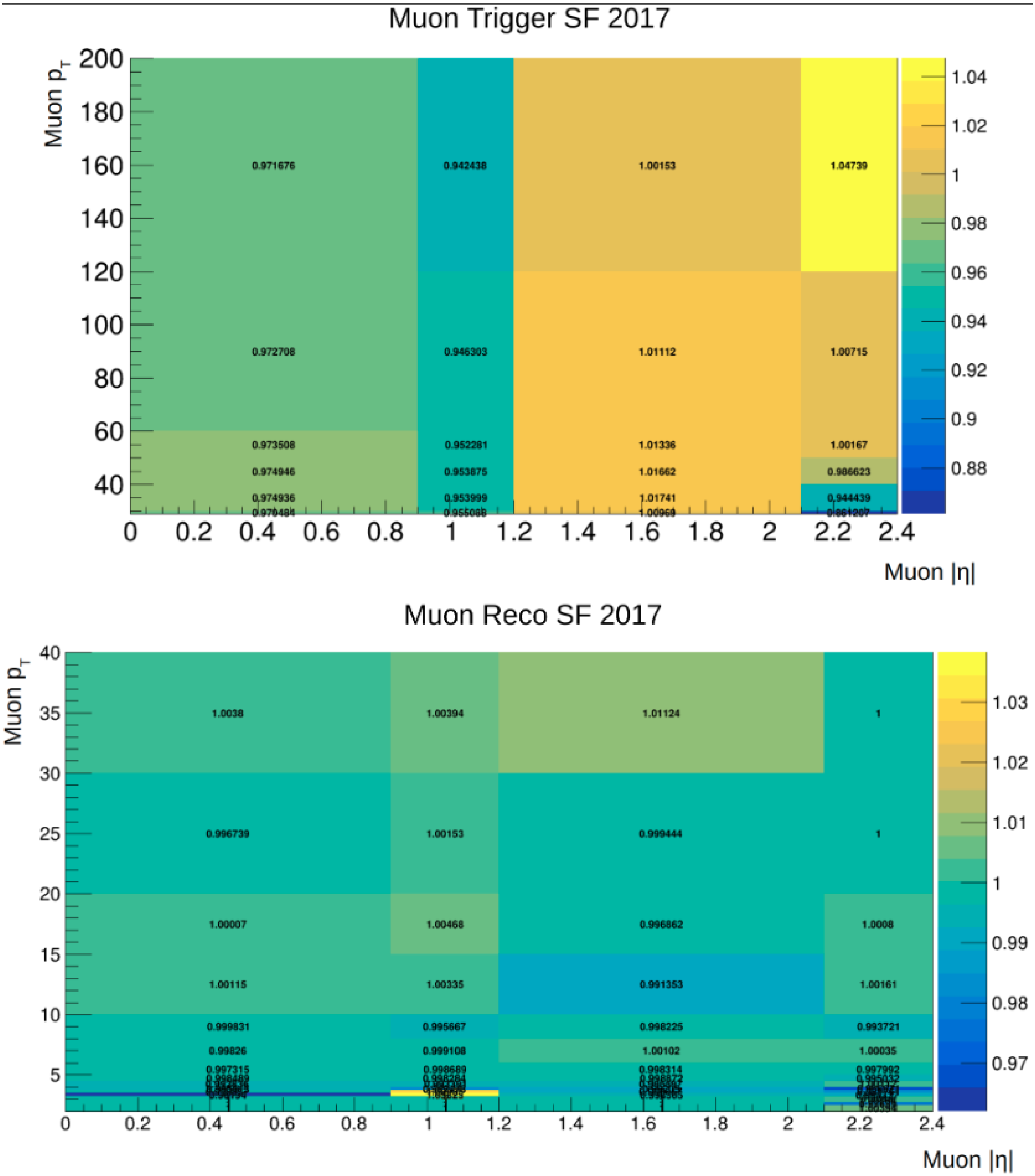


Figure A.18: Trigger (top) and reconstruction (bottom) muon SFs, computed using 2017 simulation and data [328].

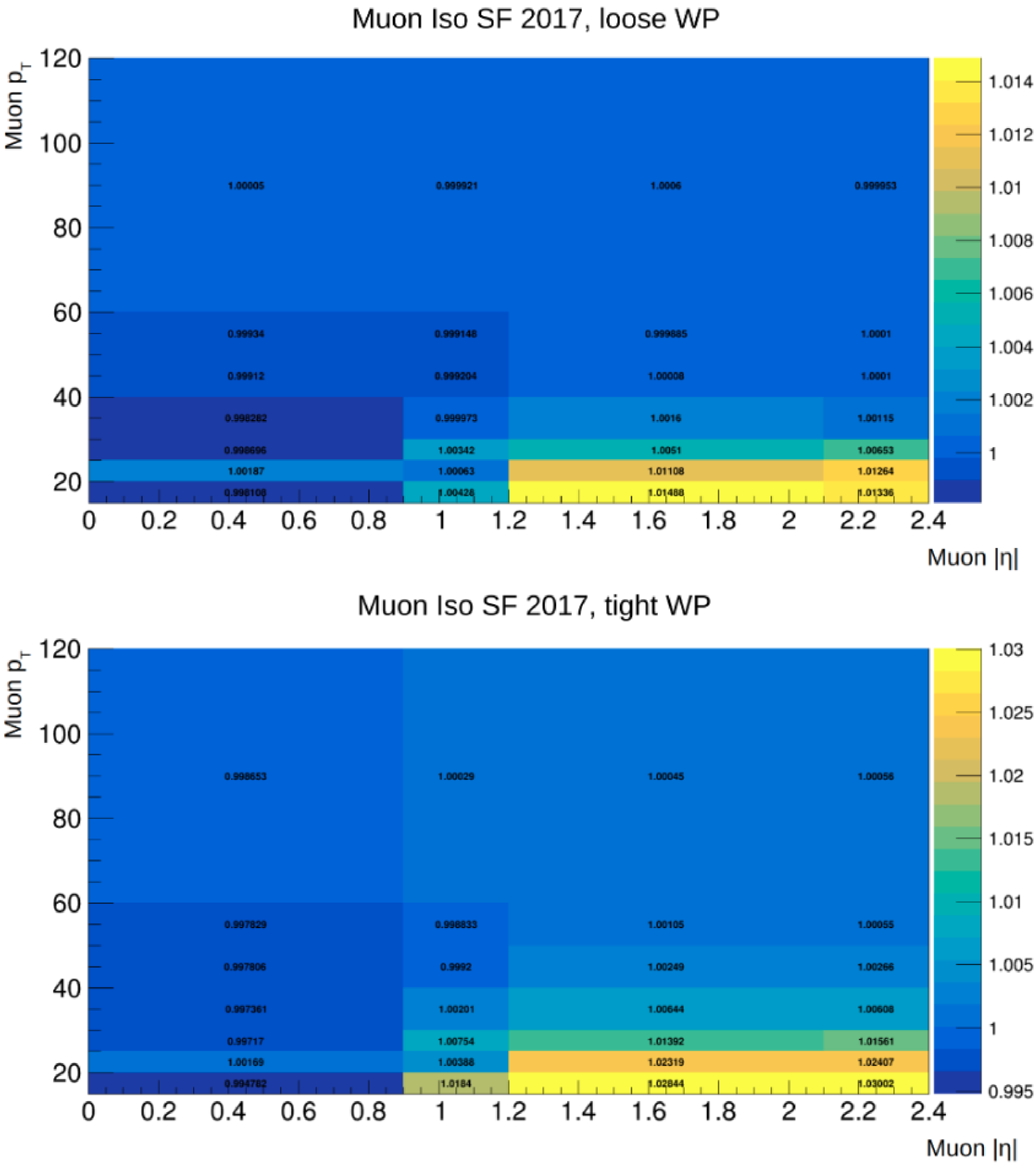


Figure A.19: Loose (top) and tight (bottom) isolation muon SFs, computed using 2017 simulation and data [328].

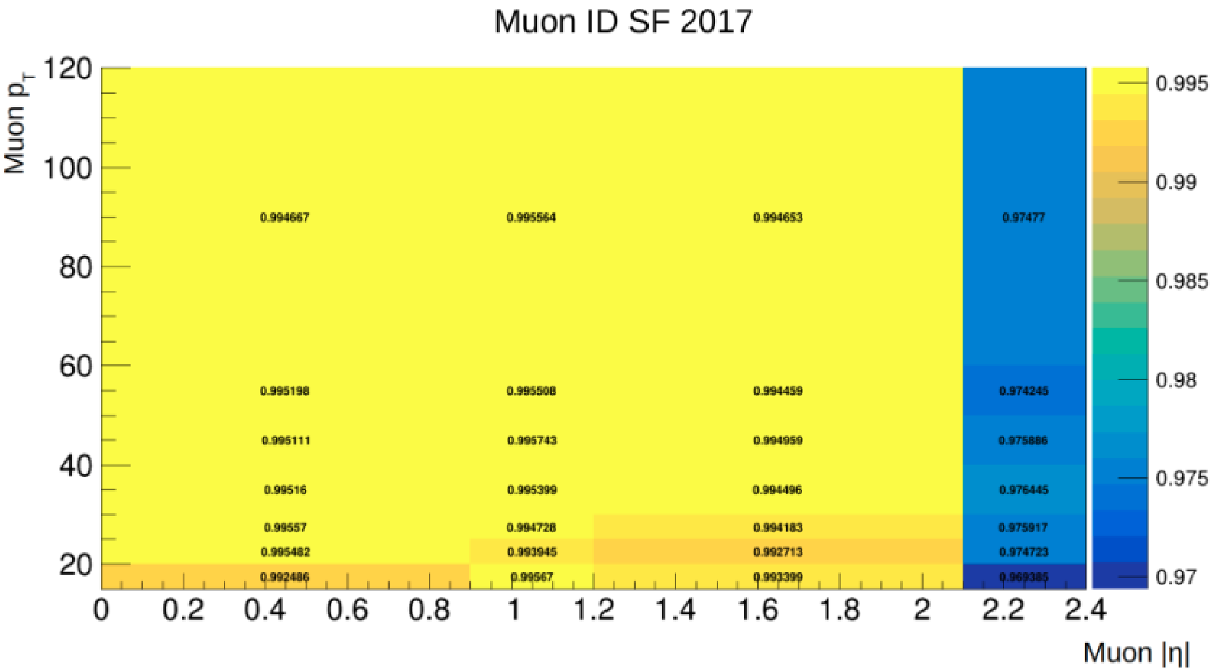


Figure A.20: ID muon SFs, computed using 2017 simulation and data [328].

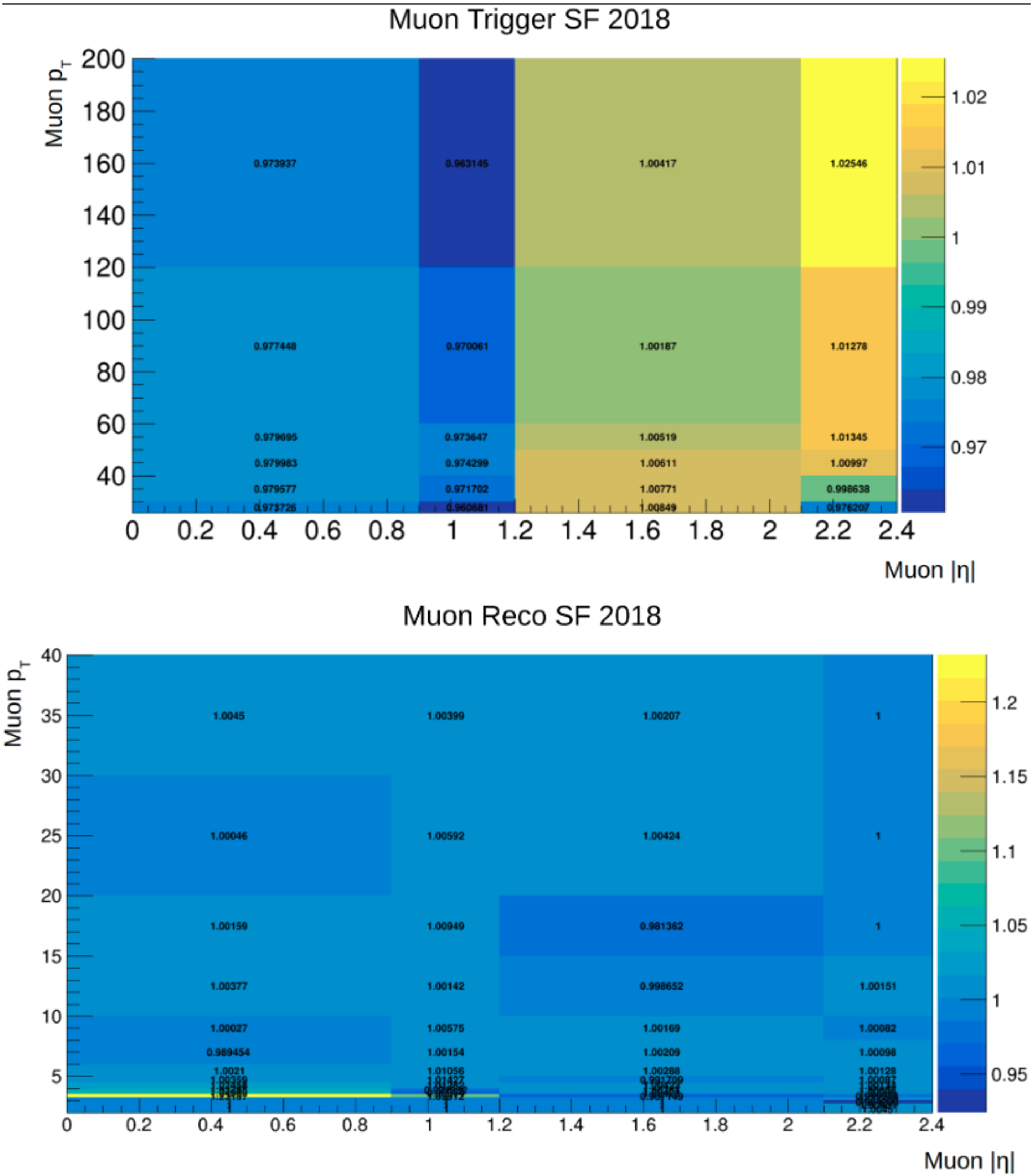


Figure A.21: Trigger (top) and reconstruction (bottom) muon SFs, computed using 2018 simulation and data [328].

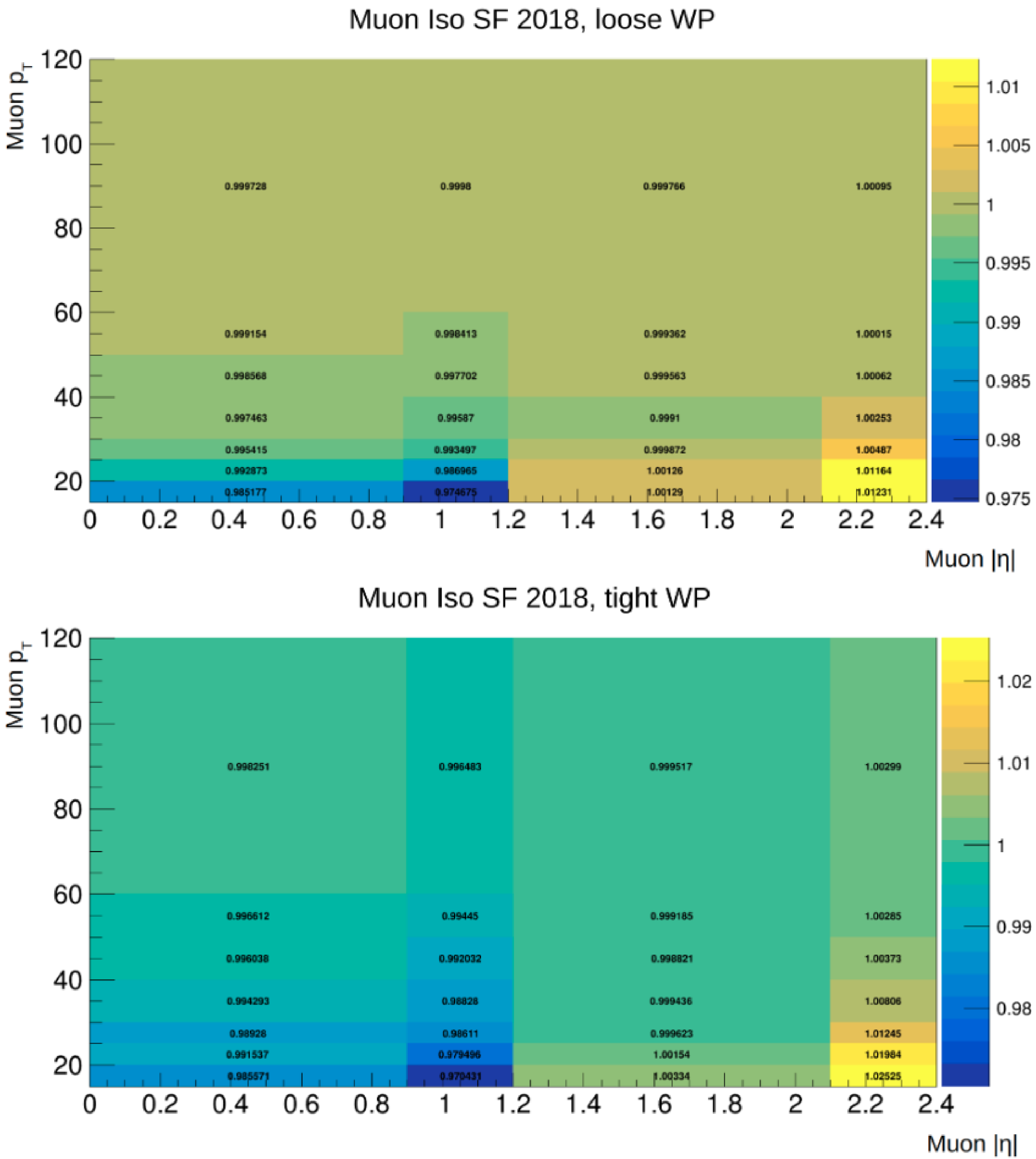


Figure A.22: Loose (top) and tight (bottom) isolation muon SFs, computed using 2018 simulation and data [328].

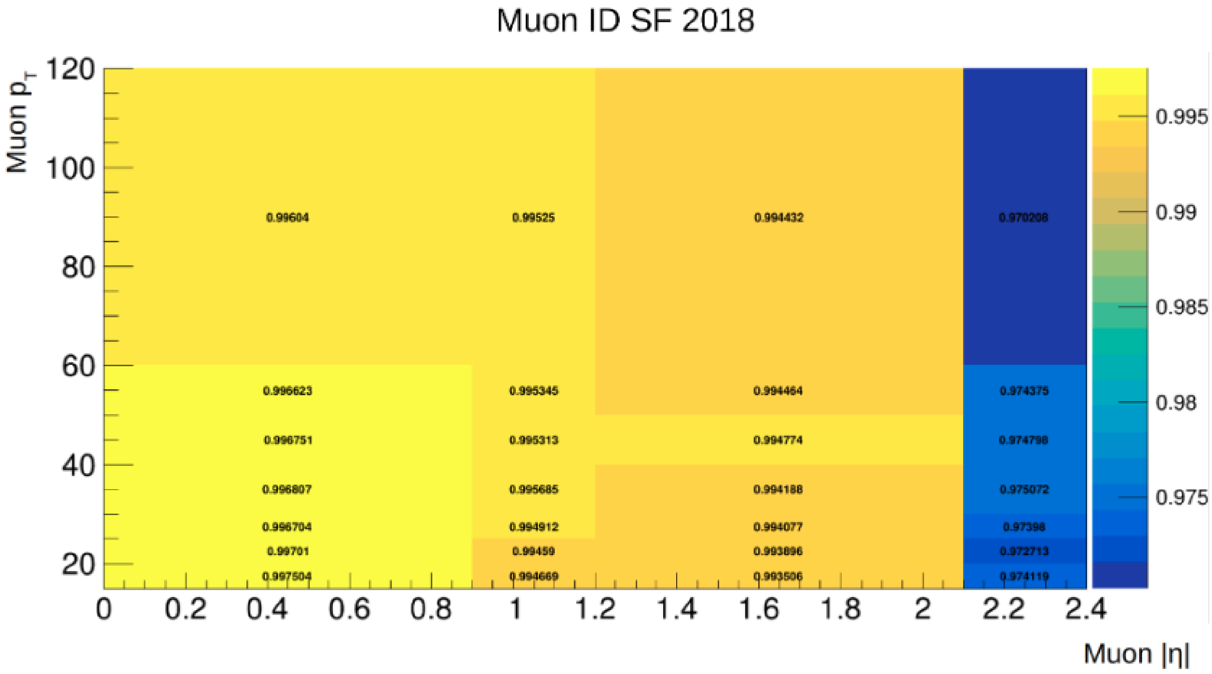


Figure A.23: ID muon SFs, computed using 2018 simulation and data [328].

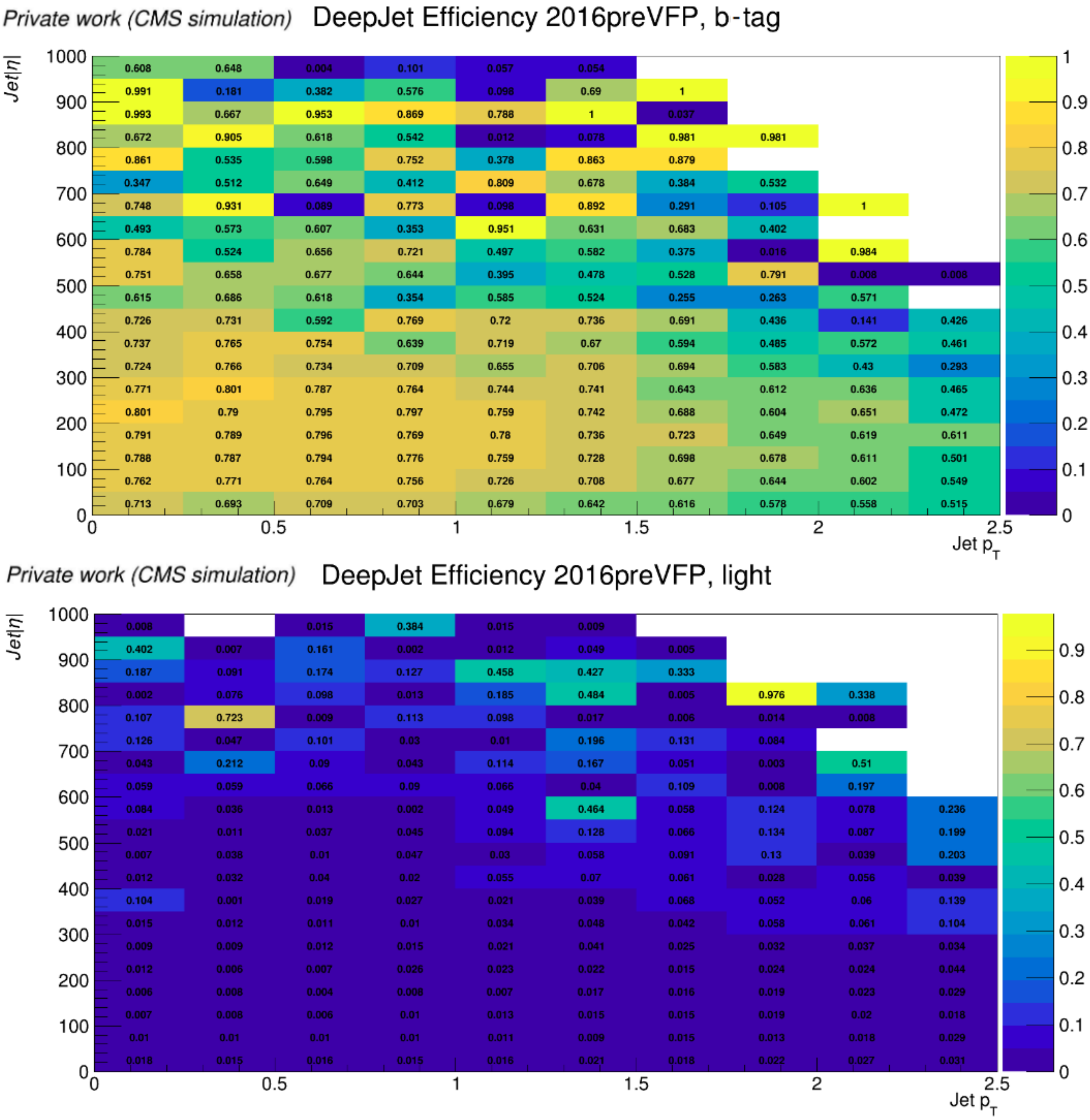


Figure A.24: Efficiencies ϵ_{b-tag} (top) and ϵ_{light} (bottom), computed using 2016 preVFP simulation with $m_{T'} = 700$ GeV.

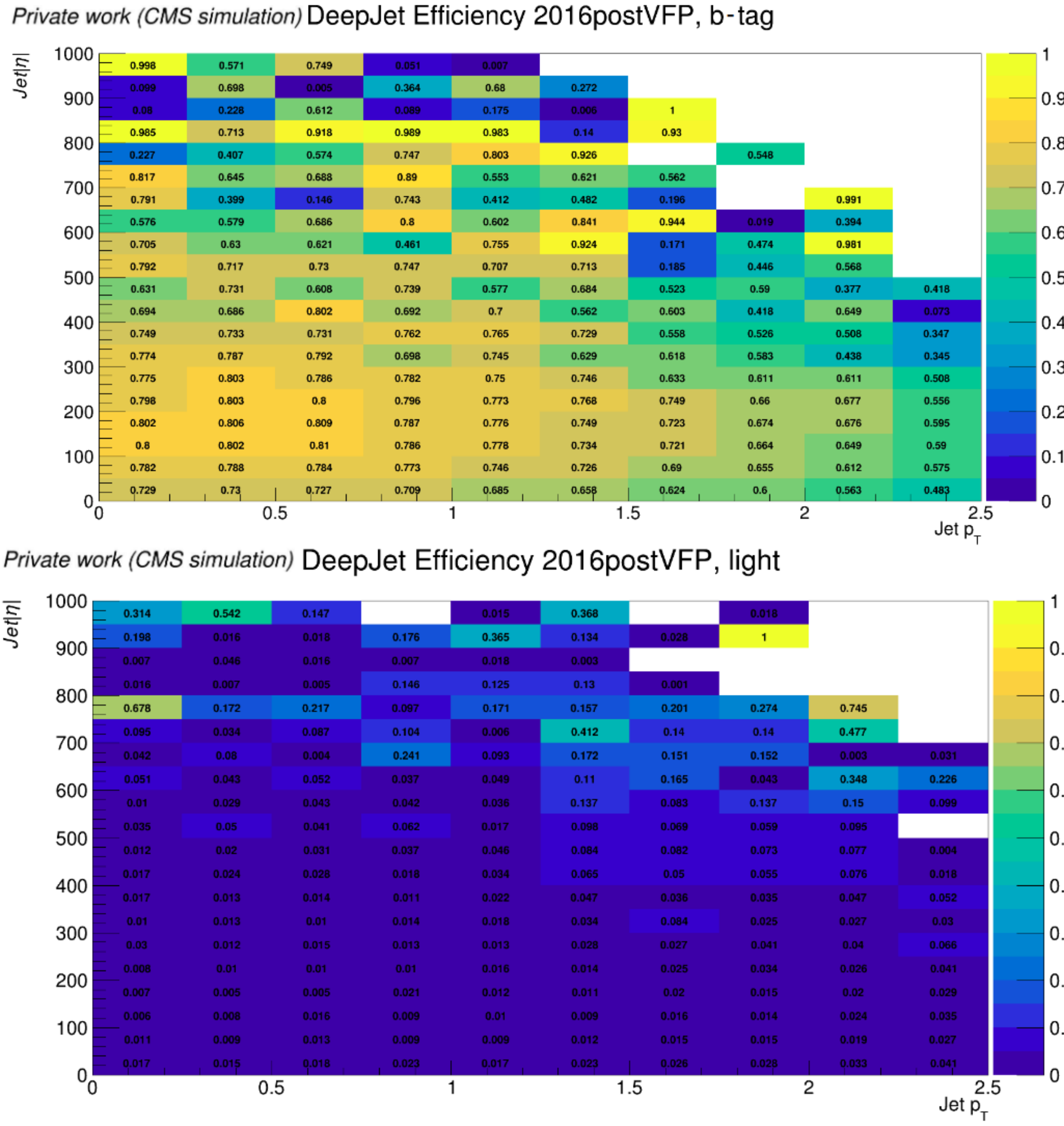


Figure A.25: Efficiencies ϵ_{b-tag} (top) and ϵ_{light} (bottom), computed using 2016 postVFP simulation with $m_{T'} = 700$ GeV.

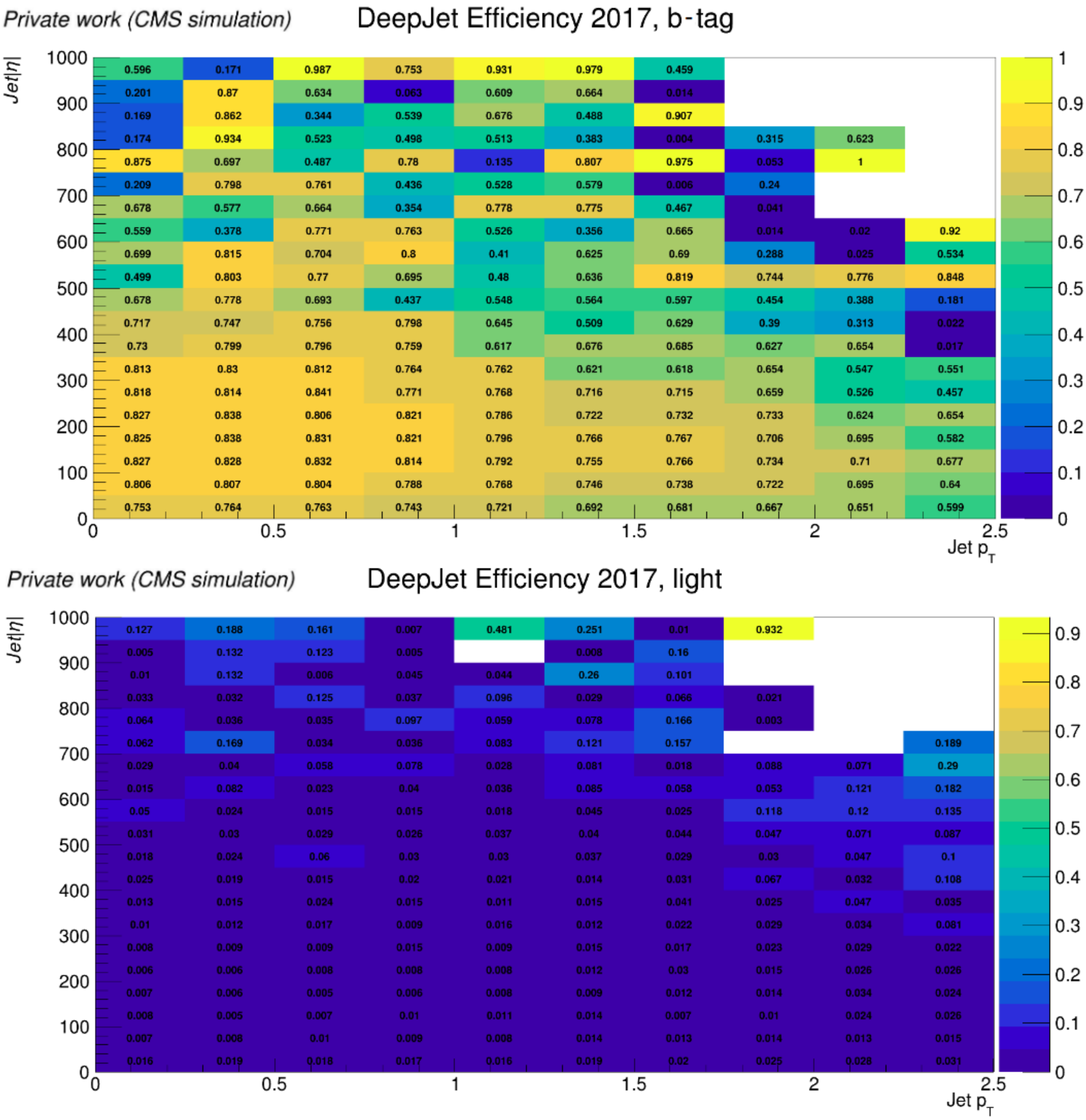


Figure A.26: Efficiencies ϵ_{b-tag} (top) and ϵ_{light} (bottom), computed using 2017 simulation with $m_{T'} = 700$ GeV.

Appendix B

Analysis strategy per mass point value of the Vector-Like Quark T'

B.1 Optimized selection using the getQuantiles method

The results of the optimized selection per mass point value of the T' are presented in Tables B.1 and B.2 for the cuts 2 and 3 respectively. They follow the procedure discussed in Chapter 4 using the getQuantiles method, and illustrated with 2018 simulated samples.

$m_{T'}$ (GeV)	Cut 2
600	$p_{T,l_1} + p_{T,l_2} > 4.345 + 0.3415 M_{T,T'} - 1.118 \times 10^{-4} M^2_{T,T'}$
625	$p_{T,l_1} + p_{T,l_2} > 12.58 + 0.3095 M_{T,T'} - 8.849 \times 10^{-5} M^2_{T,T'}$
650	$p_{T,l_1} + p_{T,l_2} > 8.147 + 0.3199 M_{T,T'} - 9.502 \times 10^{-5} M^2_{T,T'}$
675	$p_{T,l_1} + p_{T,l_2} > 3.216 + 0.3081 M_{T,T'} - 1.055 \times 10^{-4} M^2_{T,T'}$
700	$p_{T,l_1} + p_{T,l_2} > 6.468 + 0.3081 M_{T,T'} - 9.9901 \times 10^{-5} M^2_{T,T'}$
800	$p_{T,l_1} + p_{T,l_2} > -1.569 + 0.2925 M_{T,T'} - 9.248 \times 10^{-5} M^2_{T,T'}$
900	$p_{T,l_1} + p_{T,l_2} > -19.47 + 0.32 M_{T,T'} - 1.185 \times 10^{-4} M^2_{T,T'}$
1000	$p_{T,l_1} + p_{T,l_2} > -18.2 + 0.2979 M_{T,T'} - 1.158 \times 10^{-4} M^2_{T,T'}$
1100	$p_{T,l_1} + p_{T,l_2} > -10.63 + 0.2637 M_{T,T'} - 9.076 \times 10^{-5} M^2_{T,T'}$
1200	$p_{T,l_1} + p_{T,l_2} > -0.5841 + 0.2269 M_{T,T'} - 6.661 \times 10^{-5} M^2_{T,T'}$

Table B.1: Values of the optimized cut 2 per mass point value of the T'.

$m_{T'}$ (GeV)	Cut 3
600	$\lvert M_{inv}(3j) - m_{top} \rvert > 163.9 - 0.8918 M_{T,T'} + 1.956 \times 10^{-3} M^2_{T,T'} - 1.68 \times 10^{-6} M^3_{T,T'} + 4.954 \times 10^{-10} M^4_{T,T'}$
625	$\lvert M_{inv}(3j) - m_{top} \rvert > 158.1 - 0.8573 M_{T,T'} + 1.879 \times 10^{-3} M^2_{T,T'} - 1.608 \times 10^{-6} M^3_{T,T'} + 4.717 \times 10^{-10} M^4_{T,T'}$
650	$\lvert M_{inv}(3j) - m_{top} \rvert > 168.2 - 0.8982 M_{T,T'} + 1.93 \times 10^{-3} M^2_{T,T'} - 1.631 \times 10^{-6} M^3_{T,T'} + 4.742 \times 10^{-10} M^4_{T,T'}$
675	$\lvert M_{inv}(3j) - m_{top} \rvert > 162.2 - 0.8649 M_{T,T'} + 1.864 \times 10^{-3} M^2_{T,T'} - 1.579 \times 10^{-6} M^3_{T,T'} + 4.605 \times 10^{-10} M^4_{T,T'}$
700	$\lvert M_{inv}(3j) - m_{top} \rvert > 143 - 0.7486 M_{T,T'} + 1.583 \times 10^{-3} M^2_{T,T'} - 1.324 \times 10^{-6} M^3_{T,T'} + 3.835 \times 10^{-10} M^4_{T,T'}$
800	$\lvert M_{inv}(3j) - m_{top} \rvert > 112.3 - 0.6151 M_{T,T'} + 1.384 \times 10^{-3} M^2_{T,T'} - 1.208 \times 10^{-6} M^3_{T,T'} + 3.631 \times 10^{-10} M^4_{T,T'}$
900	$\lvert M_{inv}(3j) - m_{top} \rvert > -42.47 + 0.21 M_{T,T'} - 1.844 \times 10^{-4} M^2_{T,T'} + 5.344 \times 10^{-8} M^3_{T,T'}$
1000	$\lvert M_{inv}(3j) - m_{top} \rvert > -12.05 + 0.09041 M_{T,T'} - 4.074 \times 10^{-5} M^2_{T,T'}$
1100	$\lvert M_{inv}(3j) - m_{top} \rvert > 287.5 - 1.503 M_{T,T'} + 2.935 \times 10^{-3} M^2_{T,T'} - 2.368 \times 10^{-6} M^3_{T,T'} + 6.792 \times 10^{-10} M^4_{T,T'}$
1200	$\lvert M_{inv}(3j) - m_{top} \rvert > 184.9 - 0.9112 M_{T,T'} + 1.72 \times 10^{-3} M^2_{T,T'} - 1.316 \times 10^{-6} M^3_{T,T'} + 3.548 \times 10^{-10} M^4_{T,T'}$

Table B.2: Values of the optimized cut 3 per mass point value of the T'.

B.2 Expected yield in the Signal Region

The number of selected signal and background events per mass point value of the T' in the SR are presented after the optimized selection from Table B.3 to B.5 for 2016, 2017, and 2018 respectively. The procedure is discussed in Chapter 4. The $\mu\mu$, eμ, and ee channels are combined, and all the values are scaled to the integrated luminosity as presented in equation 4.2. The 'Others' column includes the $t\bar{t}$0l, DY, and $W + jets$ processes.

B.3 Control and Measurement regions

B.3.1 The ttX Control Region

The number of selected signal, background, and data events per mass point value of the T' in the ttX CR are presented from Table B.6 to B.8 for 2016, 2017, and 2018 respectively. The procedure is discussed in Chapter 5. The $\mu\mu$, eμ, and ee channels are combined, and all the values are scaled to the integrated luminosity as presented in equation 4.2. The 'Other' column includes the events which rarely pass the selection in the ttX CR.

B.3.2 The tt2l Measurement Region

The number of selected signal, background, and data events per mass point value of the T' in the tt2l MR are presented from Table B.9 to B.11 for 2016, 2017, and 2018 respectively. The procedure is discussed in Chapter 5. The $\mu\mu$, eμ, and ee channels are combined, and all the values are scaled to the integrated luminosity as presented in equation 4.2. The 'Other' column includes the events which rarely pass the selection in the tt2l MR.

B.3.3 The tt1l Measurement Regions

The number of selected signal, background, and data events per mass point value of the T' in the tt1l MRs are presented after the optimized selection from Table B.12 to B.17 for 2016, 2017, and 2018 respectively. The procedure is discussed in Chapter 5. The $\mu\mu$, eμ, and ee channels are combined, and all the values are scaled to the integrated luminosity as presented in equation 4.2. The 'Other' column includes the events which rarely pass the selection in the tt1l MRs.

$m_{T'}$ (GeV)	Signal	ϵ_{signal} (%)	$t\bar{t}$X	$t\bar{t}$1l	$t\bar{t}$2l	Multibosons	$tX(W/q)$	Others	S/B (%)
600	1.52	3.9	37.02	20.64	10.96	4.60	4.51	1.57	1.88
625	1.80	6.0	37.45	20.96	11.32	4.54	4.54	1.57	2.19
650	1.50	5.2	37.46	20.85	11.20	4.57	4.53	1.57	1.84
675	1.29	4.9	37.12	20.63	10.99	4.65	4.52	1.57	1.60
700	1.43	6.4	44.24	25.82	13.81	5.28	4.86	1.57	1.47
800	0.93	7.6	50.12	31.17	16.50	5.68	5.71	1.57	0.83
900	0.64	8.6	54.76	34.74	18.31	6.17	5.95	1.57	0.52
1000	0.47	10.1	59.75	39.93	20.25	6.66	6.24	1.57	0.35
1100	0.25	7.9	63.25	42.93	21.25	6.90	6.96	1.57	0.17
1200	0.16	8.4	63.42	42.26	21.29	6.88	6.99	1.57	0.11

Table B.3: Number of events for the signal and background processes after the optimized selection per mass point value of the T' in the SR, computed using 2016 simulation. The signal efficiency is also provided.

$m_{T'}$ (GeV)	Signal	ϵ_{signal} (%)	$t\bar{t}$X	$t\bar{t}$1l	$t\bar{t}$2l	Multibosons	$tX(W/q)$	Others	S/B (%)
600	1.64	4.4	42.11	16.31	6.50	5.94	3.76	0.00	2.15
625	1.72	5.1	42.71	16.69	6.53	5.97	3.79	0.00	2.22
650	1.48	5.2	42.63	16.49	6.50	6.02	3.77	0.00	1.92
675	1.48	5.5	42.33	16.23	6.50	5.94	3.76	0.00	1.94
700	1.49	6.3	50.71	20.51	7.63	7.13	4.14	0.00	1.63
800	1.12	8.2	57.60	24.53	9.30	7.95	4.55	0.00	1.07
900	0.59	8.2	62.77	27.13	10.14	8.26	5.44	0.00	0.52
1000	0.43	8.9	68.55	30.00	11.07	8.96	5.81	0.00	0.34
1100	0.28	8.7	72.52	32.59	11.94	9.49	6.17	0.00	0.21
1200	0.18	9.0	72.65	32.71	11.97	9.38	5.97	0.00	0.14

Table B.4: Number of events for the signal and background processes after the optimized selection per mass point value of the T' in the SR, computed using 2017 simulation. The signal efficiency is also provided.

$m_{T'}$ (GeV)	Signal	ϵ_{signal} (%)	$t\bar{t}$X	$t\bar{t}$1l	$t\bar{t}$2l	Multibosons	$tX(W/q)$	Others	S/B (%)
600	2.70	4.5	63.83	23.35	9.75	9.80	6.32	0.06	2.33
625	2.70	5.2	64.67	23.92	9.95	9.84	6.32	0.06	2.30
650	2.05	4.7	64.56	23.57	9.79	9.85	6.33	0.06	1.76
675	2.19	5.5	64.20	23.30	9.68	9.89	6.32	0.06	1.89
700	2.36	7.0	77.28	31.04	11.65	10.63	7.49	0.07	1.68
800	1.55	8.3	87.82	37.28	13.88	11.53	8.47	0.06	0.97
900	1.02	8.4	96.27	41.38	15.52	12.37	9.26	0.06	0.58
1000	0.66	7.9	105.72	46.33	17.01	13.05	9.85	0.06	0.34
1100	0.44	9.4	112.22	50.90	18.19	13.54	10.07	0.06	0.21
1200	0.27	10.0	112.35	50.61	18.19	13.60	10.09	0.06	0.13

Table B.5: Number of events for the signal and background processes after the optimized selection per mass point value of the T' in the SR, computed using 2018 simulation. The signal efficiency is also provided.

$m_{T'}$ (GeV)	Signal	$t\bar{t}$X	Non prompt	Charge flip	Multibosons	Other	Total background	Data
600	0.43	15.79	10.78	2.92	1.08	0.64	31.65	39
625	0.53	15.93	10.96	2.97	1.08	0.64	32.12	39
650	0.42	15.74	10.83	2.94	1.08	0.64	31.65	39
675	0.29	15.57	10.69	2.89	1.05	0.64	31.13	39
700	0.22	15.23	10.56	2.59	1.00	0.61	30.21	37
800	0.17	17.50	12.67	2.94	1.10	0.70	35.08	40
900	0.08	18.98	14.74	3.40	1.16	0.81	39.18	42
1000	0.05	20.17	15.96	3.62	1.21	0.84	41.85	43
1100	0.03	17.31	12.03	2.79	0.92	0.71	33.79	33
1200	0.02	17.42	12.38	2.87	0.94	0.68	34.30	33

Table B.6: Number of events for the signal and background processes after the optimized selection per mass point value of the T' in the ttX CR, comparing 2016 simulation and data. The number of data events is also provided.

$m_{T'}$ (GeV)	Signal	$t\bar{t}$X	Non prompt	Charge flip	Multibosons	Other	Total background	Data
600	0.55	18.77	9.73	1.82	1.59	0.67	33.13	45
625	0.54	18.86	9.89	1.82	1.59	0.67	33.37	45
650	0.38	18.65	9.81	1.82	1.57	0.67	32.90	45
675	0.24	18.52	9.71	1.77	1.57	0.66	32.48	45
700	0.31	17.93	10.28	1.61	1.21	0.50	31.83	42
800	0.17	20.58	12.56	1.85	1.47	0.77	37.40	49
900	0.11	22.25	14.09	2.04	1.56	0.94	40.98	52
1000	0.05	23.60	15.24	2.12	1.63	0.98	43.63	55
1100	0.03	20.38	11.76	1.54	1.12	0.77	35.61	47
1200	0.02	20.47	12.29	1.54	1.22	0.78	36.31	45

Table B.7: Number of events for the signal and background processes after the optimized selection per mass point value of the T' in the ttX CR, comparing 2017 simulation and data. The number of data events is also provided.

$m_{T'}$ (GeV)	Signal	$t\bar{t}$X	Non prompt	Charge flip	Multibosons	Other	Total background	Data
600	1.01	29.57	15.60	2.37	2.23	1.25	52.03	71
625	0.61	29.78	15.79	2.41	2.25	1.22	52.06	70
650	0.78	29.46	15.70	2.41	2.18	1.25	51.78	69
675	0.69	29.17	15.49	2.29	2.17	1.20	51.01	69
700	0.45	28.34	14.34	2.41	2.05	1.16	48.74	68
800	0.30	32.01	17.83	2.69	2.50	1.27	56.60	77
900	0.16	34.67	20.32	2.82	2.67	1.41	62.05	83
1000	0.09	36.96	22.55	2.99	2.70	1.52	66.81	91
1100	0.04	31.67	17.57	2.31	2.26	1.22	55.08	78
1200	0.03	32.10	18.05	2.39	2.14	1.18	55.86	80

Table B.8: Number of events for the signal and background processes after the optimized selection per mass point value of the T' in the ttX CR, comparing 2018 simulation and data. The number of data events is also provided.

$m_{T'}$ (GeV)	Signal	$t\bar{t}$2l	Non prompt	$tX(W/q)$	DY	Other	Total background	Data
600	1.52	10770.64	868.40	615.83	260.58	92.13	12609.09	12255
625	1.85	10961.99	887.19	625.54	226.34	93.44	12832.36	12459
650	1.70	10905.70	880.02	624.08	260.10	93.15	12764.76	12400
675	1.24	10775.32	868.84	617.75	260.21	92.53	12615.89	12256
700	1.33	13137.64	1092.69	727.26	322.23	111.23	15392.38	14995
800	0.92	15649.31	1352.52	813.34	427.66	127.06	18370.81	17988
900	0.62	17308.99	1541.10	889.19	477.78	139.57	20357.25	19997
1000	0.39	19238.14	1763.04	960.14	510.11	152.87	22624.68	22334
1100	0.28	20416.80	1885.69	999.29	535.94	160.91	23978.92	23696
1200	0.15	20418.74	1869.85	1006.73	528.04	161.45	23984.96	23694

Table B.9: Number of events for the signal and background processes after the optimized selection per mass point value of the T' in the tt2l MR, comparing 2016 simulation and data. The number of data events is also provided.

$m_{T'}$ (GeV)	Signal	$t\bar{t}$2l	Non prompt	$tX(W/q)$	DY	Other	Total background	Data
600	2.14	12626.87	920.20	745.60	292.52	111.24	14698.59	15172
625	1.93	12849.53	938.71	755.69	292.49	112.43	14950.77	15421
650	1.52	12787.62	932.54	753.27	292.49	112.26	14879.70	15344
675	1.43	12633.04	920.03	747.61	284.18	111.53	14697.82	15166
700	1.45	15395.40	1151.64	877.03	351.29	132.42	17909.22	18575
800	1.10	18302.16	1416.80	986.34	472.70	151.58	21330.68	22219
900	0.67	20186.06	1609.24	1061.99	557.79	164.88	23580.63	24605
1000	0.44	22404.68	1834.59	1148.38	616.08	180.19	26184.36	27443
1100	0.26	23779.95	1945.74	1190.50	645.58	189.80	27751.83	29076
1200	0.16	23787.30	1948.84	1193.85	634.47	190.71	27755.34	29087

Table B.10: Number of events for the signal and background processes after the optimized selection per mass point value of the T' in the tt2l MR, comparing 2017 simulation and data. The number of data events is also provided.

$m_{T'}$ (GeV)	Signal	$t\bar{t}$2l	Non prompt	$tX(W/q)$	DY	Other	Total background	Data
600	2.64	17908.16	1355.56	1097.65	401.66	168.58	20934.26	21110
625	2.51	18226.65	1382.93	1107.71	401.66	170.93	21292.39	21436
650	2.22	18138.31	1374.55	1105.02	401.66	170.48	21192.25	21352
675	1.97	17917.44	1357.71	1097.63	394.09	169.11	20937.95	21102
700	2.23	21895.69	170..81	1278.86	517.17	202.53	25597.30	25811
800	1.64	26122.98	2104.07	1447.01	658.55	230.43	30564.68	31095
900	1.01	28879.89	2398.24	1557.51	770.26	251.18	33858.09	34511
1000	0.65	32129.75	2743.92	1681.66	830.33	275.51	37661.32	38451
1100	0.45	34127.89	2913.05	1742.42	894.26	290.14	39968.22	40770
1200	0.26	34137.84	2918.26	1749.93	870.65	290.74	39967.68	40778

Table B.11: Number of events for the signal and background processes after the optimized selection per mass point value of the T' in the tt2l MR, comparing 2018 simulation and data. The number of data events is also provided.

$m_{T'}$ (GeV)	Signal	Non prompt	$t\bar{t}$X	Charge flip	Other	Total background	Data
600	0.62	239.40	8.70	3.33	1.74	253.79	204
625	0.69	243.87	8.82	3.36	1.75	258.48	208
650	0.50	242.80	8.81	3.36	1.74	257.21	207
675	0.48	239.00	8.74	3.28	1.75	253.26	204
700	0.45	306.20	10.84	4.11	2.43	324.02	262
800	0.33	388.55	12.97	5.10	2.83	409.79	340
900	0.22	444.16	14.69	5.63	3.25	467.94	396
1000	0.12	510.74	16.60	6.13	3.82	537.41	455
1100	0.09	553.65	17.85	6.57	4.07	582.24	490
1200	0.07	550.89	17.92	6.68	4.05	579.61	487

Table B.12: Number of events for the signal and background processes after the optimized selection per mass point value of the T' in the tt1l MR of the SR, comparing 2016 simulation and data. The number of data events is also provided.

$m_{T'}$ (GeV)	Signal	Non prompt	$t\bar{t}$X	Charge flip	Other	Total background	Data
600	0.70	240.34	9.08	1.69	1.70	254.21	268
625	0.67	244.67	9.24	1.74	2.40	258.73	274
650	0.50	242.85	9.23	1.74	2.40	256.72	271
675	0.52	238.74	9.12	1.70	2.39	252.47	268
700	0.46	307.45	11.46	2.16	3.25	324.78	345
800	0.32	389.16	13.82	2.75	3.77	409.83	435
900	0.23	451.76	15.59	3.18	4.27	475.03	496
1000	0.17	515.60	17.86	3.38	4.87	541.88	551
1100	0.09	560.41	19.28	3.69	5.12	588.59	602
1200	0.05	557.17	19.27	3.69	5.09	585.27	605

Table B.13: Number of events for the signal and background processes after the optimized selection per mass point value of the T' in the tt1l MR of the SR, comparing 2017 simulation and data. The number of data events is also provided.

$m_{T'}$ (GeV)	Signal	Non prompt	$t\bar{t}$X	Charge flip	Other	Total background	Data
600	1.10	366.40	13.31	3.33	2.97	387.11	379
625	0.89	373.15	13.56	3.33	3.12	394.04	389
650	0.49	370.81	13.48	3.33	3.01	391.12	386
675	0.59	365.31	13.35	3.33	2.98	385.57	378
700	0.88	467.35	16.84	3.75	3.88	492.69	480
800	0.51	588.22	20.12	4.90	4.85	618.59	634
900	0.36	669.02	22.76	5.48	5.56	703.17	721
1000	0.29	761.94	25.75	5.97	6.91	800.85	814
1100	0.12	826.60	27.65	6.23	7.08	867.67	879
1200	0.08	816.66	24.71	6.27	5.44	853.17	869

Table B.14: Number of events for the signal and background processes after the optimized selection per mass point value of the T' in the tt1l MR of the SR, comparing 2018 simulation and data. The number of data events is also provided.

$m_{T'}$ (GeV)	Signal	Non prompt	$t\bar{t}$X	Charge flip	Other	Total background	Data
600	0.23	92.74	4.16	1.26	0.60	98.99	83
625	0.20	94.01	4.19	1.29	0.59	100.29	83
650	0.17	93.09	4.14	1.29	0.59	99.28	83
675	0.11	91.28	4.08	1.26	0.58	97.32	82
700	0.13	92.13	4.10	1.40	0.49	98.27	77
800	0.05	118.24	5.01	1.53	0.68	125.51	104
900	0.03	141.97	5.80	1.66	0.79	150.25	125
1000	0.01	159.86	6.41	1.88	0.96	169.12	136
1100	0.01	120.96	5.29	1.48	0.75	128.48	105
1200	0.007	124.93	5.35	1.48	0.78	132.53	107

Table B.15: Number of events for the signal and background processes after the optimized selection per mass point value of the T' in the tt1l MR of the ttX CR, comparing 2016 simulation and data. The number of data events is also provided.

$m_{T'}$ (GeV)	Signal	Non prompt	$t\bar{t}$X	Charge flip	Other	Total background	Data
600	0.20	99.32	4.39	0.59	0.78	105.28	100
625	0.20	100.62	4.52	0.64	0.91	106.89	104
650	0.13	100.13	4.47	0.63	0.91	106.28	101
675	0.10	98.38	4.41	0.59	0.91	104.40	99
700	0.07	101.15	4.46	0.60	0.81	107.11	93
800	0.04	127.31	5.48	0.73	0.97	134.53	108
900	0.04	149.68	6.43	0.92	1.03	158.11	143
1000	0.02	167.55	7.19	1.00	1.09	176.86	166
1100	0.01	126.49	5.97	0.73	0.87	134.07	117
1200	0.01	131.24	6.08	0.73	0.87	138.92	120

Table B.16: Number of events for the signal and background processes after the optimized selection per mass point value of the T' in the tt1l MR of the ttX CR, comparing 2017 simulation and data. The number of data events is also provided.

$m_{T'}$ (GeV)	Signal	Non prompt	$t\bar{t}$X	Charge flip	Other	Total background	Data
600	0.26	139.98	6.95	0.94	1.11	149.24	144
625	0.22	140.96	7.03	0.93	1.12	150.26	145
650	0.11	140.09	6.95	0.94	1.11	149.20	146
675	0.11	137.81	6.89	0.94	1.10	146.84	143
700	0.06	137.62	6.87	0.87	1.19	146.61	141
800	0.07	175.52	8.50	1.17	1.27	186.53	180
900	0.05	212.74	9.77	1.30	1.48	225.34	217
1000	0.03	238.13	10.77	1.48	1.62	252.02	249
1100	0.01	180.59	9.16	1.26	1.54	192.55	189
1200	0.01	187.80	9.32	1.26	1.53	199.92	202

Table B.17: Number of events for the signal and background processes after the optimized selection per mass point value of the T' in the tt1l MR of the ttX CR, comparing 2018 simulation and data. The number of data events is also provided.

Appendix C

Signal extraction per mass point value of the Vector-Like Quark T'

C.1 Systematic uncertainties

The impact of the shape systematic uncertainties are presented for the signal from Figures C.1 to C.14. The procedure is discussed in Chapter 5. The variations are presented for a nominal mass of the VLQ T' $m_{T'} = 700$ GeV in the SR, with 2018 simulated samples. The RelativeSample and RelativeBal JES uncertainties introduce a variation of the distributions by 60% at maximum, with large statistical fluctuations. The rest of the uncertainties introduce a variation of the distributions by 15% at maximum. These results are similar for all the years, regions, and mass point values of the T'.

C.2 Impact of the nuisance parameters

The impact of the nuisance parameters for the fits are presented from Figures C.15 to C.26 for 2016, 2017, and 2018 combined. The procedure is discussed in Chapter 5. The fit includes the $\mu\mu$, eμ, and ee channels, as well as the SR and ttX CR, and are presented per mass point value of the VLQ T'. For $m_{T'} = 700$ GeV, the impacts are presented separately for 2016, 2017, and 2018. Only Asimov datasets are considered, with a signal injection at $r = 1$. The variable $M_{T,T'}$ serves as the discriminant observable.

C.3 Likelihood fit

The profiled likelihood fits are presented per mass point value of the T' from Tables C.1 to C.4 for 2016, 2017, and 2018 separately, then combined for the three years. The procedure is discussed in Chapter 5. The fits include the $\mu\mu$, eμ, and ee channels, as well as the SR and ttX CR, with $m_{T'} = 700$ GeV. Only Asimov datasets are considered, with a signal injection at $r = 1$. The variable $M_{T,T'}$ serves as the discriminant observable.

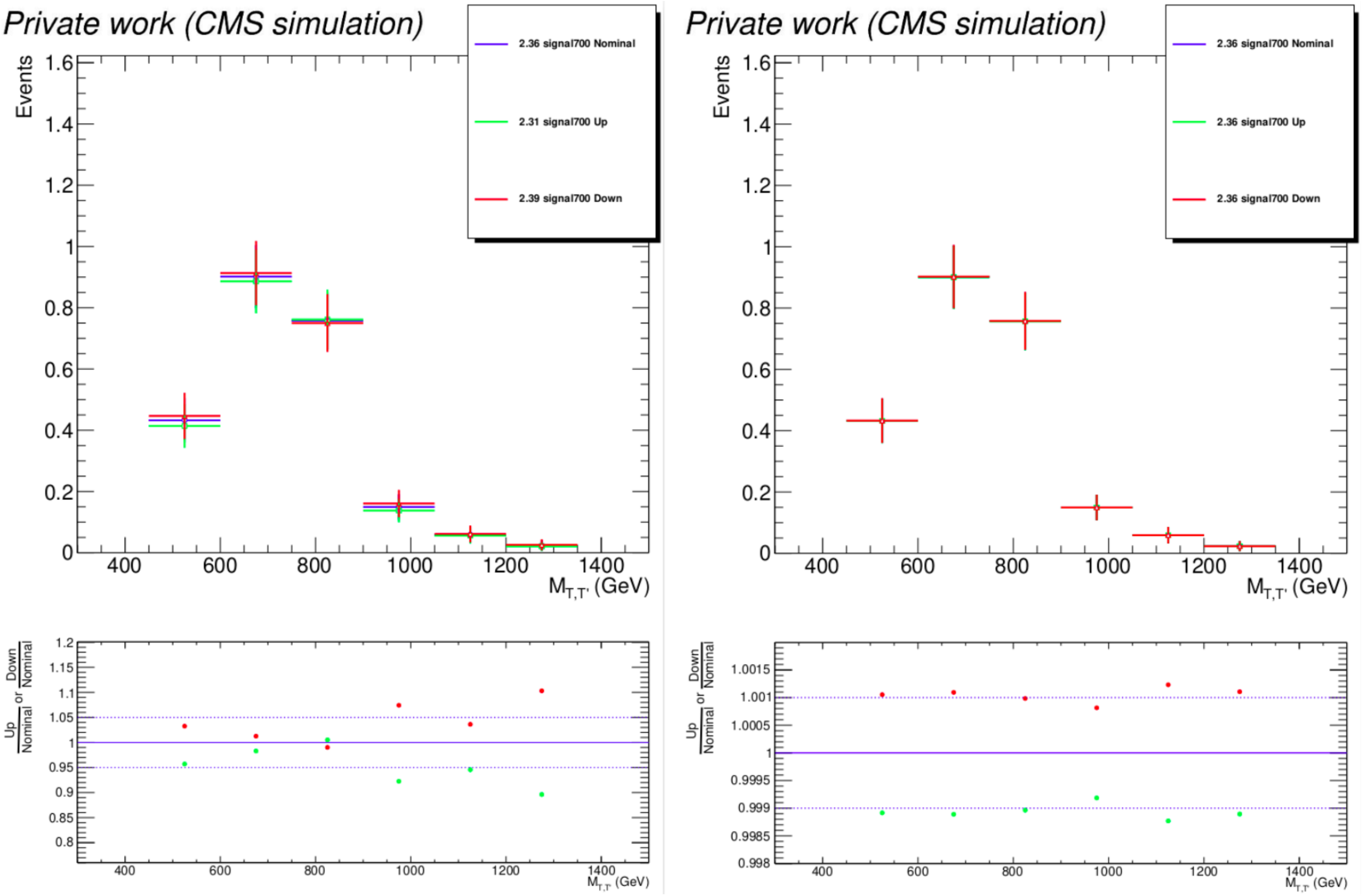


Figure C.1: Distribution of $M_{T,T'}$ for the nominal (blue), up (green) and down (red) variations of the PU (left) and prefiring (right) uncertainties, computed using 2018 simulation for the signal with $m_{T'} = 700$ GeV.

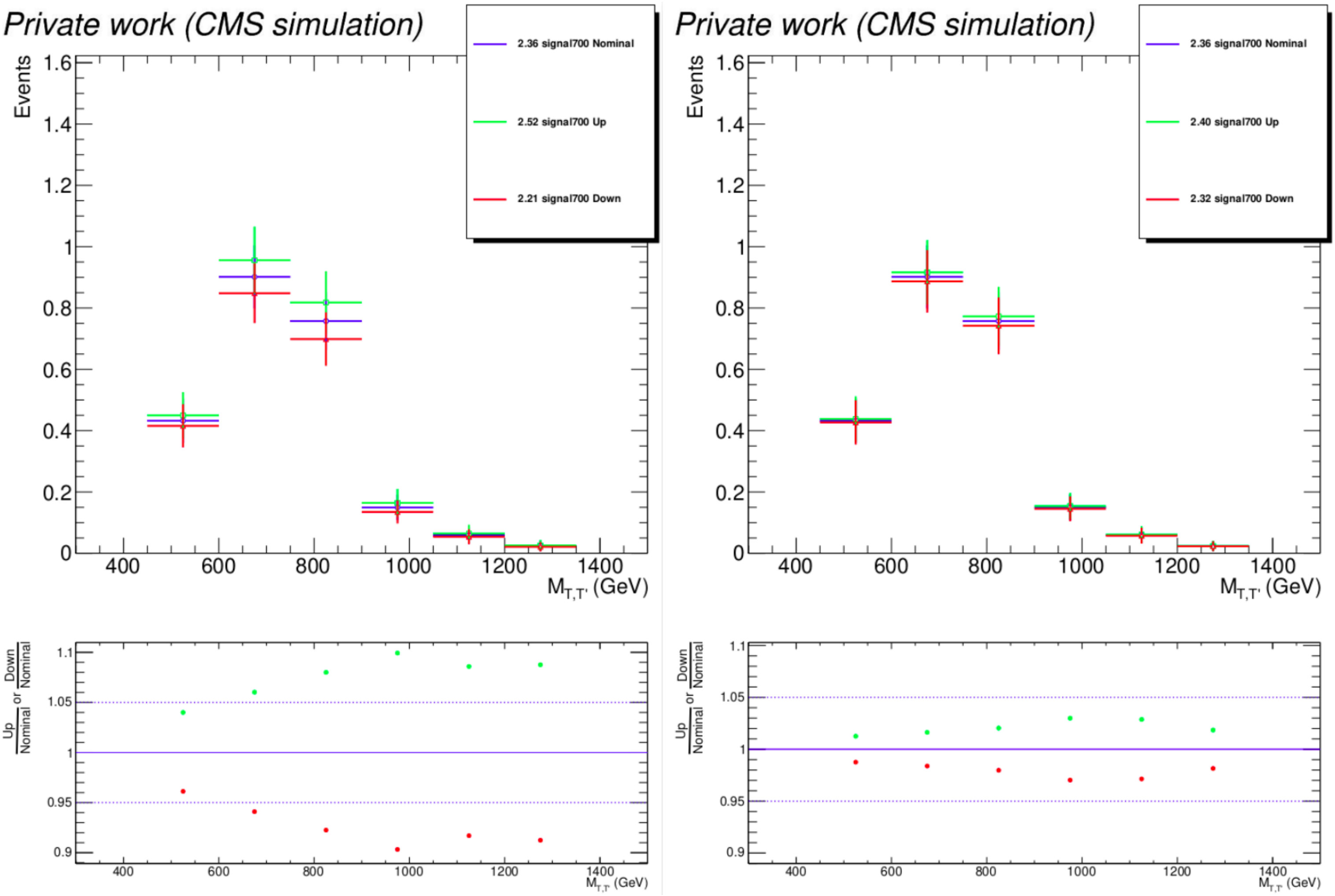


Figure C.2: Distribution of $M_{T,T'}$ for the nominal (blue), up (green) and down (red) variations of the b-tagging uncertainty for the heavy flavors, computed using 2018 simulation for the signal, with $m_{T'} = 700$ GeV and both the correlated (left) and uncorrelated (right) components of the uncertainty.

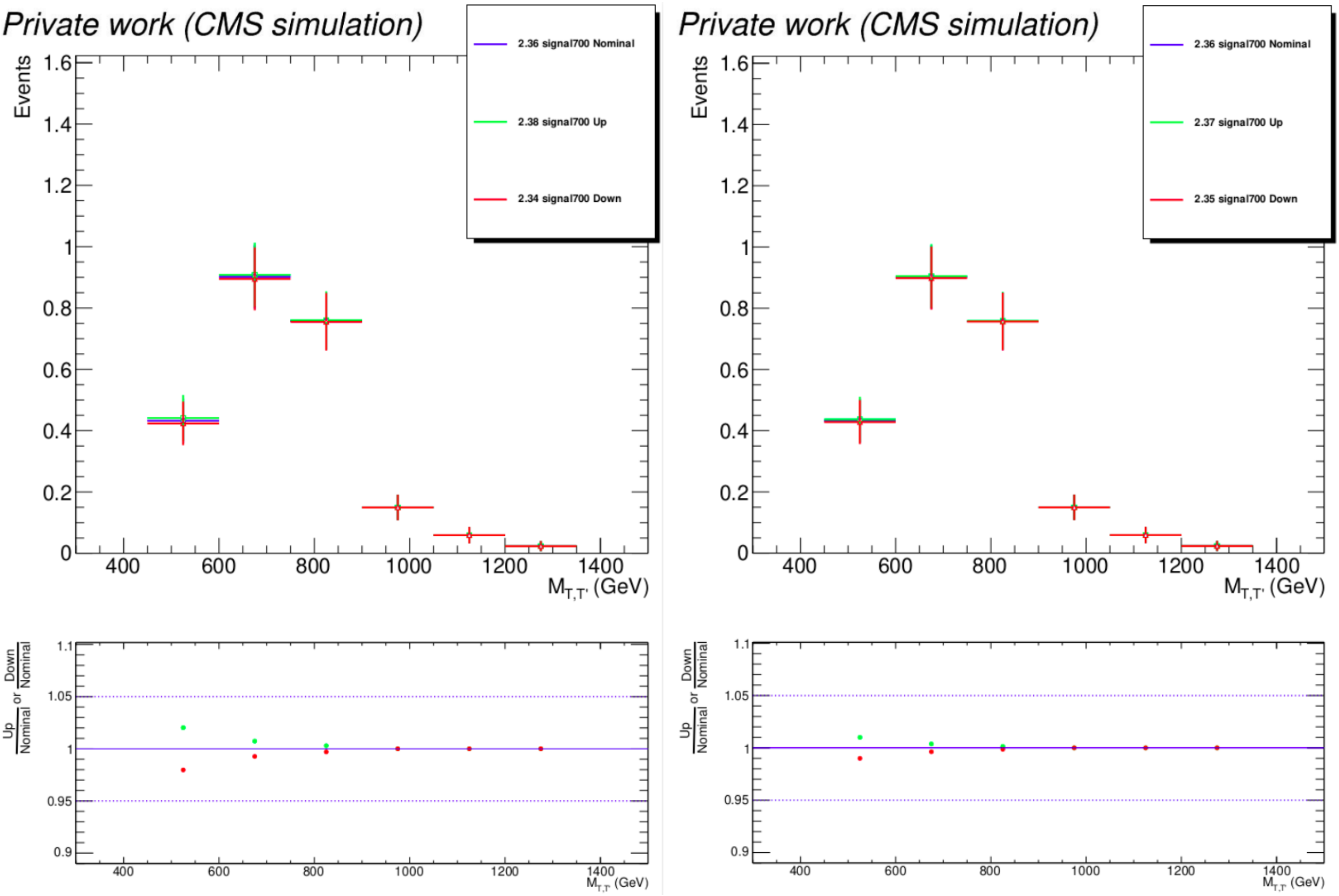


Figure C.3: Distribution of $M_{T,T'}$ for the nominal (blue), up (green) and down (red) variations of the b-tagging uncertainty for the light flavors, computed using 2018 simulation for the signal, with $m_{T'} = 700$ GeV and both the correlated (left) and uncorrelated (right) components of the uncertainty.

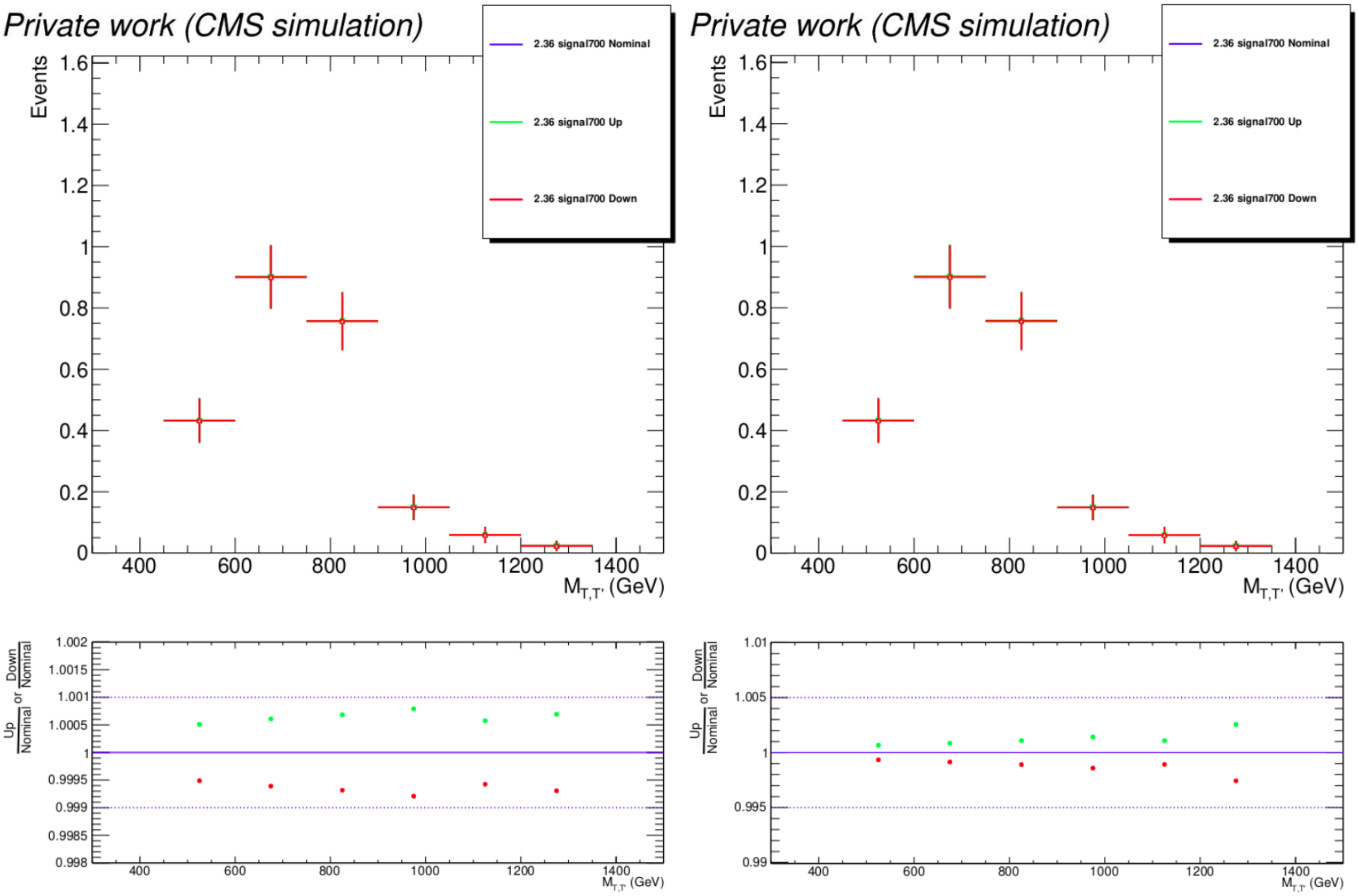


Figure C.4: Distribution of $M_{T,T'}$ for the nominal (blue), up (green) and down (red) variations of the electron trigger uncertainty, computed using 2018 simulation for the signal, with $m_{T'} = 700$ GeV and both the correlated (left) and uncorrelated (right) components of the uncertainty.

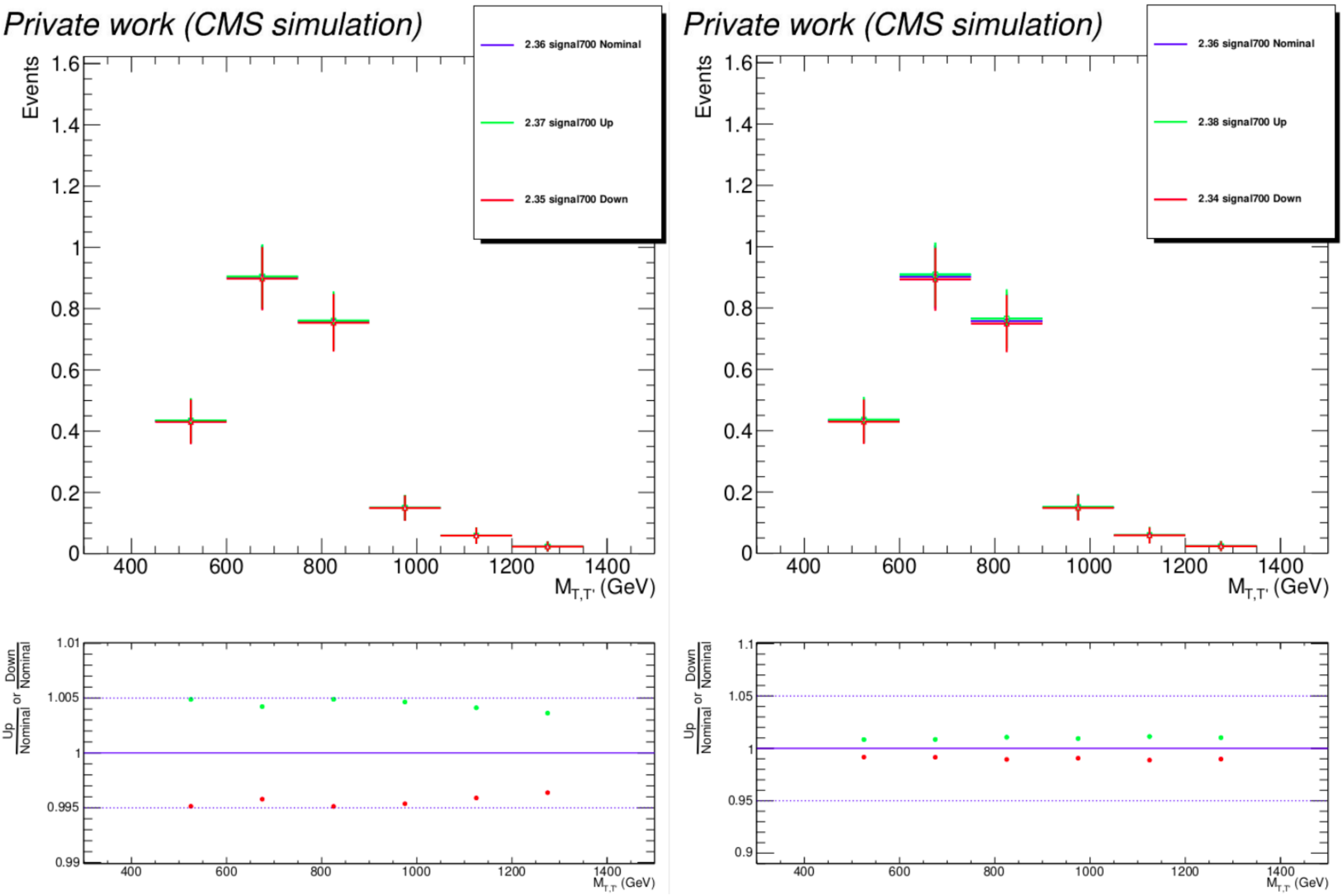


Figure C.5: Distribution of $M_{T,T'}$ for the nominal (blue), up (green) and down (red) variations of the electron reconstruction (left) and ID (right) uncertainties, computed using 2018 simulation for the signal with $m_{T'} = 700$ GeV.

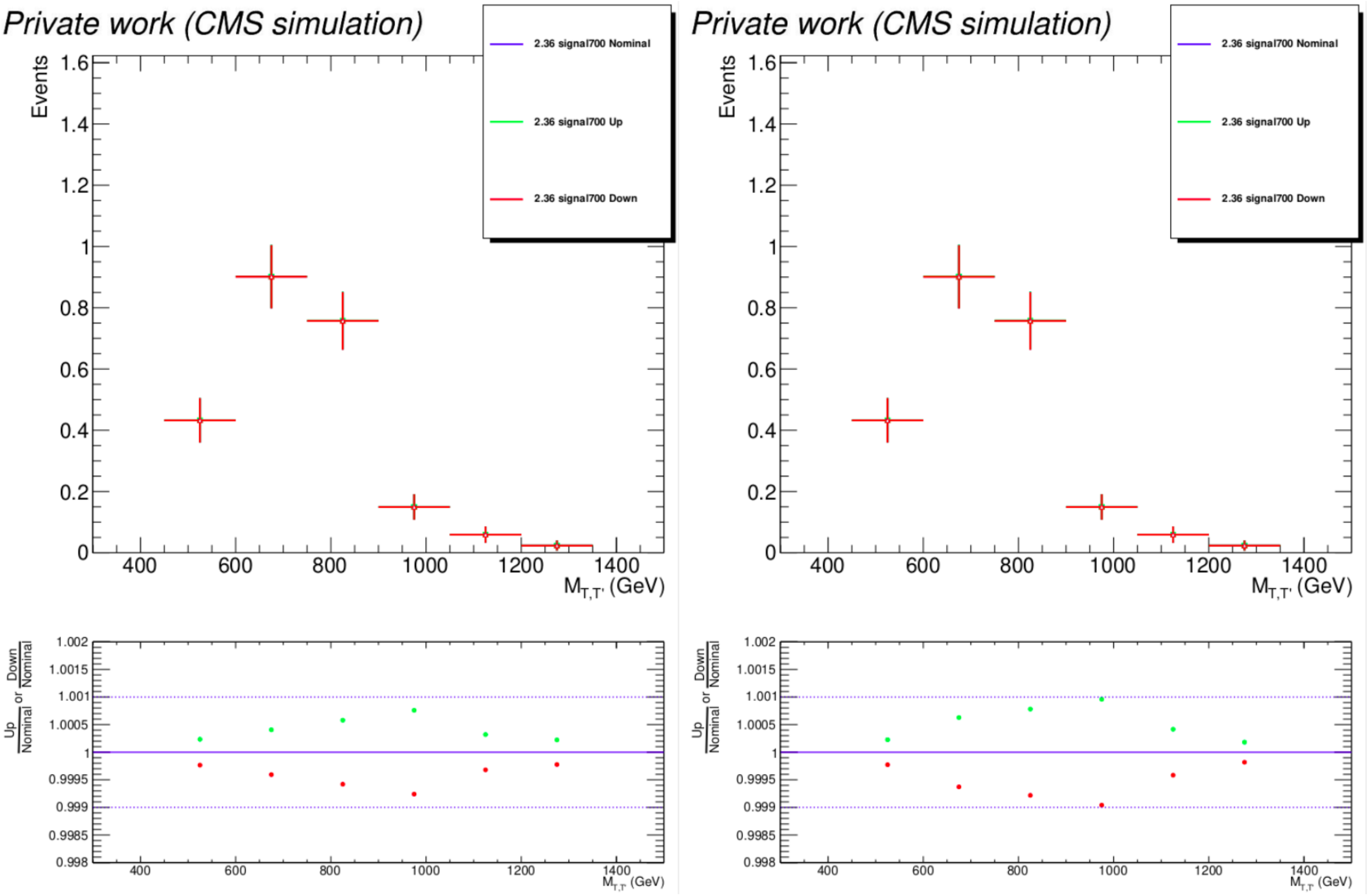


Figure C.6: Distribution of $M_{T,T'}$ for the nominal (blue), up (green) and down (red) variations of the muon trigger uncertainty, computed using 2018 simulation for the signal, with $m_{T'} = 700$ GeV and both the correlated (left) and uncorrelated (right) components of the uncertainty.

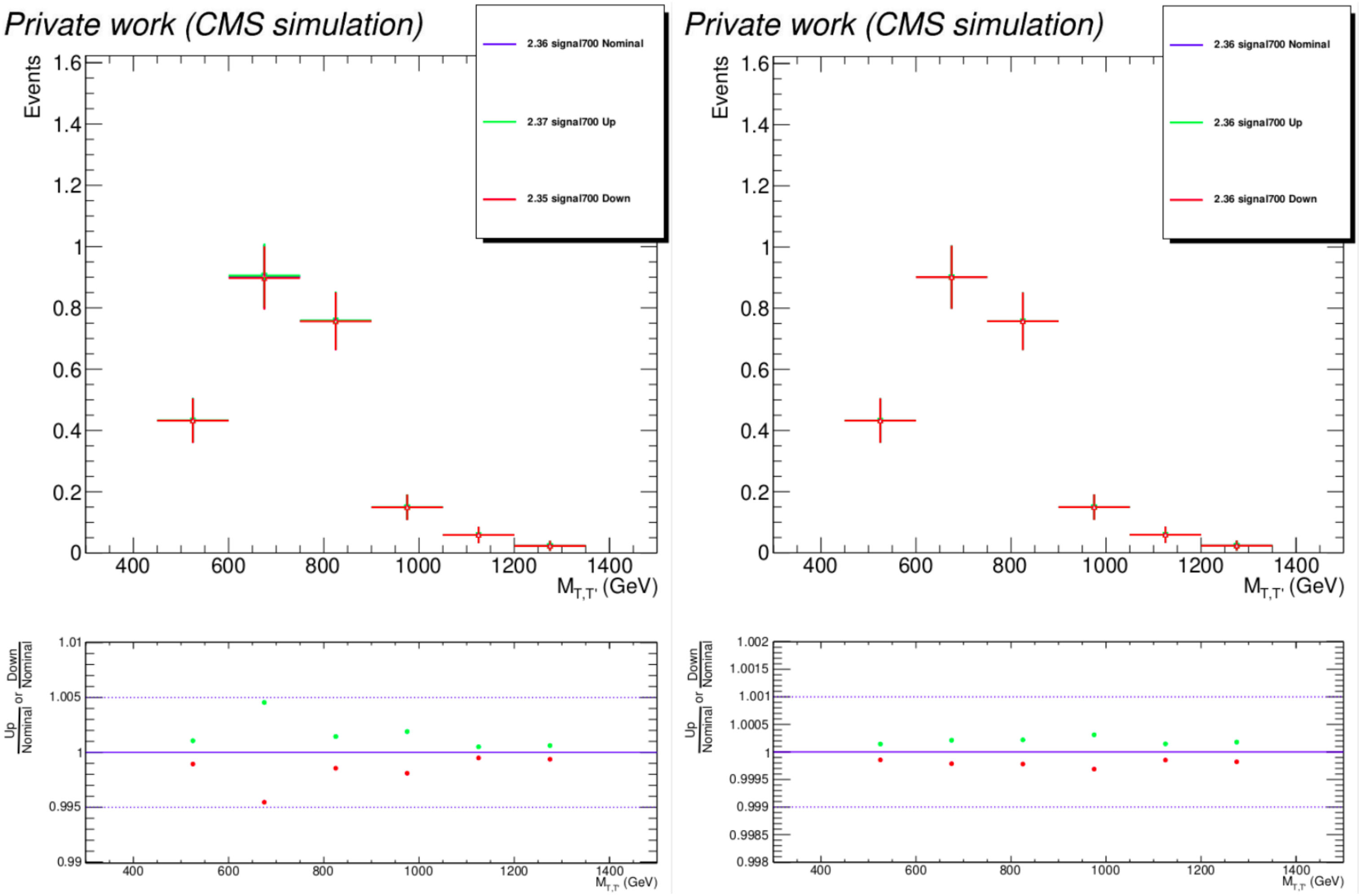


Figure C.7: Distribution of $M_{T,T'}$ for the nominal (blue), up (green) and down (red) variations of the muon reconstruction uncertainty, computed using 2018 simulation for the signal, with $m_{T'}$ = 700 GeV and both the correlated (left) and uncorrelated (right) components of the uncertainty.

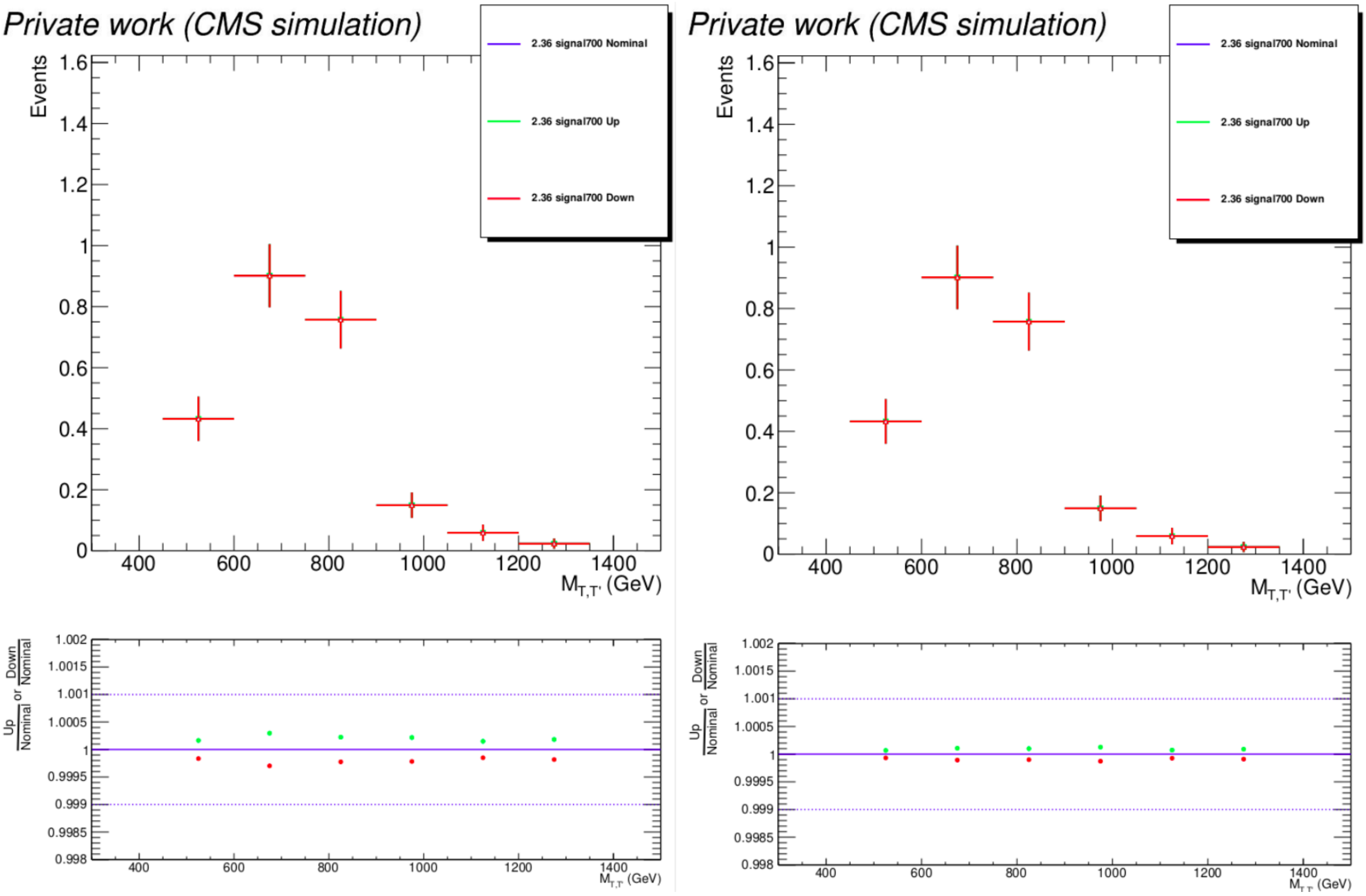


Figure C.8: Distribution of $M_{T,T'}$ for the nominal (blue), up (green) and down (red) variations of the muon ID uncertainty, computed using 2018 simulation for the signal, with $m_{T'} = 700$ GeV and both the correlated (left) and uncorrelated (right) components of the uncertainty.

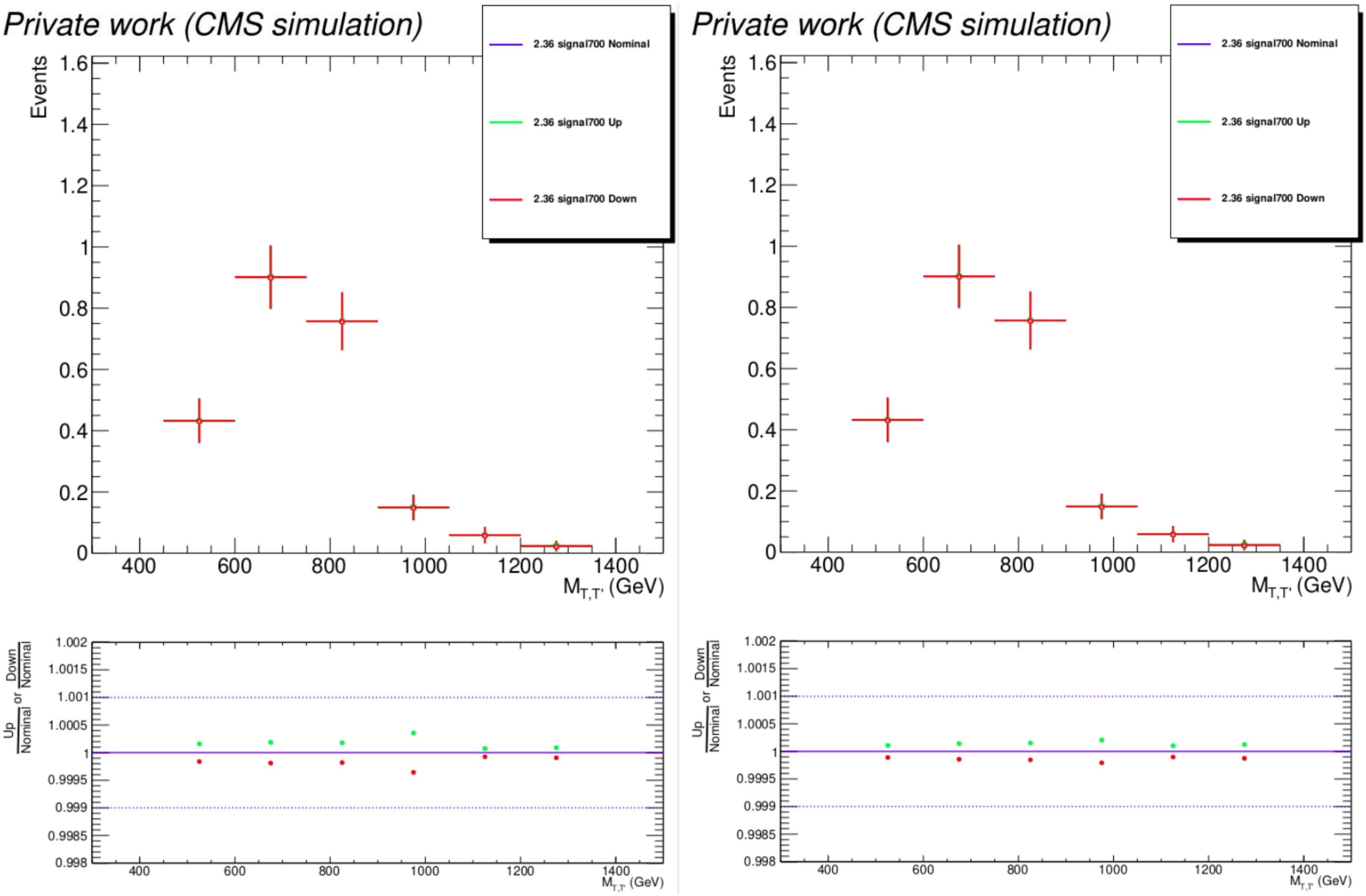


Figure C.9: Distribution of $M_{T,T'}$ for the nominal (blue), up (green) and down (red) variations of the muon isolation uncertainty, computed using 2018 simulation for the signal, with $m_{T'} = 700$ GeV and both the correlated (left) and uncorrelated (right) components of the uncertainty.

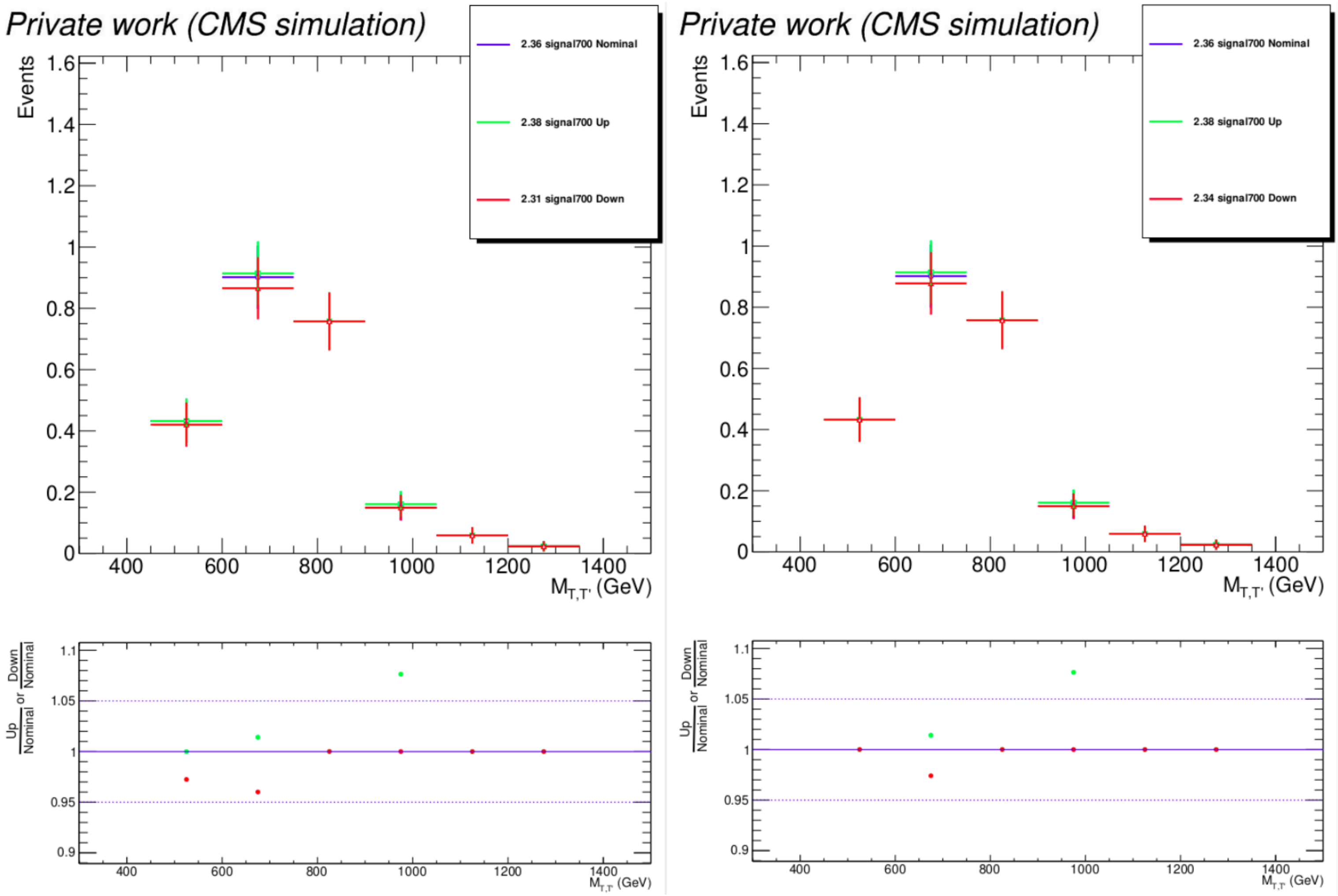


Figure C.10: Distribution of $M_{T,T'}$ for the nominal (blue), up (green) and down (red) variations of the Absolute JES uncertainty, computed using 2018 simulation for the signal, with $m_{T'} = 700$ GeV and both the correlated (left) and uncorrelated (right) components of the uncertainty.

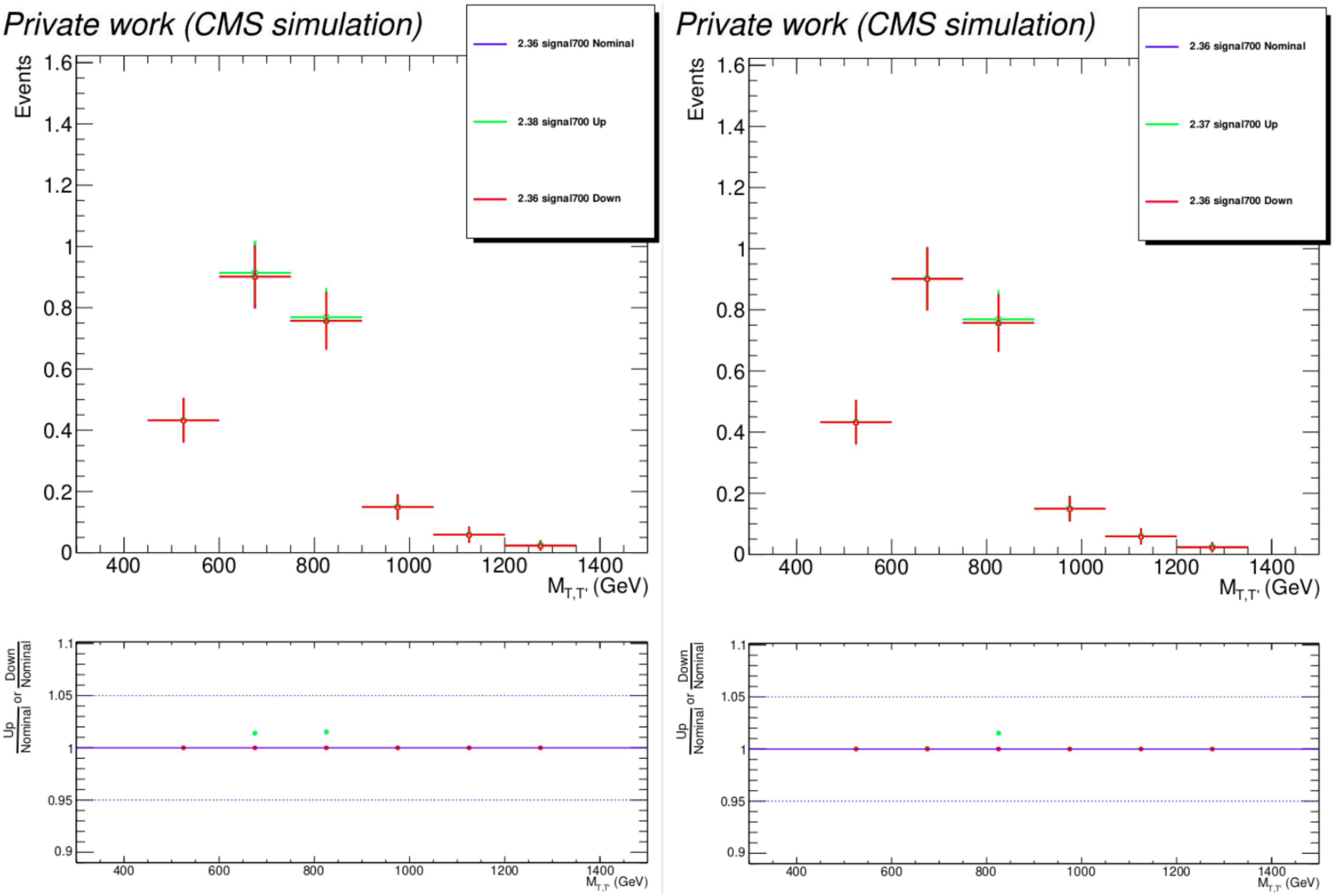


Figure C.11: Distribution of $M_{T,T'}$ for the nominal (blue), up (green) and down (red) variations of the BBEC1 JES uncertainty, computed using 2018 simulation for the signal, with $m_{T'} = 700$ GeV and both the correlated (left) and uncorrelated (right) components of the uncertainty.

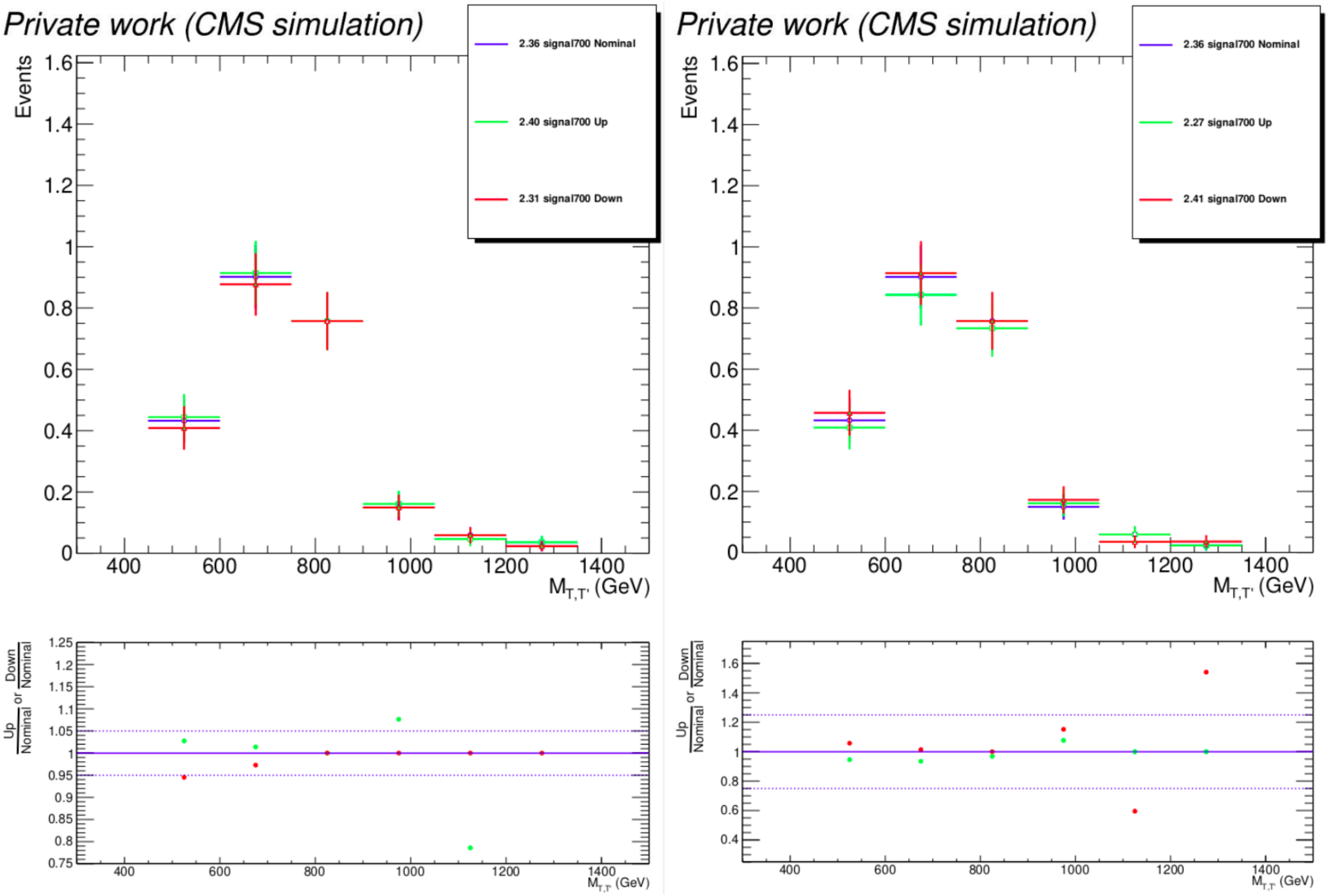


Figure C.12: Distribution of $M_{T,T'}$ for the nominal (blue), up (green) and down (red) variations of the FlavorQCD (left) and RelativeSample (right) JES uncertainties, computed using 2018 simulation for the signal with $m_{T'} = 700$ GeV.

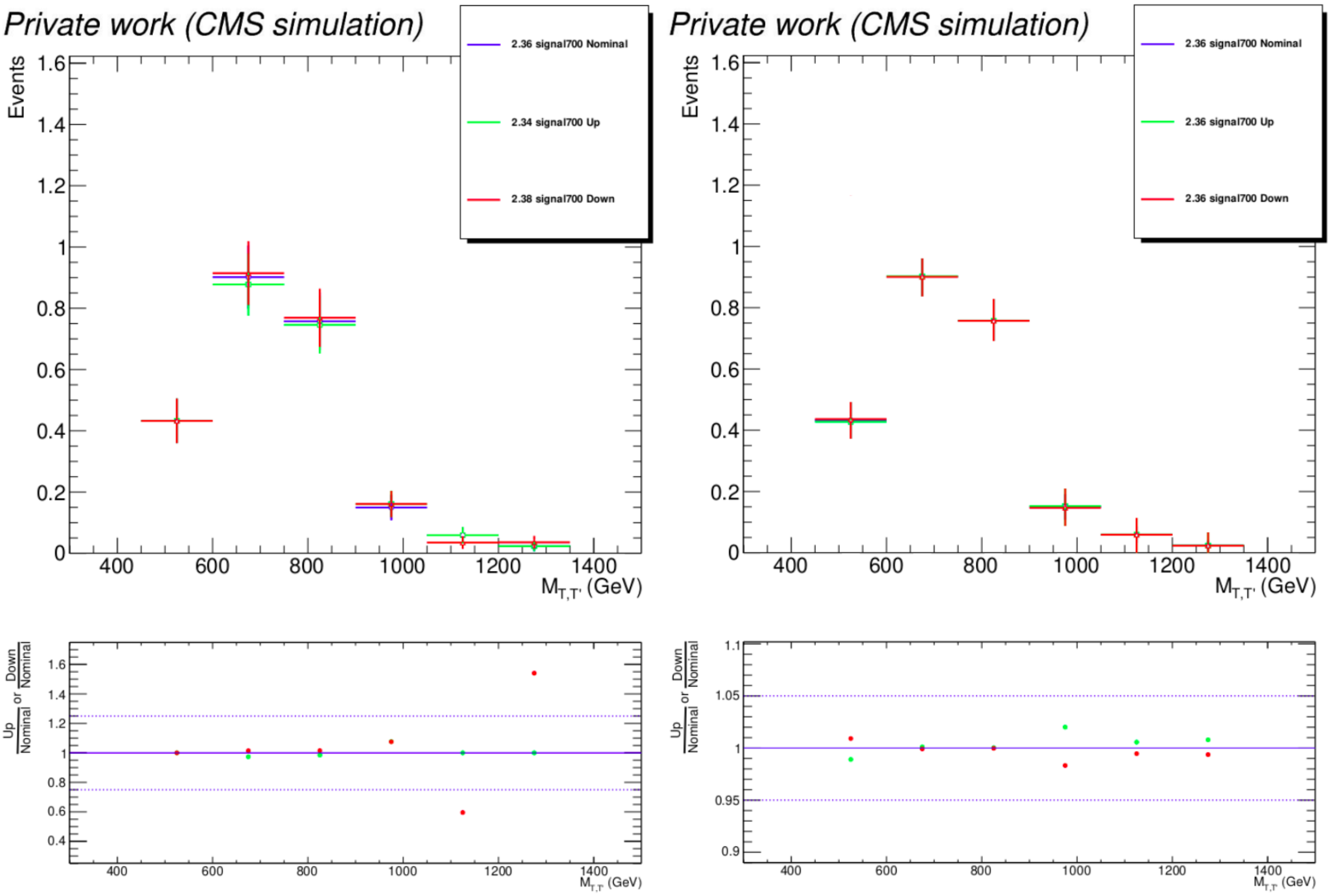


Figure C.13: Distribution of $M_{T,T'}$ for the nominal (blue), up (green) and down (red) variations of the RelativeBal JES (left) and the renormalization and factorization scale (right) uncertainties, computed using 2018 simulation for the signal with $m_{T'} = 700$ GeV.

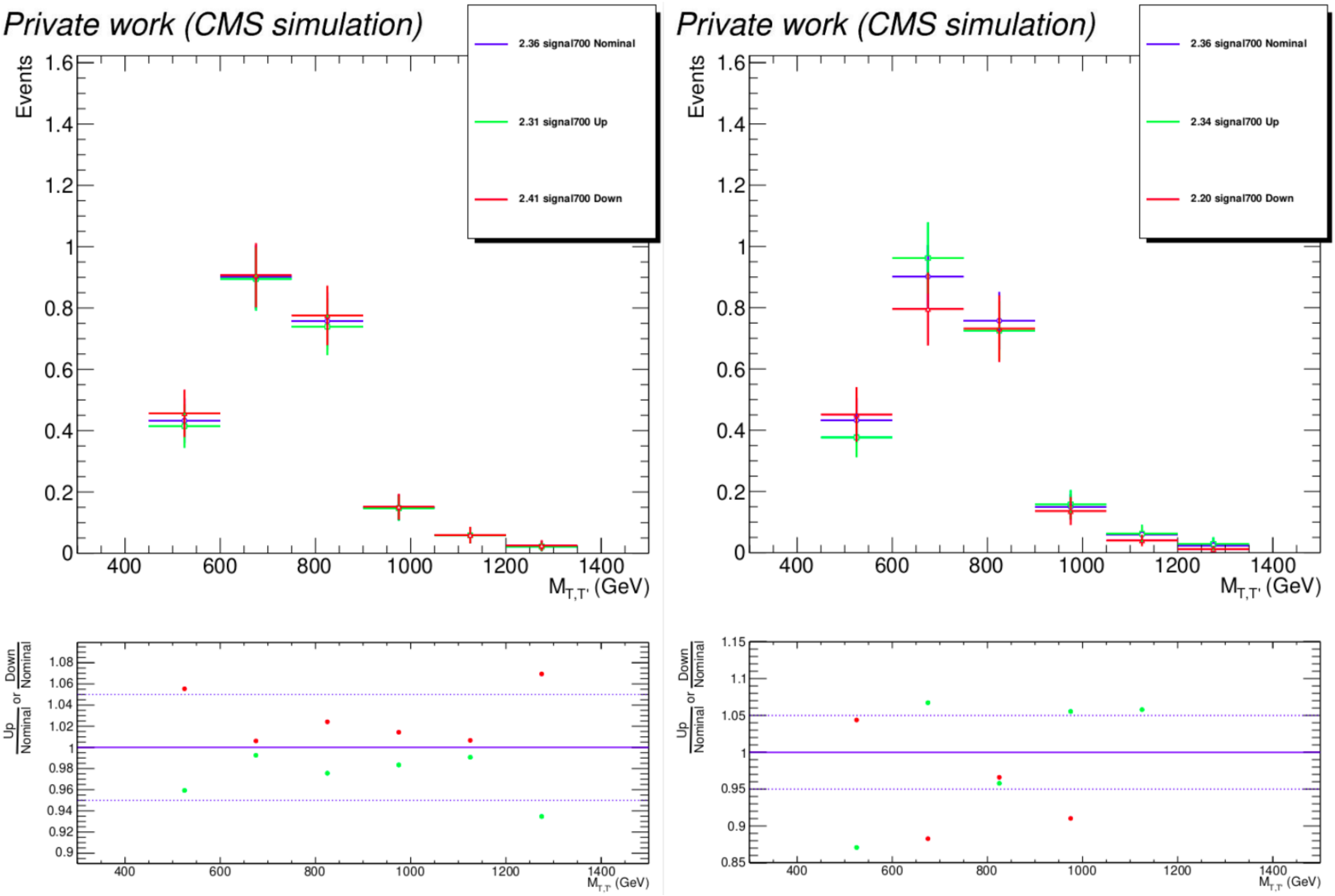


Figure C.14: Distribution of $M_{T,T'}$ for the nominal (blue), up (green) and down (red) variations of the ISR (left) and FSR (right) uncertainties, computed using 2018 simulation for the signal with $m_{T'} = 700$ GeV.

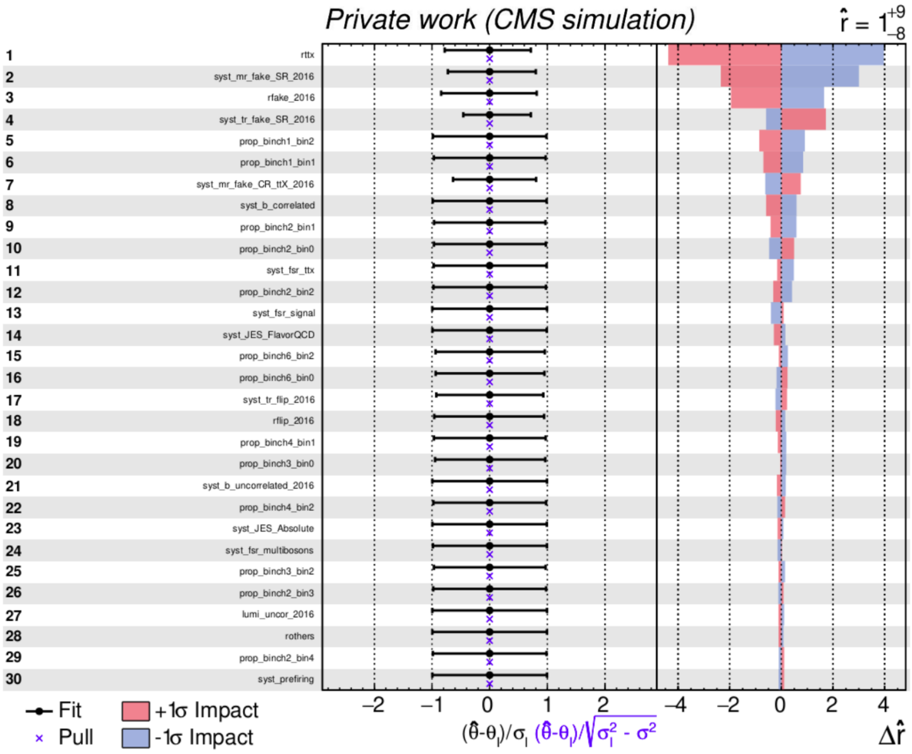


Figure C.15: Impact of the nuisance parameters for the Run 2 period with all the channels and regions combined, computed using 2016 simulation with a signal injection of $r = 1$ and $m_{T'} = 700$ GeV. The 30 first most important impacts are included.

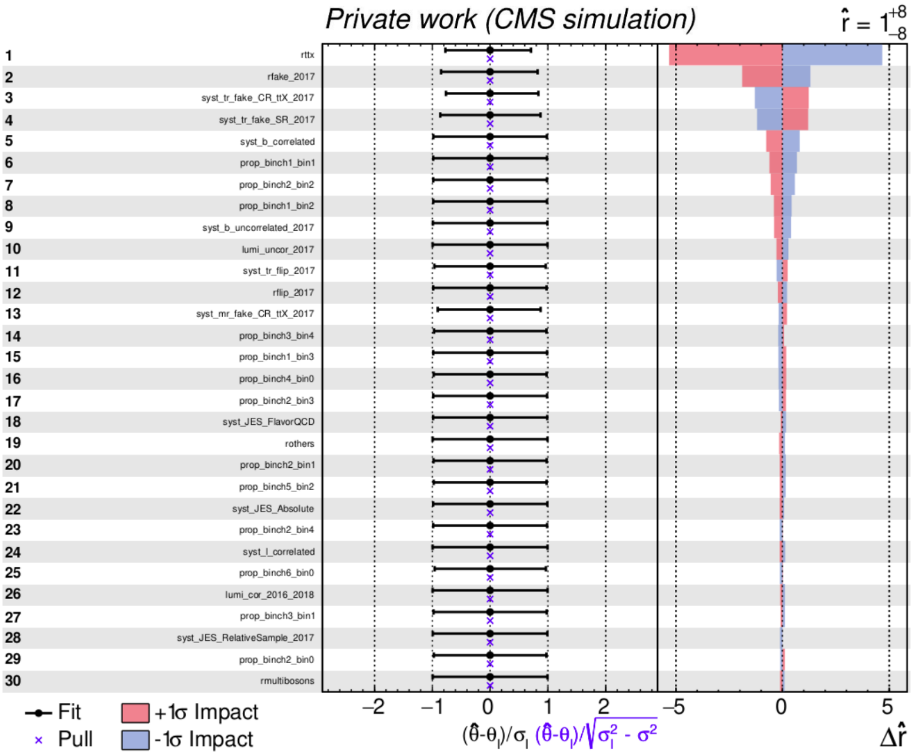


Figure C.16: Impact of the nuisance parameters for the Run 2 period with all the channels and regions combined, computed using 2017 simulation with a signal injection of $r = 1$ and $m_{T'} = 700$ GeV. The 30 first most important impacts are included.

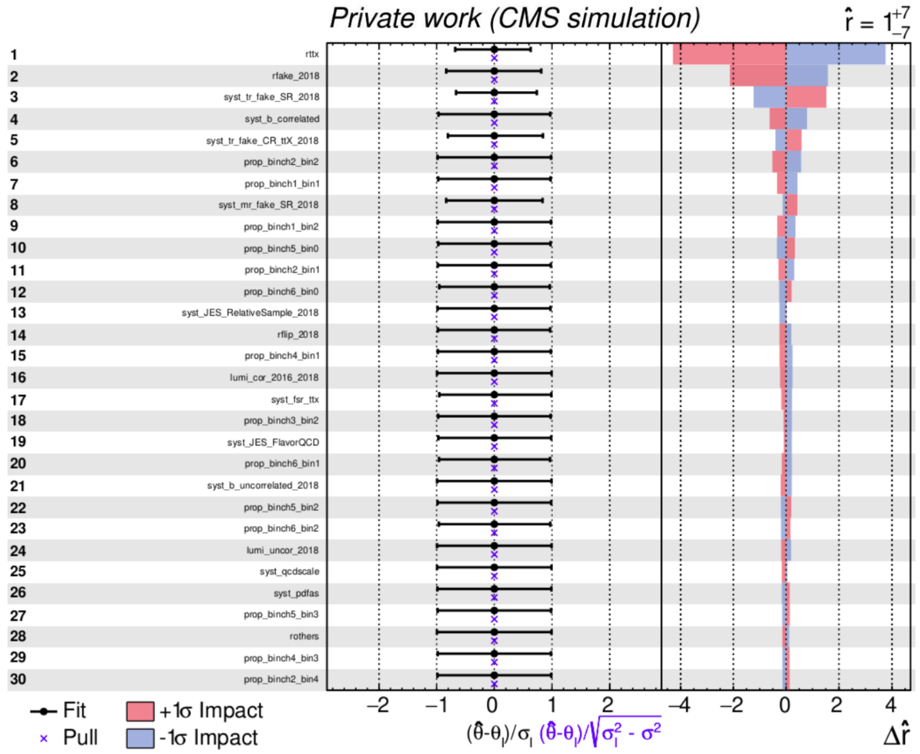


Figure C.17: Impact of the nuisance parameters for the Run 2 period with all the channels and regions combined, computed using 2018 simulation with a signal injection of $r = 1$ and $m_{T'} = 700$ GeV. The 30 first most important impacts are included.

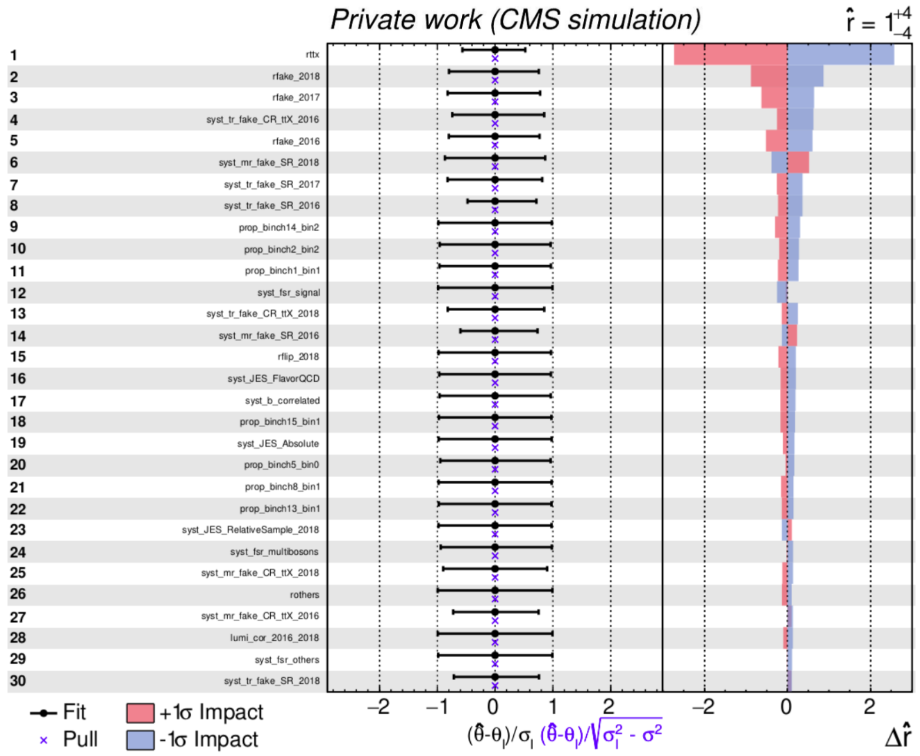


Figure C.18: Impact of the nuisance parameters for the Run 2 period with all the channels and regions combined, computed using Run 2 simulation with a signal injection of $r = 1$ and $m_{T'} = 600$ GeV. The 30 first most important impacts are included.

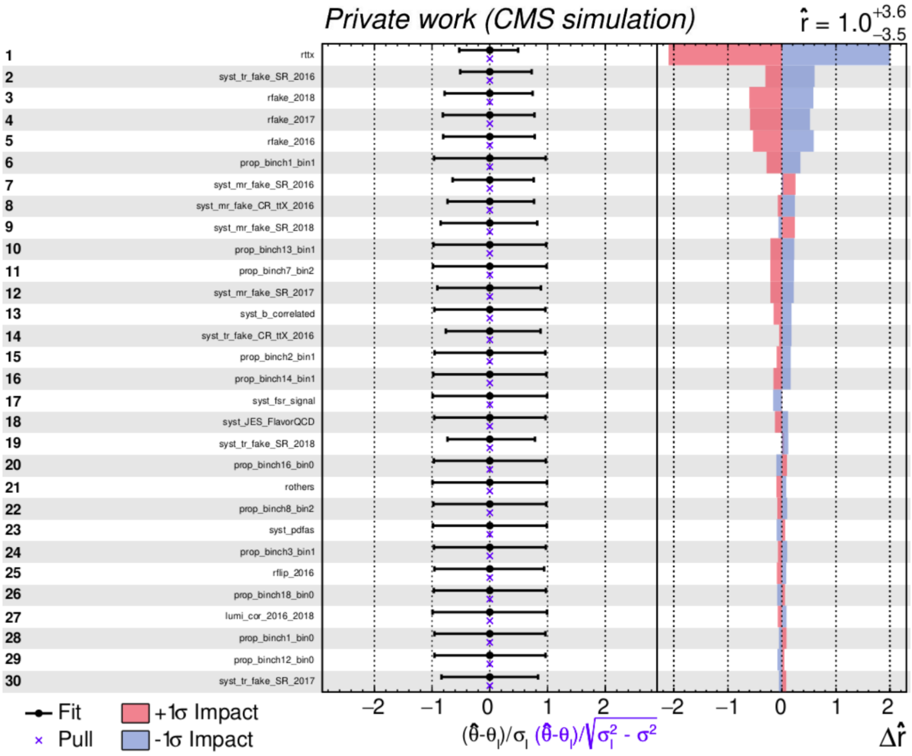


Figure C.19: Impact of the nuisance parameters for the Run 2 period with all the channels and regions combined, computed using Run 2 simulation with a signal injection of $r = 1$ and $m_{T'} = 625$ GeV. The 30 first most important impacts are included.

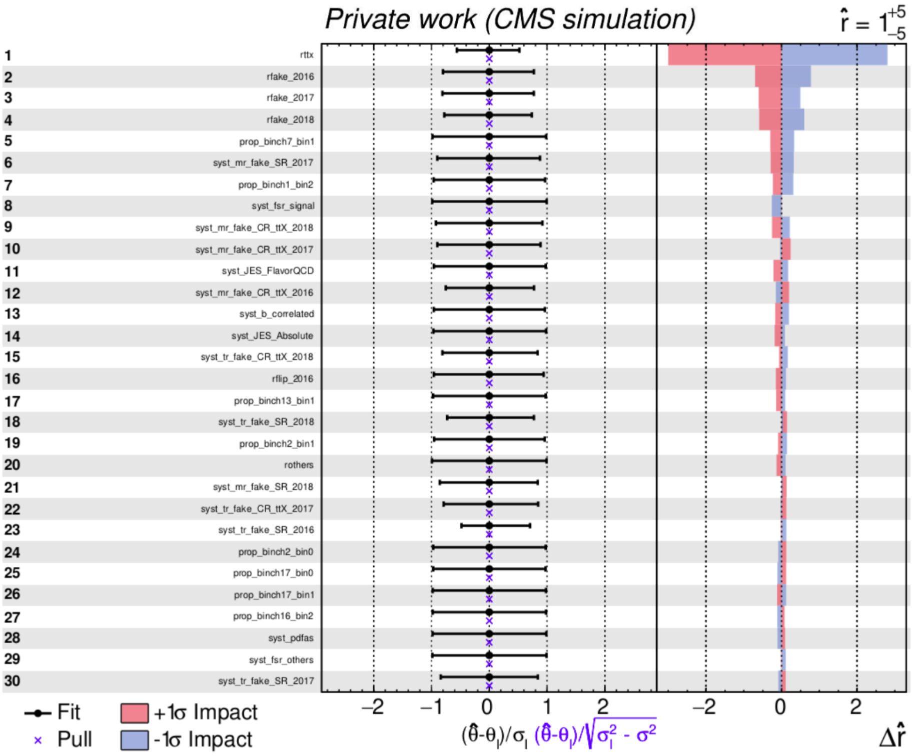


Figure C.20: Impact of the nuisance parameters for the Run 2 period with all the channels and regions combined, computed using Run 2 simulation with a signal injection of $r = 1$ and $m_{T'} = 650$ GeV. The 30 first most important impacts are included.

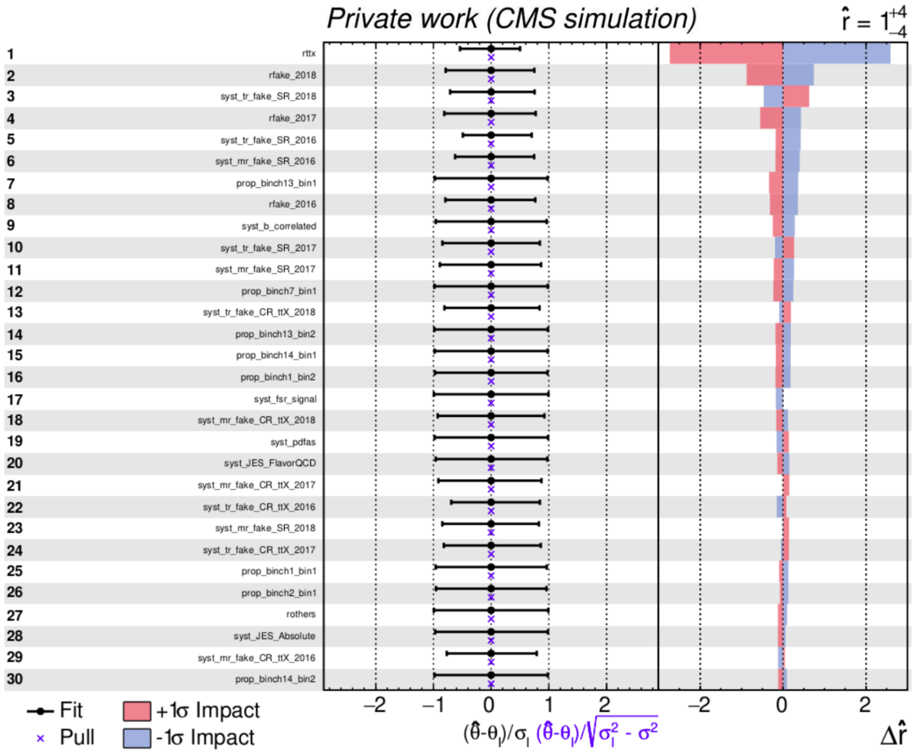


Figure C.21: Impact of the nuisance parameters for the Run 2 period with all the channels and regions combined, computed using Run 2 simulation with a signal injection of $r = 1$ and $m_{T'} = 675$ GeV. The 30 first most important impacts are included.

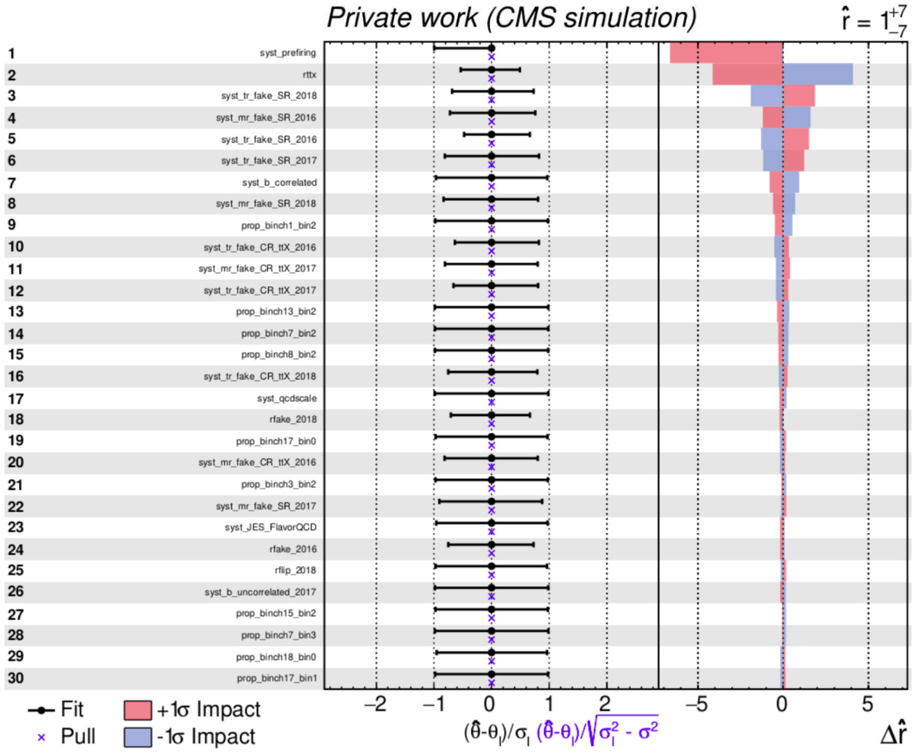


Figure C.22: Impact of the nuisance parameters for the Run 2 period with all the channels and regions combined, computed using Run 2 simulation with a signal injection of $r = 1$ and $m_{T'} = 800$ GeV. The 30 first most important impacts are included.

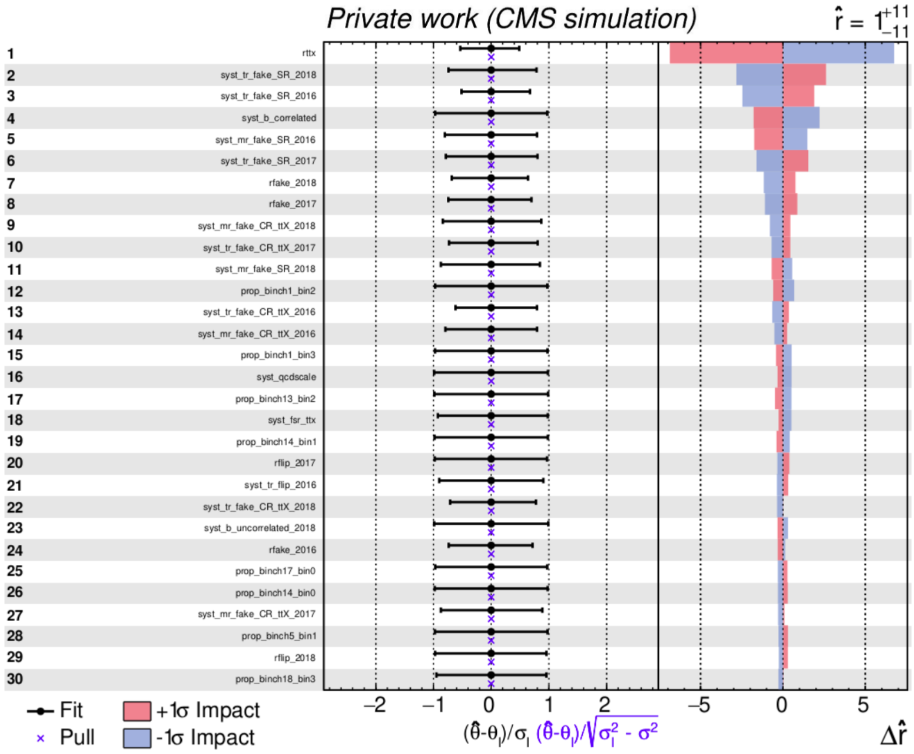


Figure C.23: Impact of the nuisance parameters for the Run 2 period with all the channels and regions combined, computed using Run 2 simulation with a signal injection of $r = 1$ and $m_{T'} = 900$ GeV. The 30 first most important impacts are included.

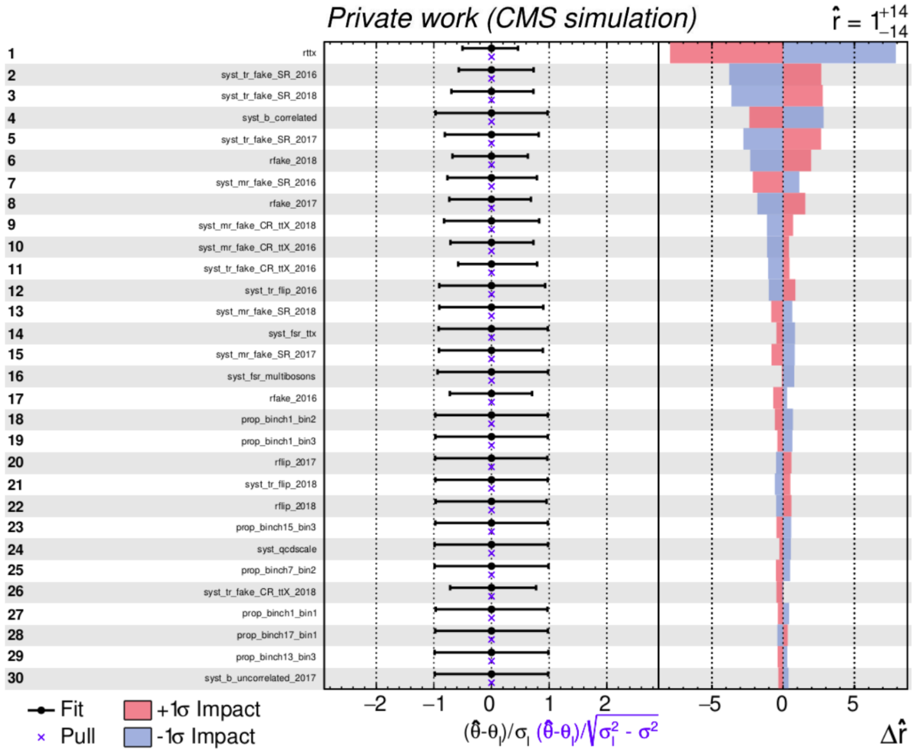


Figure C.24: Impact of the nuisance parameters for the Run 2 period with all the channels and regions combined, computed using Run 2 simulation with a signal injection of $r = 1$ and $m_{T'} = 1000$ GeV. The 30 first most important impacts are included.

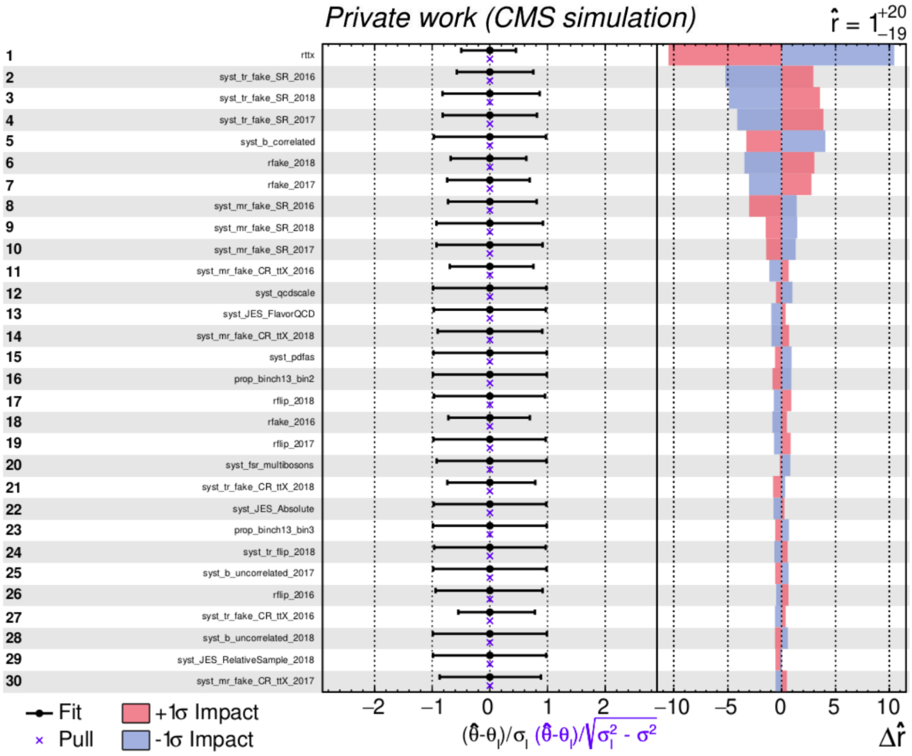


Figure C.25: Impact of the nuisance parameters for the Run 2 period with all the channels and regions combined, computed using Run 2 simulation with a signal injection of $r = 1$ and $m_{T'} = 1100$ GeV. The 30 first most important impacts are included.

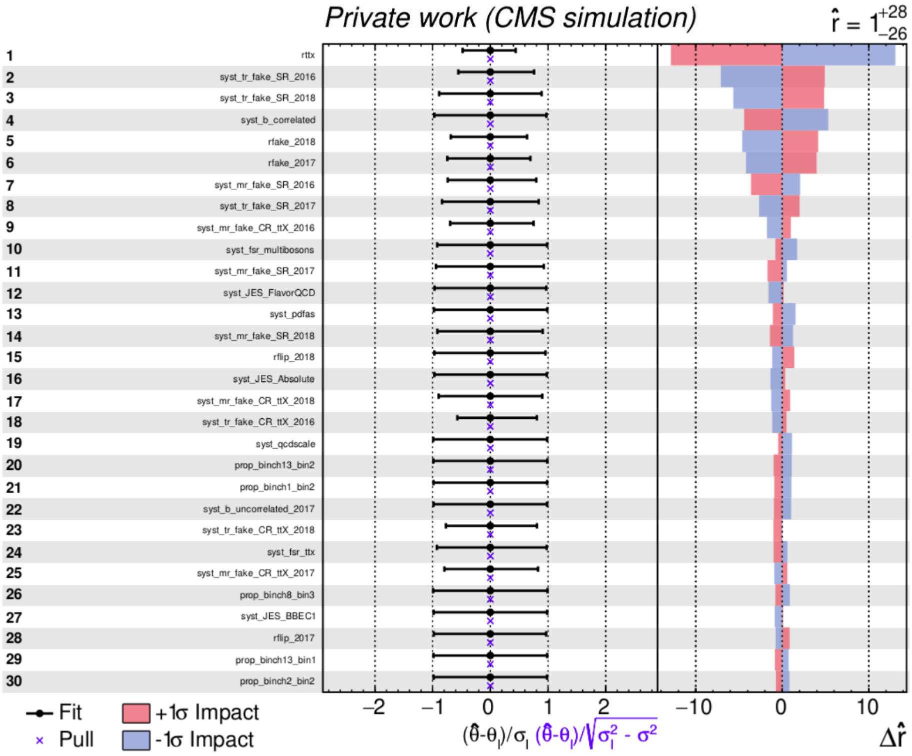


Figure C.26: Impact of the nuisance parameters for the Run 2 period with all the channels and regions combined, computed using Run 2 simulation with a signal injection of $r = 1$ and $m_{T'} = 1200$ GeV. The 30 first most important impacts are included.

$m_{T'}$ (GeV)	Signal strength
600	$r = 1.0^{+7.4}_{-7.3} = 1.0^{+5.7}_{-5.8}$ (Syst) $^{+4.8}_{-4.4}$ (Stat)
625	$r = 1.0^{+6.5}_{-6.2} = 1.0^{+5.0}_{-5.0}$ (Syst) $^{+4.0}_{-3.7}$ (Stat)
650	$r = 1.0^{+7.7}_{-7.6} = 1.0^{+5.9}_{-6.1}$ (Syst) $^{+4.9}_{-4.5}$ (Stat)
675	$r = 1.0^{+9.7}_{-9.6} = 1.0^{+7.7}_{-7.9}$ (Syst) $^{+6.0}_{-5.4}$ (Stat)
700	$r = 1.0^{+8.6}_{-8.1} = 1.0^{+6.6}_{-6.5}$ (Syst) $^{+5.5}_{-4.9}$ (Stat)
800	$r = 1.0^{+11.7}_{-10.8} = 1.0^{+8.9}_{-8.6}$ (Syst) $^{+7.6}_{-6.7}$ (Stat)
900	$r = 1.0^{+17.6}_{-17.6} = 1.0^{+12.9}_{-14.0}$ (Syst) $^{+11.9}_{-10.6}$ (Stat)
1000	$r = 1.0^{+22.7}_{-22.6} = 1.0^{+16.3}_{-17.8}$ (Syst) $^{+15.8}_{-14.0}$ (Stat)
1100	$r = 1.0^{+37.3}_{-37.2} = 1.0^{+25.4}_{-29.1}$ (Syst) $^{+27.3}_{-23.3}$ (Stat)
1200	$r = 1.0^{+48.8}_{-50.5} = 1.0^{+27.5}_{-37.8}$ (Syst) $^{+40.2}_{-33.6}$ (Stat)

Table C.1: Signal strength obtained with the likelihood fit per mass point value of the T', computed using 2016 simulation.

$m_{T'}$ (GeV)	Signal strength
600	$r = 1.0^{+7.6}_{-7..7} = 1.0^{+6.0}_{-6.5}$ (Syst) $^{+4.6}_{-4.2}$ (Stat)
625	$r = 1.0^{+5.7}_{-5.4} = 1.0^{+4.1}_{-4.0}$ (Syst) $^{+3.9}_{-3.5}$ (Stat)
650	$r = 1.0^{+7.0}_{-6.6} = 1.0^{+5.1}_{-5.0}$ (Syst) $^{+4.8}_{-4.3}$ (Stat)
675	$r = 1.0^{+7.5}_{-7.1} = 1.0^{+5.4}_{-5.4}$ (Syst) $^{+5.1}_{-4.6}$ (Stat)
700	$r = 1.0^{+8.3}_{-8.1} = 1.0^{+6.4}_{-6.5}$ (Syst) $^{+5.2}_{-4.8}$ (Stat)
800	$r = 1.0^{+11.0}_{-10.5} = 1.0^{+8.4}_{-8.3}$ (Syst) $^{+7.1}_{-6.4}$ (Stat)
900	$r = 1.0^{+22.3}_{-21.8} = 1.0^{+17.8}_{-18.0}$ (Syst) $^{+13.5}_{-12.3}$ (Stat)
1000	$r = 1.0^{+25.0}_{-23.0} = 1.0^{+18.1}_{-17.4}$ (Syst) $^{+17.2}_{-15.1}$ (Stat)
1100	$r = 1.0^{+35.0}_{-32.0} = 1.0^{+24.4}_{-23.5}$ (Syst) $^{+25.0}_{-21.8}$ (Stat)
1200	$r = 1.0^{+46.2}_{-39.1} = 1.0^{+27.4}_{-24.3}$ (Syst) $^{+37.1}_{-30.6}$ (Stat)

Table C.2: Signal strength obtained with the likelihood fit per mass point value of the T', computed using 2017 simulation.

$m_{T'}$ (GeV)	Signal strength
600	$r = 1.0^{+5.6}_{-5.5} = 1.0^{+4.5}_{-4.6}$ (Syst) $^{+3.2}_{-3.0}$ (Stat)
625	$r = 1.0^{+5.7}_{-5.6} = 1.0^{+4.5}_{-4.5}$ (Syst) $^{+3.5}_{-3.2}$ (Stat)
650	$r = 1.0^{+7.3}_{-7.4} = 1.0^{+6.0}_{-6.3}$ (Syst) $^{+4.2}_{-3.9}$ (Stat)
675	$r = 1.0^{+6.0}_{-5.8} = 1.0^{+4.7}_{-4.7}$ (Syst) $^{+3.8}_{-3.5}$ (Stat)
700	$r = 1.0^{+7.0}_{-6.9} = 1.0^{+5.6}_{-5.7}$ (Syst) $^{+4.2}_{-3.9}$ (Stat)
800	$r = 1.0^{+10.9}_{-10.7} = 1.0^{+9.0}_{-9.1}$ (Syst) $^{+6.0}_{-5.5}$ (Stat)
900	$r = 1.0^{+15.6}_{-14.9} = 1.0^{+12.5}_{-12.2}$ (Syst) $^{+9.3}_{-8.5}$ (Stat)
1000	$r = 1.0^{+23.1}_{-21.8} = 1.0^{+18.2}_{-17.6}$ (Syst) $^{+14.2}_{-12.9}$ (Stat)
1100	$r = 1.0^{+30.0}_{-27.3} = 1.0^{+22.1}_{-20.6}$ (Syst) $^{+20.4}_{-17.9}$ (Stat)
1200	$r = 1.0^{+43.7}_{-39.1} = 1.0^{+30.5}_{-28.0}$ (Syst) $^{+31.3}_{-27.2}$ (Stat)

Table C.3: Signal strength obtained with the likelihood fit per mass point value of the T', computed using 2018 simulation.

$m_{T'}$ (GeV)	Signal strength
600	$r = 1.0^{+4.1}_{-4.1} = 1.0^{+3.5}_{-3.5}$ (Syst) $^{+2.3}_{-2.2}$ (Stat)
625	$r = 1.0^{+3.6}_{-3.5} = 1.0^{+2.9}_{-2.8}$ (Syst) $^{+2.2}_{-2.0}$ (Stat)
650	$r = 1.0^{+4.6}_{-4.5} = 1.0^{+3.8}_{-3.8}$ (Syst) $^{+2.6}_{-2.5}$ (Stat)
675	$r = 1.0^{+4.5}_{-4.4} = 1.0^{+3.6}_{-3.6}$ (Syst) $^{+2.7}_{-2.5}$ (Stat)
700	$r = 1.0^{+4.8}_{-4.7} = 1.0^{+4.0}_{-4.0}$ (Syst) $^{+2.8}_{-2.6}$ (Stat)
800	$r = 1.0^{+6.9}_{-6.7} = 1.0^{+5.8}_{-5.6}$ (Syst) $^{+3.8}_{-3.6}$ (Stat)
900	$r = 1.0^{+11.3}_{-10.9} = 1.0^{+9.3}_{-9.2}$ (Syst) $^{+6.3}_{-6.0}$ (Stat)
1000	$r = 1.0^{+14.5}_{-13.9} = 1.0^{+11.5}_{-11.2}$ (Syst) $^{+8.8}_{-8.3}$ (Stat)
1100	$r = 1.0^{+20.4}_{-19.4} = 1.0^{+15.5}_{-15.0}$ (Syst) $^{+13.3}_{-12.3}$ (Stat)
1200	$r = 1.0^{+28.3}_{-26.1} = 1.0^{+20.1}_{-18.8}$ (Syst) $^{+19.9}_{-18.1}$ (Stat)

Table C.4: Signal strength obtained with the likelihood fit per mass point value of the T', computed using Run 2 simulation.