ECOLE POLYTECHNIQUE DE BRUXELLES

BRUXELLES BRUSSEL

ULB

UNIVERSITÉ LIBRE DE BRUXELLES

Searching for New Physics in Dilepton Events
A Study of the Photon-Induced Process

The CMS Experiment - CERN

Mémoire présenté en vue de l'obtention du diplôme
de Master Ingénieur Civil Physicien

Maxence Draguet

Promoteur
Professeur Jean-Marc Sparenberg

Co-Promotrice
Professeur Barbara Clerbaux

Mentor
Laurent Thomas

Service
Inter-University Institute For High Energies

Année académique
2018 - 2019

Abstract

This Master thesis presents an analysis of the data collected during the full Run II (2016, 2017 and 2018) by the CMS detector, located at the proton-proton collider, the LHC, at CERN. Events with high energy dileptons are studied, with a focus on photon-induced (PI) lepton pair production. Protons, at high energy, are formed by valence quarks (u and d) and an intricate sea of quarks, gluons, and photons, described by the parton distributions functions. In proton-proton collisions at the LHC, two of these photons can interact to form a pair of leptons: the PI process. However, other processes such as the Drell-Yan (DY), the $t\bar{t}$ and the diboson WW also contribute to high energy dilepton final states and dominate the PI. Exploiting differences at the event level, the PI is selected from these backgrounds by demanding a low missing transverse energy significance, no b-jets, at most a single jet, and a low number of charged hadrons in the vertex fit. These conditions emanate from the absence of neutrinos and quarks in the final state as well as the relative stability of the protons in a PI interaction. The selected signal estimated using PI simulation is fitted to the data to deliver mass-dependent corrections to the theoretical cross section. Two procedures are implemented. The first one uses the difference in topology between the PI and the DY in the $\cos\theta$ distributions, due to the forward-backward asymmetry of the latter. The second one relies on the number of charged hadrons from the vertex fit. The corrections are found to be in good agreement with recently published results, pointing out an overestimation originating from finite size effects in the proton being neglected. A BSM search is then introduced by displaying the invariant mass spectrum in the PI enriched region. No significant deviations are observed within statistical incertitude, though a deeper study is needed to properly account for systematics.

Keywords:

Photon-induced | Drell-Yan | Dilepton | Standard Model Measurement | CMS Exotic physics

Résumé

Ce mémoire propose une analyse des données collectées lors du Run II (de 2016 à 2018) de l'expérience CMS, localisée au collisionneur de protons du CERN, le LHC. Les évènements dileptons de haute énergie sont étudiés et plus spécifiquement le processus de création de paire leptonique induit par des photons contenus dans les protons. En effet, à haute énergie, le proton est formé de ses quarks de valence (u et d) et une mer de quarks, gluons et photons. Cette composition est décrite de manière statistique par les fonctions de distribution de partons. Lors d'une collision proton-proton du LHC, deux de ces photons peuvent interagir et mener à la formation d'une paire leptonique : il s'agit du processus photon-induit (PI). Ce signal est cependant fortement dominé par d'autres processus pouvant conduire à un état final de dilepton, tels que le Drell-Yan (DY), le $t\bar{t}$ et le diboson WW. La sélection du PI peut être obtenue en tirant profit de différences au niveau de l'évènement. En effet, pour sélectionner le signal, il suffit de demander une faible signifiance d'énergie transverse manquante, pas de b-jets, au maximum un jet et un faible nombre de hadrons chargés liés au fit du vertex. Ces conditions sont inspirées par l'absence de neutrinos et de quarks dans l'état final du canal étudié ainsi que la relative stabilité des protons lors d'une interaction PI. Ayant isolé une zone enrichie en signal, un fit des simulations du PI aux données amène à des poids de correction sur la section efficace théorique à différentes masses. Deux procédures sont appliquées pour ce faire. La première tire profit de la différence de topologie entre les PI et DY des distributions en $\cos\theta$, due à l'asymétrie avant-arrière caractéristique du DY. La seconde s'intéresse au nombre de hadrons chargés. Ces corrections sont en bon accord avec des résultats récents issus de la littérature montrant une surestimation liée à une mauvaise inclusion des effets de taille finie du proton. Le travail se conclut avec une recherche de physique au-delà du Modèle Standard explorant le spectre de masse invariante dans la région enrichie en PI. Aucune déviation significative n'est observée comptant l'incertitude statistique et la systématique est laissée à une étude plus complète.

Mots-Clefs :

**Photon-induit | Drell-Yan | dilepton | mesure Modèle Standard | CMS
Physique exotique**

Acknowledgements

I would like to start by deeply thanking Professor Barbara Clerbaux for her continuous help in the last two years. For sending me to CERN on a summer internship, supporting the multitude of applications I have started and that will lead me to Oxford and, finally, for her precious remarks and guidance in the elaboration of the present work. My opening to the field of Particle Physics truly was paved by her valuable support.

Next in line, Laurent Thomas to whom I am deeply grateful for his mentoring in this project. From generating (awesome) simulation files to clearing up and ordering the overwhelming flux of information opening the doors of CMS unleashes on the unprepared student. I wish him the best in his bright new familial adventures.

My thanks to Professor Sparenberg for his help in making this project accessible to a physics engineer, overshadowing the work and his support in my many candidatures. His introductions to the field of quantum mechanics have left a vibrant interest that grew into a passion.

Many thanks also to Cécile Caillol who supervised my internship at CERN and introduced me to the joys of C++ and the wonders of ROOT as well as the scheme to carry out experimental Particle Physics analysis. This truly was a dream come true.

Psychologically, I would like to thank my roommates, Boris, Florian, Marin, and Francesca, who had to support the various stages of madness coming with the elaboration of this thesis. Who would have believed that a simple cup of coffee in good company could prove the last bastion of sanity?

Finally, I wish to express my complete gratitude to my family for carrying me all along the way, exposing me to so many wonderful culture and supporting every dreams.

Contents

Introduction

The current landscape in the field of Particle Physics is a deeply puzzling one. The Standard Model (SM) has, in the last decades, proven remarkably successful in describing the various constituents of matter as well as the strong, the weak and electromagnetic interactions. It however utterly fails to address some other aspects of the Universe, such as dark matter and dark energy. This provides a fertile ground for a plethora of ideas to grow, attempting to enlarge our understanding by deeper, beyond the Standard Model theories ranging from Supersymmetry to Grand Unified Theory.

A common denominator of these various extensions is that they increase the fundamental set of pieces: the elementary particles. Experiments, such as the Compact Muon Solenoid at CERN, are actively searching for traces of these illusive components. Having a thorough understanding of SM processes is thus of paramount importance to distinguish them. This Master thesis centres around that objective. It consists in an analysis of the CMS Run II (2016-2018) data to specifically uncover the contribution of the photon-induced (PI) process to the production of highly energetic lepton pairs. This final state is particularly appreciated by experimentalists as leptons are easy to identify, measure, reconstruct and also suffer from less backgrounds, an interesting set of characteristics demanded for precision measurements.

In the first chapter, entitled "The Standard Model of Particle Physics and Beyond", both the current theoretical context as well as the need and the method to uncover exotic particles are introduced. The channel chosen in this study, the PI signal, is thoroughly presented in a section dedicated to its Standard Model cross section computation at leading order and that of its irreducible background, the Drell-Yan. A specificity in the dynamics of the final state in processes involving a spin-1 mediator with vectorial and axial couplings, called "forward-backward asymmetry", is proven to be a decisive distinction between the two channels.

Some characteristics of the LHC machine and the CMS detector impacting the study are then presented in the chapter "The CMS Experiment at CERN". They are complemented by a short description of the reconstruction process in the detector as well as the different parts of the apparatus, with a focus on information relevant to the analysis.

The chapter "Models and Simulations" moves on to the procedure for carrying out the physical analysis at the generated level (the ideal theoretical world in which the process occurs) and the reconstructed level (the practical level corresponding to the data observed which is unfortunately impacted by the state of the detector). At the LHC, the high energy protons collided are composed of valence quarks (u and d) and an intricate sea of quarks, gluons, and photons. The formalism of parton distribution functions, a statistical representation of the proton content as a distribution of particle species with a certain likelihood of possessing a given fraction of the proton momentum, is briefly explained. Generated level simulations are then presented to search for relative differences in kinematics between the DY and PI as well as a practical study of forward-backward asymmetry to verify the discussion of the first chapter. The behaviour of both cross sections with invariant mass under interesting cuts is then explored. The chapter concludes with the reconstructed level simulation and the procedure to derive normalisation scale factors to match the different simulations to the data.

The core of the analysis begins in the chapter "Selecting the PI". The objective, as indicated by the title, is to derive a set of cuts to isolate the PI signal from other backgrounds. Different options are explored by going back to the main differences between the various processes at the level of the events and the impacts of the cuts proposed are studied. It concludes with a discussion on the quality of the simulations, particularly concerning the number of charged hadrons matched to the vertex with an analysis of some subtle effects, such as the importance of pile-up, misidentification of vertex traces and colour interactions between proton remnants.

The last chapter, "Study of the PI", consists in a statistical analysis fitting the selected signal, estimated using PI simulation, to the data to deliver mass-dependent corrections to the theoretical cross section. Two procedures are implemented. The first one uses the difference in topology between the PI and the DY in the $\cos\theta$ distributions, due to the forward-backward asymmetry of the latter. The second one relies on the number of charged hadrons from the vertex fit. Finally, the resulting mass spectrum in the PI enriched region is briefly explored to search for a potential BSM signal.

This Master thesis sits only at the beginning of a larger scope effort to uncover exotic phenomena. It focuses more on the SM observations of interacting photon partons of the proton at hadron colliders, a subject still in nascent phase. After some conclusions on the work itself, the last part offers different outlooks and prospects.

CHAPTER 1

The Standard Model of Particle Physics and Beyond

The field of Particle Physics lead, without the shred of a doubt, to some of the most important achievements of modern physics. It embodies a marriage of the different profound theories to have shaken our very understanding of the world we are a part of. Joining relativity into a quantum description of field dynamics proven indeed extremely fruitful to tackle the immense variety of physical elements at their most fundamental level and the way they interact with each other.

This chapter opens with an overview of the Standard Model. A high-level tour of this gauge theory is proposed with its fundamental blocks unveiled, though no quantitative elements are introduced. It is followed by a discussion on the main limitations of the model as well as some of the ideas actively explored for extending its realm. Finally, a discussion on new heavy gauge bosons is presented with a focus on the particular dilepton final state. Some important theoretical results regarding expected SM processes in proton-proton collisions are then introduced to conclude the chapter: the Drell-Yan and the photon-induced processes.

1.1 Overview of the Standard Model

The marriage of special relativity, quantum mechanics and field theory into Quantum Field Theory (QFT) lead to the construction of a set of theoretical pieces that can be joined to form what is commonly called the Standard Model (SM). It consists in a quantum field gauge theory based on the symmetry group $SU(3) \times SU(2) \times U(1)$ narrating the story of interacting fields. Physical interactions are mediated by bosons, integer spin particles, and three forces are successfully described by the SM, whereas the last one, gravity, is not included. They are presented in order of decreasing intensity in table 1.1.

Matter itself is structured by fermions, half-integer spin particles, that can be regrouped, as presented in figure 1.1, into two different types: quarks and leptons. Each of these is formed of three generations, with only the first one being stable (they indeed are the lightest of their respective type). Leptons consist, in order of generation, of electrons, muons, and taus, all having, in units of the electron charge, charge -1, as well as their uncharged neutrino counterparts. Quarks gather the up/down, charm/strange

Force	Range	Relative Strength	Mediator Bosons
Strong	10^{-15} m	1	8 gluons (g)
Electromagnetic	∞	10^{-3}	Photon (γ)
Weak	10^{-18} m	10^{-14}	$W^{\pm}$, Z
Gravity	∞	10^{-43}	*Hypothetical*: graviton (G)

Table 1.1 – Fundamental forces in Nature and their SM mediators for the first three, the last one being hypothetical.

and top/bottom couples with respective charges $\frac{2}{3}$/-$\frac{1}{3}$, where these fractional values constitute one of their peculiarity. They are also the only ones to be directly sensitive to the strong force. However, due to a mechanism described by the appropriate quantum field theory of this interaction, baptised quantum chromodynamics (QCD), they group to form hadrons where the strong charge, called "colour" in this context, cancels out. Hadrons are either mesons if they consist of a quark and an antiquark or baryons if they are composed of three quarks (some more exotic compounds may have been observed such as tetraquarks, an assembly of four quarks, and so on).

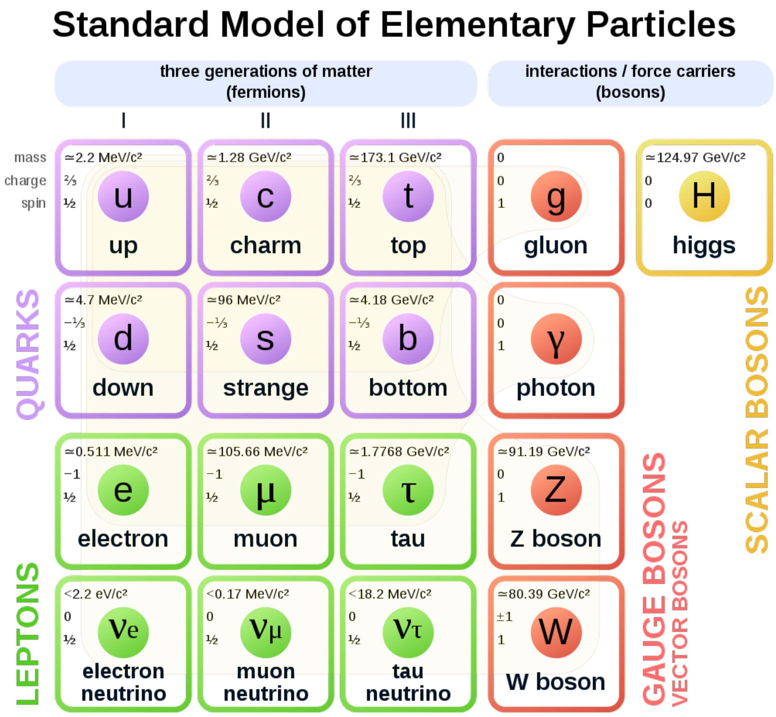


Figure 1.1 – Particles in the SM.

Explicit mass terms in the SM formulation are forbidden by the invariance under $SU(3)_C \times SU(2)_L \times U(1)_Y$ (C for colour, L for chirality and Y for hypercharge) that constrains the theory [1, 2, 3]. Indeed, any fermionic mass term would assume the form $m\bar{\psi}\psi$ and thus couple left- and right-handed fields thereby breaking the gauge symmetry of the SM under which left-handed fermions transform as doublets and right-handed fermions as singlets. The solution to this problem requires a complex breaking of symmetry and the introduction of new scalar field leading to an extra-particle: the Brout-Englert-Higgs scalar boson (H) emerging from its eponym mechanism [4, 5].

1.2 The limitations of the Standard Model

Over the last forty years, predictions from the Standard Model have had tremendous successes in describing fundamental aspects of the world with great precision. The existence of the W and Z bosons, the gluons, the top and the charm quarks was confirmed and in 2012 the last missing piece of the SM, the Brout-Englert-Higgs boson, was discovered. Nevertheless, it is commonly accepted that the SM can only constitute a low energy approximation of a deeper and larger theory, extending beyond the Standard Model (BSM) physics. It indeed utterly fails to address some important aspects of the Universe. This disappointing realisation brought about the complex current situation where a fundamental model is capable of an astonishing precision while completely ignoring a vast landscape of phenomena. Among the main missing parts of the SM are:

- Dark matter and dark energy, which existence is suggested by cosmological observations, are not accounted for by the SM.
- The Standard Model does not offer any explanation regarding the matter-antimatter asymmetry of the Universe.
- Neutrinos are not fully described by the SM which does not offer any explanation regarding their actual and very low masses as indicated by "neutrino oscillation", a phenomenon allowing a neutrino of a certain flavour to spontaneously acquire another one. The absence of right-handed neutrinos and left-handed antineutrinos is also a puzzling specificity.
- The hierarchy problem. This refers to the need for an extraordinary fine tuning required for loop corrections to cancel each other out so that the mass of the H boson may match observations. There also seems to be no explanation regarding the large difference in scale between the electroweak and gravitational forces, the latter operating at a scale 10^{17} times larger than the former.
- The Standard Model does not predict the mass of fermions. They are in fact given as parameters to the theory. They, together with other variables in the model, number nineteen free constants with seemingly arbitrary values.
- Gravity. This fundamental force is indeed quite different from the other three, mainly due to its interconnection with the geometry of space-time as indicated by General Relativity. Hence, it cannot so simply be added to the SM as an interaction. An other important deviation is its relatively much weaker strength at the electroweak scale (region of energies $\approx$ 100 GeV) - it indeed only becomes somewhat significant at what is called the Plank scale, for energies in the order of 10^{19} GeV. The last issue with this interaction is the impossibility to renormalise a theory of quantum gravity due to its hypothetical mediator, the graviton, being a spin-2 particle leading to diverging loop corrections containing these bosons that cannot be absorbed.

1.3 New physics: Beyond the Standard Model

The various difficulties of the Standard Model suggest the existence of a deeper theory that would englobe the different observations into a larger framework. A wild landscape of miscellaneous extensions are currently being investigated with not a single experimental result favouring any direction. Most notable among these are:

- Supersymmetry (SUSY) [6]. The primary idea is to have a new symmetry of the action connecting each bosonic (fermionic) field to a "superpartner" fermionic (bosonic) field into a "supermultiplet". This process leads to a family of models with the simplest one, entitled the "Minimal Supersymmetric Standard Model" (MSSM), doubling the number of fundamental particles by allocating a superpartner to each SM particle (called sfermion for fermion, gaugino for gauge bosons, and so on). It offers a natural dark matter candidate: the lightest neutral SUSY particle. One of the main challenge of some supersymmetric models is the lost of the proton stability, a prediction that has not currently been supported by any experiment.

- Grand Unified Theory (GUT), proposing an extension of the SM gauge group. Again, this family of models lies behind the idea of extending the symmetry group ruling the action. The SM would in this scenario only constitute a low energy residuum of a more fundamental group that broke down. Most notable, a $SU(5)$ extension was proposed in 1974 [7] but also assert the instability of the proton and does not accommodate for massive neutrinos. Building upon this attempt, a $SO(10)$ model was suggested as well as an E_6 one, supported by string-theory models [8]. Some members of this family also provide an attempt at unifying the three coupling constants into a single entity at high energy. With the breaking of the fundamental group symmetry, extra $U(1)$ symmetries should appear and introduce extra neutral gauge bosons called Z' [9, 10]. Any such boson with a mass in the TeV range could be detected at CERN if it couples to quarks and leptons.

- Extra Dimensions. This family of models postulates the existence of compact extra spatial dimensions to explain the apparent weakness of gravity by allowing the graviton to propagate along this new space, effectively diluting the effect of its interaction with the common SM components that are constrained to the familiar 4-dimension space-time. These models therefore have the advantage of bringing the Planck scale down, potentially as low as a few TeV.

A current objective of the Particle Physics community is to search for any trace of exotic or anomalous physics with respect to the SM that could help indicate, or at least restrain, the families of models to keep under consideration. In this respect, the direct observation of a new particle predicted by some deeper theory would be a decisive discovery. In hadron colliders, such as the Large Hadron Collider (LHC) at CERN, this new physics could manifest itself through many different scenarios ranging from dark matter and dark sector, SUSY superpartners, heavy exotic resonances, exotic long-lived signatures, vector-like quarks (new quarks having a vectorial coupling), and so on [11, 12].

1.4 Searches in the dilepton invariant mass spectrum

A particularly privileged scenario at the LHC is to select events with a lepton-antilepton pair of the same generation in the final state and to search the dilepton invariant mass spectrum for deviations from the predictions of the SM. This quest is motivated both theoretically and experimentally. Theoretically since many models predict the existence of exotic channels with a pair of highly energetic leptons in the final state. Experimentally because lepton identification, reconstruction and measurement are globally more efficient to perform than their quark/hadron counterparts and events with such configuration are less sensitive to background. It thus offers up a clean channel to search for new resonances or deviations from the expected SM processes due to the hypothetical presence of exotic heavy neutral gauge bosons

Z'. Many searches are being performed, with no significant deviations from the SM expectations observed yet and limits on production of new resonances cross sections placed [13, 14, 15, 16, 17].

However, various Standard Model processes can also lead to such a final state. The dominating one is the so called Drell-Yan (DY) and is explored in section 1.5. Other important processes leading to a final state with leptons are the WW (leading for example to $WW \to l\nu l'\bar{\nu}$, where l stands for leptons and ν for neutrinos), the WZ (with $WZ \to l\nu l'l'$), the ZZ (to four leptons or two leptons and two neutrinos), the $t\bar{t}$, and the tW [18, 19, 20]. Note that these channels do not exactly lead to the same final configuration but may appear to do so if some particles are missed or misidentified.

Finally, another SM process can lead to a dilepton pair: the photon-induced lepton pair production (PI). Leptons production occurs from interacting photons from the colliding protons. The process is described in section 1.6 and constitutes an irreducible contributions to the dilepton mass spectrum, as its final state is precisely the same. Understanding this channel is quite challenging as the photon content of the proton is difficult to access. In most BSM studies, the process is included as an uncertainty on the better understood DY [21], though in recent years some direct searches for the PI signal were performed [22] and the measured cross section has been observed to be consistent with theoretical predictions [23]. Among the kinematical variables, the acoplanarity (defined in section 4.2.3), the invariant mass, and the transverse momentum have also been verified to loosely match theoretical predictions [24]. However, the incertitude of these studies, either due to the limited data set available (not the entire Run II) or the lower centre of mass energy of $\sqrt{s} = 7$ TeV, leaves some room for improvement. Errors are typically dominated by statistical limitations that can be tackled by collecting more events in the data. Another important source of limitations is the description of the proton and its photon content which has seen a significant improvement in recent years, as explained is section 3.3. The present analysis therefore aims to study the PI contribution in the region of large dilepton invariant mass, searching for new resonances. To reach this, PI selection criteria are developed and the full Run II dataset at a centre of mass of $\sqrt{s} = 13$ TeV is analysed.

1.5 The Drell-Yan process

One of the most intense process in high energy proton collisions offering a lepton-antilepton pair as a final state is the so called Drell-Yan process, suggested in 1970 by Sidney Drell and Tung-Mow Yan [25, 26]. It constitutes the leading contribution to lepton production in high dilepton mass (above 20 GeV). Schematically represented by its leading order Feynman diagrams in figure 1.2, it consists in the annihilation of a quark and an antiquark belonging to two different protons. The intermediary state can involve the SM weak neutral gauge boson Z or an off-shell photon γ^*, both ultimately decaying to form the dilepton state.

This process is theoretically well understood and the following section aims to derive its cross section, following the steps described in [15], as well as some specificities of the kinematics of the leptons that will prove crucial in the next parts of this work. What follows is a simplified list of the conventions employed:

- 4-vectors are written as p^μ (p_μ) for the contravariant (covariant) case.
- 3-vectors are given an arrow $\vec{p}$.

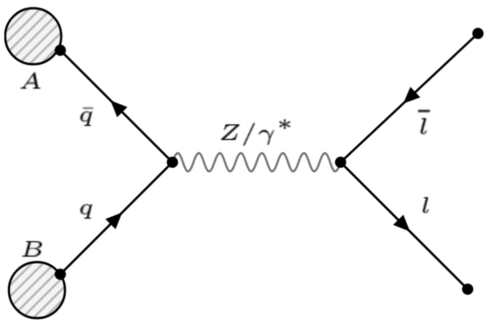


Figure 1.2 – The Drell-Yan process.

- The Minkowsky space-time metric $\eta^{\mu\nu}$ is set with signature sign(η) = (1, -1, -1, -1).
- Indices repeated in contravariant and covariant ways are summed over. Greek letters indicate 4-dimensional indices, 0 being time and 1, 2, 3 space. Latin letters stand for the 3-dimensional space indices.
- The natural unit system is adopted, where $\hbar = c = \epsilon_0 = 1$.

The amplitude of both diagrams from figure 1.2 is proportional to the strength of the dominating interaction: the fine-structure constant $\alpha \approx 1/137$. One thus expects the cross section to be proportional to the square of the coupling constant, as the diagrams have each two vertices. Given the intense energies studied, an assumption will be employed: that of ultra-relativistic limit emanating from the extremely high centre of mass energy. This allows to ignore the mass of fermions compared to their momentum in the computation: $E^2 = m^2 + p^2$ and $E \sim p$. The mass of the Z boson however nears a non-negligible 91 GeV, hence the approximation does not apply in its case.

1.5.1 Two-body scattering cross section

The differential cross section of a two-body scattering (denoted by q and $\bar{q}$) is given by [2]:

$$d\sigma = \frac{1}{(2E_q)(2E_{\bar{q}})|\vec{v_q} - \vec{v_{\bar{q}}}|} |\mathcal{M}|^2 \, d\Pi_{\text{LIPS}}, \tag{1.1}$$

with Π_{LIPS} the Lorentz-invariant phase space (LIPS)

$$d\Pi_{\text{LIPS}} \equiv (2\pi)^4 \delta \left(\sum_{\text{All}} p \right) \times \prod_{\text{final states j}} \frac{d^3 p_j}{(2\pi)^3 2E_{p_j}}, \tag{1.2}$$

where by convention entering 4-momenta are positive and exiting ones are negative. Hence the $\delta \left(\sum_{\text{All}} p \right)$ term imposes energy and 3-momentum conservations. The $|\mathcal{M}|^2$ contains the matrix element, to be computed using the Feynman approach. The frame chosen for computation is the centre of mass frame of the quark-antiquark pair ($q\bar{q}$). The z-axis is chosen so that it points in the same direction as the quark, the y-axis is perpendicular to the plane containing the quark and the lepton directions, and the x-axis points in the direction of the lepton while making the whole system dextrogyral.

The angle θ is defined to lie between the quark and lepton directions, as shown in figure 1.3. It will prove important in the definition of the "forward-backward asymmetry" later in this thesis. Working in

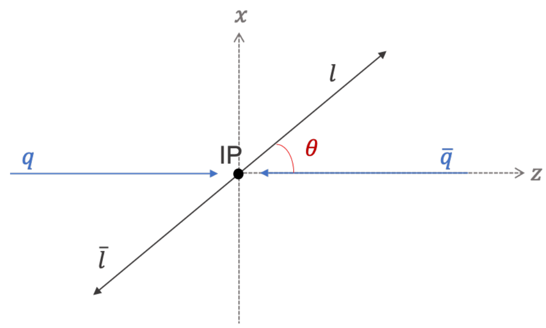


Figure 1.3 – The $q\bar{q}$ centre of mass frame to describe the Drell-Yan process.

this particular frame simplifies considerably equation 1.1, as the masses of the fermionic particles are here negligible (and their anti-counterparts possess the same irrelevant mass). One therefore reduces the problem to the following set of 4-momenta:

$$p_q = \left(\frac{E_{CM}}{2}, 0, 0, \frac{E_{CM}}{2}\right), \qquad p_l = \left(\frac{E_{CM}}{2}, \frac{E_{CM}}{2}\sin\theta, 0, \frac{E_{CM}}{2}\cos\theta\right), \tag{1.3}$$

$$p_{\bar{q}} = \left(\frac{E_{CM}}{2}, 0, 0, -\frac{E_{CM}}{2}\right), \qquad p_{\bar{l}} = \left(\frac{E_{CM}}{2}, -\frac{E_{CM}}{2}\sin\theta, 0, -\frac{E_{CM}}{2}\cos\theta\right), \tag{1.4}$$

where in particular $\vec{p_q}$ = - $\vec{p_{\bar{q}}}$ and $\vec{p_l}$ = - $\vec{p_{\bar{l}}}$, given the centre of mass frame of total energy E_{CM}. Taking profit of the δ-function in the LIPS term, one can integrate over one of the momentum without considering the matrix element. A convenient change of variable can then be imposed to reach a simplified form of equation 1.1, valid in the case of particle-antiparticle decay/creation in the centre of mass frame [2]:

$$\left(\frac{d\sigma}{d\Omega}\right)_{CM} = \frac{1}{64\pi^2 E_{CM}^2}|\mathcal{M}|^2. \tag{1.5}$$

Only the modulus squared of the matrix element is left to compute. The two fermions in both the initial and final states have unknown polarisations. Any initial and final spin configurations are valid and should be averaged over (2 choices for each quark). Only colourless pairs of quark can give a Z or γ so they should be required to carry opposite colours (there are three possible choices). Thus, averaging over the $2 \times 2 \times 3 = 12$ different valid configurations leads to:

$$|\mathcal{M}|^2 = \frac{1}{12}\sum_{s_q}\sum_{s_{\bar{q}}}\sum_{s_l}\sum_{s_{\bar{l}}} |\mathcal{M}_{s_q,s_{\bar{q}},s_l,s_{\bar{l}}}(q+\bar{q} \to l+\bar{l})|^2. \tag{1.6}$$

1.5.2 The Drell-Yan cross section

Computing this last result requires a careful setting of the problem and a methodical approach to consider all terms. Appendix A goes through each step of the derivation until the final cross section. In the following, Q_l is the lepton charge, Q_q the quark one, and $g_{V_{l/q}}$ and $g_{A_{l/q}}$ are the Z boson vectorial and axial couplings with either a lepton l or a quark q. One obtains successively the cross section for a photon propagator:

$$\left(\frac{d\sigma_\gamma}{d\Omega}\right)_{CM} = \frac{Q_l^2 Q_q^2 e^4}{24(4\pi)^2 E_{CM}^2}\left[(1+\cos\theta)^2 + (1-\cos\theta)^2\right]. \tag{1.7}$$

That for a Z-boson propagator:

$$\left(\frac{d\sigma_Z}{d\Omega}\right)_{CM} = \frac{Q_l^2 Q_q^2 e^4}{24(4\pi)^2 E_{CM}^2}|\mathcal{R}|^2 \Big\{[(g_{V_l}^2 + g_{A_l}^2)(g_{V_q}^2 + g_{A_q}^2) + 4g_{V_l}g_{A_l}g_{V_q}g_{A_q}](1+\cos\theta)^2 + [(g_{V_l}^2 + g_{A_l}^2)(g_{V_q}^2 + g_{A_q}^2) - 4g_{V_l}g_{A_l}g_{V_q}g_{A_q}](1-\cos\theta)^2\Big\}. \tag{1.8}$$

And finally one for the interference between both diagrams:

$$\left(\frac{d\sigma_{\text{Int}}}{d\Omega}\right)_{CM} = \frac{Q_l^2 Q_q^2 e^4}{24(4\pi)^2 E_{CM}^2}\text{Re}\{\mathcal{R}\}\left\{2[g_{V_l}g_{V_q} + g_{A_l}g_{A_q}](1+\cos\theta)^2 + 2[g_{V_l}g_{V_q} - g_{A_l}g_{A_q}](1-\cos\theta)^2\right\}. \tag{1.9}$$

Note in particular how a flip of $\theta \to \theta + \pi$ leaves equation 1.7 invariant but modifies in an observable fashion equations 1.8 and 1.9, a peculiarity referred as "forward-backward asymmetry" [27]. This only happens if none of the couplings are null for the first one and if g_{A_l} and g_{A_q} are non-null for the second one. Clearly, the asymmetry finds its origin in the mixture of axial and vectorial couplings of the neutral gauge boson Z.

The total cross section is simply the sum of equation 1.7, 1.8 and 1.9:

$$\left(\frac{d\sigma_{\text{DY}}}{d\Omega}\right)_{CM} = \frac{Q_l^2 Q_q^2 e^4}{24(4\pi)^2 E_{CM}^2}\left[c_1\,(1+\cos^2\theta) + c_2\,\cos\theta\right], \tag{1.10}$$

where

$$c_1 = 1 + 2\,\text{Re}\{\mathcal{R}\}g_{V_l}g_{V_q} + |\mathcal{R}|^2\left(g_{V_l}^2 + g_{A_l}^2\right)\left(g_{V_q}^2 + g_{A_q}^2\right), \tag{1.11}$$

$$c_2 = 4\,\text{Re}\{\mathcal{R}\}g_{A_l}g_{A_q} + 8\,|\mathcal{R}|^2 g_{V_l}g_{A_l}g_{V_q}g_{A_q}. \tag{1.12}$$

Note also the $\left(\frac{e^2}{4\pi}\right)^2 = \alpha^2$ dependence, as expected from the coupling discussion. The total cross section is simply the integral over the solid angle Ω of equation 1.10:

$$\sigma_{\text{DY}} = \int_\Omega \frac{d\sigma_{\text{DY}}}{d\Omega}d\Omega = \frac{4\pi}{9}\left(\frac{e^2}{4\pi}\right)^2\frac{Q_q^2 Q_l^2}{E_{CM}^2}\,c_1 \tag{1.13}$$

where Q_l^2 is equal to 1, given the nature of leptons.

The presence of a $\cos\theta$ in equation 1.10 term clearly marks the forward-backward asymmetry if c_2 is non-null. This asymmetry can be expressed as the variable:

$$A_{\text{FB}} = \frac{\sigma_{\text{F}} - \sigma_{\text{B}}}{\sigma_{\text{F}} + \sigma_{\text{B}}} \tag{1.14}$$

where σ_{F} and σ_{B} are the forward ($\theta \in [0, \pi/2]$) and backward ($\theta \in [\pi/2, \pi]$) cross sections, obtained by integrating the differential one on the solid angle Ω contained over these regions. One easily uncovers:

$$A_{\text{FB}} = \frac{3}{8}\frac{c_2}{c_1} \tag{1.15}$$

The presence of the asymmetry will prove to be a very useful characteristic to differentiate the Drell-Yan process from the photon-induced one. Indeed, next section shows this asymmetry is inexistent for the latter case.

1.6 The photon-induced lepton pair production process

The two tree level diagrams are represented in figure 1.4. A few differences with respect to the DY will arise in this computation, primarily due to the presence of a fermionic propagator and the sums over initial polarisation states of the photons, as they are now external particles.

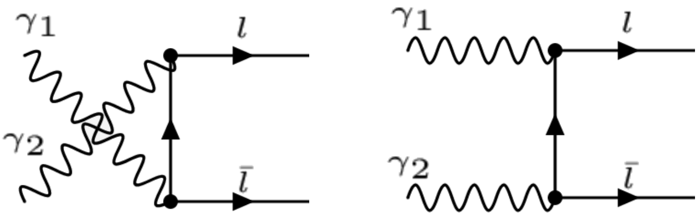


Figure 1.4 – The two tree level diagrams for the photon-induced lepton pair production.

One could go through each step of the calculation, starting from the Feynman diagrams to express the matrix elements and squaring them. There is however a much simpler approach to compute the result based on Compton scattering, shown in figure 1.5.

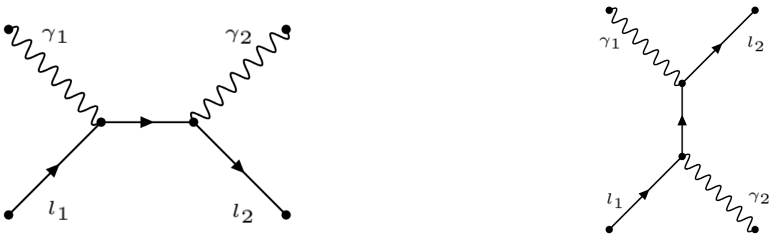


Figure 1.5 – The two tree level diagrams for Compton scattering.

Indeed, by crossing symmetry, the amplitude can be obtained from the Compton one by replacing Compton variables according to:

$$p_{\gamma_1} \to p_{\gamma_1}, \quad p_{\gamma_2} \to -p_{\gamma_2}, \quad p_{l_1} \to -p_{\bar{l}}, \quad p_{l_2} \to p_l \tag{1.16}$$

Knowing the value of the modulus squared matrix element for Compton scattering [3]:

$$\frac{1}{4}\sum_{\text{pol.}i,s_l,s_{\bar{l}}} |\mathcal{M}_{\text{Compton}}|^2 = 2e^4 \left[\frac{p_{l_1}\cdot p_{\gamma_2}}{p_{l_1}\cdot p_{\gamma_1}} + \frac{p_{l_1}\cdot p_{\gamma_1}}{p_{l_1}\cdot p_{\gamma_2}} + 2m^2\left(\frac{1}{p_{l_1}\cdot p_{\gamma_1}} - \frac{1}{p_{l_1}\cdot p_{\gamma_2}}\right) + m^4\left(\frac{1}{p_{l_1}\cdot p_{\gamma_1}} - \frac{1}{p_{l_1}\cdot p_{\gamma_2}}\right)^2\right], \tag{1.17}$$

one gets the corresponding amplitude for the photon-induced lepton pair creation by applying the change of variables prescribed and an overall sign-flip (due to the crossing symmetry):

$$\frac{1}{4}\sum_{\text{pol.}i,s_l,s_{\bar{l}}}|\mathcal{M}_{\text{PI}}|^2 = -2e^4\left[\frac{(-p_{\bar{l}}).(-p_{\gamma 2})}{(-p_{\bar{l}}).p_{\gamma 1}} + \frac{(-p_{\bar{l}}).p_{\gamma 1}}{(-p_{\bar{l}}).(-p_{\gamma 2})} + 2m^2\left(\frac{1}{(-p_{\bar{l}}).p_{\gamma 1}} - \frac{1}{(-p_{\bar{l}}).(-p_{\gamma 2})}\right)\right.$$
$$\left. + m^4\left(\frac{1}{(-p_{\bar{l}}).p_{\gamma 1}} - \frac{1}{(-p_{\bar{l}}).(-p_{\gamma 2})}\right)^2\right]. \tag{1.18}$$

And thus obtain:

$$\frac{1}{4}\sum_{\text{pol.}i,s_l,s_{\bar{l}}}|\mathcal{M}_{\text{PI}}|^2 = 2e^4\left[\frac{p_{\bar{l}}.p_{\gamma 2}}{p_{\bar{l}}.p_{\gamma 1}} + \frac{p_{\bar{l}}.p_{\gamma 1}}{p_{\bar{l}}.p_{\gamma 2}} + 2m^2\left(\frac{1}{p_{\bar{l}}.p_{\gamma 1}} + \frac{1}{p_{\bar{l}}.p_{\gamma 2}}\right) - m^4\left(\frac{1}{p_{\bar{l}}.p_{\gamma 1}} + \frac{1}{p_{\bar{l}}.p_{\gamma 2}}\right)^2\right], \tag{1.19}$$

a result equivalent to that mentioned in [28].

Now, choosing an ultra-relativistic approach by switching the lepton mass m off to 0 and a set of 4-momenta:

$$p_{\gamma_1} = \left(\frac{E_{CM}}{2}, 0, 0, \frac{E_{CM}}{2}\right), \qquad p_l = \left(\frac{E_{CM}}{2}, \frac{E_{CM}}{2}\sin\theta, 0, \frac{E_{CM}}{2}\cos\theta\right), \tag{1.20}$$
$$p_{\gamma_2} = \left(\frac{E_{CM}}{2}, 0, 0, -\frac{E_{CM}}{2}\right), \qquad p_{\bar{l}} = \left(\frac{E_{CM}}{2}, -\frac{E_{CM}}{2}\sin\theta, 0, -\frac{E_{CM}}{2}\cos\theta\right), \tag{1.21}$$

equation 1.19 can be re-expressed into:

$$\frac{1}{4}\sum_{\text{pol.}i,s_l,s_{\bar{l}}}|\mathcal{M}_{\text{PI}}|^2 = 2e^4\left[\frac{E_{CM}^2(1-\cos\theta)}{E_{CM}^2(1+\cos\theta)} + \frac{E_{CM}^2(1+\cos\theta)}{E_{CM}^2(1-\cos\theta)}\right] = 4e^4\frac{1+\cos^2\theta}{1-\cos^2\theta} \tag{1.22}$$

This concludes the matrix element computation and the photon-induced lepton pair creation differential cross section can now be obtained placing the expression 1.22 in equation 1.5:

$$\left(\frac{d\sigma_{\text{PI}}}{d\Omega}\right)_{CM} = \frac{e^4}{(4\pi)^2E_{CM}^2}\frac{1+\cos^2\theta}{1-\cos^2\theta}. \tag{1.23}$$

Observe the absence of any forward-backward asymmetry, a quite logical result given the symmetry of the initial state (two photons). This expression is also proportional to the α^2 factor.

Summary of the Chapter

After a short overview of the Standard Model and the modern landscape of Particle Physics, the main limitations of the current theory were presented and some actively researched extensions of the SM introduced. This BSM physics should lead to new exotic particles that could be observed at the LHC. A favourable final state for these searches consist in lepton pairs as these events are easy to identify and isolate. Two important backgrounds, the Drell-Yan and photon-induced lepton pair production, were further studied and their cross sections derived. The analysis indicates a significant difference in topology residing in the forward-backward asymmetry, a peculiarity brought about by the axial and vectorial couplings of the neutral gauge boson Z that tends to orientate the negatively charged lepton in the same direction as the quark at the source of the scattering affair.

CHAPTER 2

The CMS Experiment at CERN

This section aims to offer the necessary insight into the operation of the Compact Muon Solenoid (CMS) experiment at CERN as well as some basis on the accelerator complex required to generate the physical events and data necessary for the research programme. Accelerators are a remarkable tool in the context of Particle Physics. They offer up projectiles at specific energies allowing for the study of fundamental processes to take place in a low noise environment. Section 2.1 is dedicated to the Large Hadron Collider, section 2.2 describes the CMS experiment, and section 2.3 unveils the event reconstruction procedure at CMS.

2.1 The LHC infrastructure

In this thesis, only proton-proton interactions are considered which is why the following short description of the accelerator complex is focus on the chain of machines operated for such particles. Figure 2.1 nonetheless displays a more complete overview of the CERN infrastructure. The LHC is indeed also capable of providing lead beams to produce proton-proton, lead-lead and lead-proton collisions.

For protons, the story starts at their source: extraction from a ionised gas of hydrogen atoms. They are then subject to several consecutive stages of acceleration: a linear accelerator, LINAC2, to accelerate them to 50 MeV, the Proton Synchrotron Booster (PSB) to get them up to 1.4 GeV, the Proton Synchrotron (PS) to reach an energy of up to 25 GeV followed by the Super Proton Synchrotron (SPS), bringing the particles to a staggering 450 GeV. Only then do they possess the right energy to be injected into the LHC itself by two beampipes making them circulate in opposite directions along the 26.7 km long tunnel, about 100 m underground (the largest accelerator ever built). Thanks to the last stage of acceleration, provided by sixteen radiofrequency cavities in the LHC itself, the machine has been, since 2015, capable of reaching a maximum energy of 6.5 TeV per particle, with beams circulating for several hours inside the pipes until their purity is deemed insufficient and are then dumped. The protons themselves are organised into, nominally, $n_b = 2808$ bunches of $N_p \sim 10^{11}$ particles [29]

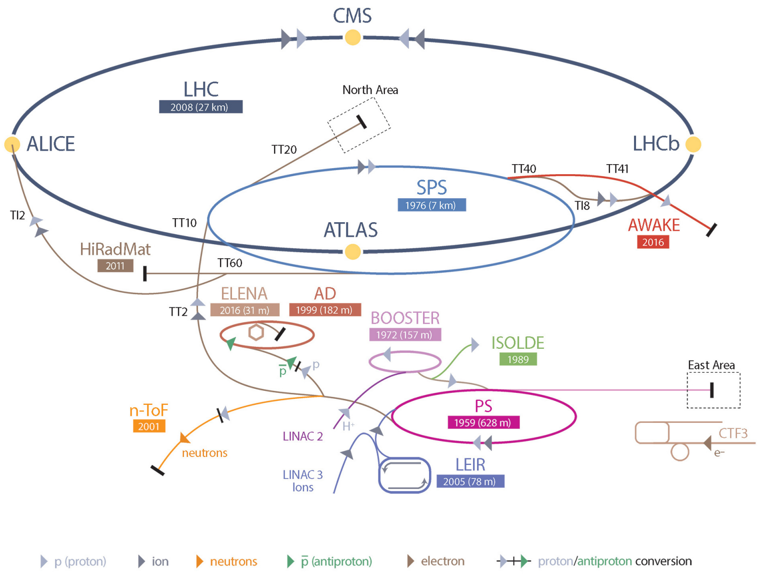


Figure 2.1 – The accelerator complex at CERN.

colliding every 25 nanoseconds with those of the other pipe in four locations where detectors - CMS, ATLAS, ALICE and LHCb - are installed.

A simple and important equation when designing an accelerator complex capable of reaching a certain energy is the relation between momentum p, curvature of the trajectory ρ and a constant and parallel magnetic field intensity B imposed by the machine [30]:

$$p = qB\rho \tag{2.1}$$

where q is the charge of the species of particle exploited. This relationship limits the maximal momentum (thus kinetic energy) reachable for a certain species, knowing the maximal magnetic field produced by available systems (currently, complex superconducting magnets cooled to 1.9 K by liquid He are capable of reaching 8T) and the curvature attainable (limited by the geometry of the tunnel where the machine is installed - most of them are operated in underground facilities to limit background noises and guarantee sufficient stability).

A parameter of most importance when operating a collider is the instantaneous luminosity $\mathcal{L}$: the proportionality factor between the event rate (the number of event of type i per unit of time) and the cross section for the same process σ_i:

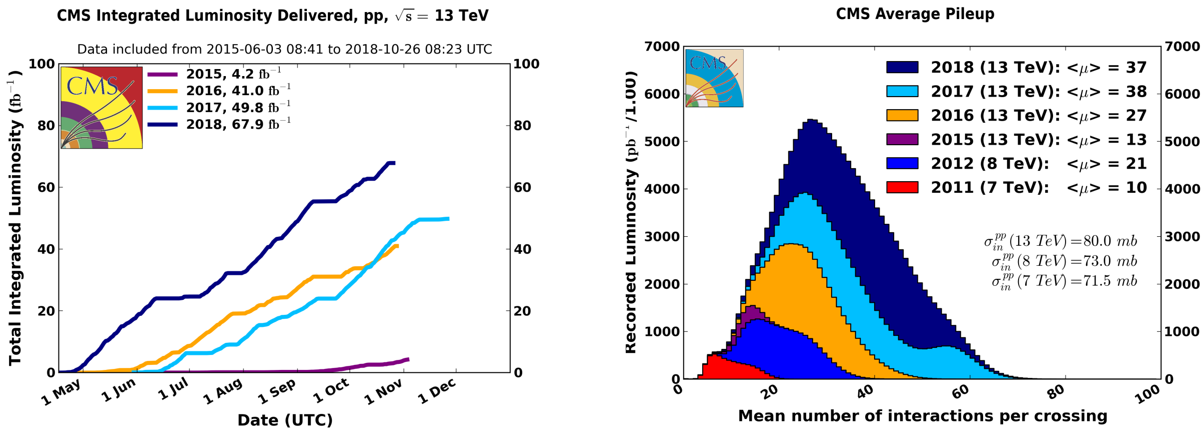
$$\frac{dN_i}{dt} = \mathcal{L}\,\sigma_i. \tag{2.2}$$

In the context of the LHC where the two beams contain roughly the same number of protons per bunches regularly spaced, the instantaneous luminosity can be expressed as [29]:

$$\mathcal{L} = N_p^2 n_b f \frac{\gamma}{4\pi\epsilon_n\beta^*} F, \tag{2.3}$$

where f is the revolution frequency of a bunch (design value of 11,245 Hz), γ is the relativistic Lorentz factor, ϵ_n the normalised transverse beam emittance (design value 3.75 μm), β^* the betatron function at the interaction point (IP), and F a geometric factor reducing the luminosity because of a small crossing angle between the two beams at the IP (note that this angle can be neglected in most of the physical discussion of events). The nominal value of the instantaneous luminosity is 10^{34} cm^{-2} s^{-1}.

The luminosity is commonly integrated over a given period of time to form the integrated luminosity, the dominant parameter for an analysis covering several lapses of data taking phases. Its value is key when attempting to measure the cross section based on the selection of events in the data, as affirmed by equation 2.2. A good control on this parameter is in fact so dire that most experiments perform online measurements. The CMS collaboration monitors the luminosity in data taking phases at a rate of 1 Hz with a statistical accuracy of about 1% and a systematic accuracy of 5% by exploiting signals from the forward hadron calorimeter (HF) [31]. The results of this online reading are compared to offline measurements in the data cleaning phase. As indicated in figure 2.2a, the luminosity of the machine is being pushed higher with each new data taking phase (a barn, b, corresponds to 10^{-28} m^2).



(a) *CMS integrated luminosity over Run II.* (b) *CMS average PU from 2011 to 2018.*

Figure 2.2 – CMS luminosity and pile-up. Extracted from the CMS public information page: https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults.

Another crucial parameter originating from the design of the machine is the pile-up (PU). When two bunches reach the IP, it is very likely to have several proton-proton interactions occurring. These events *pile up* in the data recorded by the detector since their residues are simultaneously present in the apparatus. This undesirable effect is unfortunately unavoidable when attempting to run the accelerator at high instantaneous luminosity. Complex algorithms in the reconstruction phase of the events detected are required to isolate the piled-up events. Note in figure 2.2b how the increase in luminosity during the different years of Run II is matched by an increase in average PU.

In the next years, as shown in figure 2.3, the LHC and its associated experiments as well as the chain feeding the machine are to undergo an upgrade phase scheduled to last from 2019 to the start of 2021, called long shutdown 2 (LS2). Following this, the LHC is expected to deliver a total luminosity of 300 fb^{-1} by the end of 2023.

By the middle of the next decade, the LHC will undergo a major upgrade named the High-Luminosity LHC (HL-LHC) with an objective target of 3,000 fb^{-1} [12]. This will be matched by an estimated average PU of 200, a challenging situation for reconstruction to be tackled by a dramatic improvement

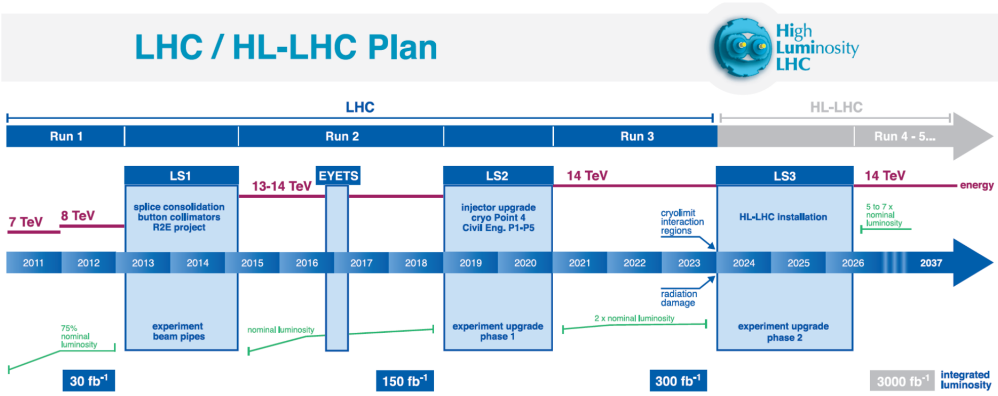


Figure 2.3 – Schedule of the LHC operation [32].

in forward acceptance and in tracking hardware (reaching higher granularity) and software as well a better calorimeters. The reward will be an increased sensitivity to BSM physics.

2.2 The CMS experiment

This thesis exploits data collected by the CMS detector, displayed in figure 2.4. It consists in a multi-purpose apparatus with muons playing an important role, as their identification and momentum measurement are at the heart of the design of this 14,000-ton detector [31]. Being cylindrical, it can be separated in two parts: a succession of concentric cylinders forming the "barrel" and two "endcaps" closing the geometry with consecutive pierced disks to leave room for the vacuum chamber. Due to limits in available space, the detector is remarkably compact with a solenoidal superconducting magnet forming the skeleton of this 28.7 m long machine with a 15 m diameter.

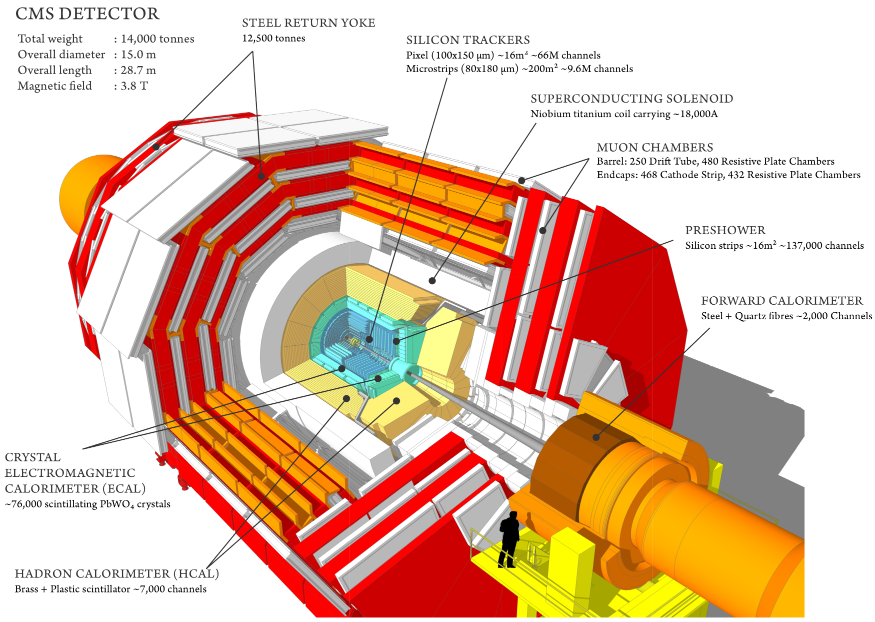


Figure 2.4 – A schematic view of the CMS detector and its various components.

At the centre of the apparatus lies the nominal interaction point. It is surrounded by an inner tracking system measuring the trajectory of charged particles. Calorimeters are the next devices, conveying a measurement of the particles energy by absorbing (most of) it either in the electromagnetic calorimeter (ECAL, for electrons, positrons and photons) or the hadronic one (HCAL, for hadrons). The next parts are the powerful magnetic solenoid and finally the muon detection system. Some typical particle trajectories are schematised in figure 2.5.

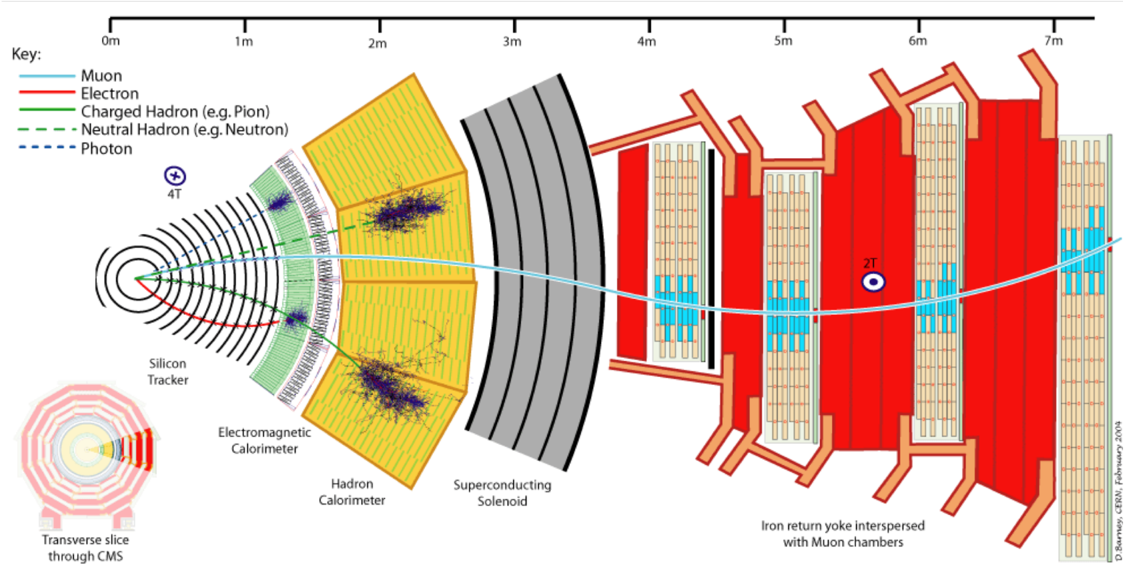


Figure 2.5 – A transverse slice of the CMS experiment with a schematic view of the path of various particles.

The coordinate system of the CMS experiment places, at the IP, the z-axis in the direction of the beam, the x-axis radially towards the centre of the LHC, and the y-axis so that it is dextrogyral. The polar angle θ lies between the direction of the proton beam, z, and the direction of the residual particle considered in the laboratory frame and the azimuthal angle ϕ is in the $x-y$ plane. The polar angle can be re-expressed as the pseudo-rapidity η, which is defined as:

$$\eta = -\ln\left(\tan\frac{\theta}{2}\right), \tag{2.4}$$

and takes its value between $(-\infty, \infty)$, corresponding respectively to a θ of π and 0. Its name comes from the fact it constitutes a good approximation of the rapidity y for particles in the ultra-relativistic limit ($E \gg m$):

$$\eta = \frac{1}{2}\ln\left(\frac{|\vec{p}| + p_L}{|\vec{p}| - p_L}\right), \tag{2.5}$$

$$y = \frac{1}{2}\ln\left(\frac{E + p_z}{E - p_z}\right), \tag{2.6}$$

since the mass can be ignored. The physical interest of the rapidity resides in the invariance of Δy under boosts along the z-direction.

2.2.1 The tracking system

The inner tracking system surrounding the IP measures the trajectory of charged particles thanks to the ionisation they impose on the medium. This device also reconstructs secondary vertices, an important input in cases where an interaction point displaced with respect to the nominal one are deemed of interest. For example, in B-physics one aims to study the b quark that typically decays some hundreds μm away from the nominal IP due to its 1.3 10^{-12} s lifetime. It is made of high granularity silicon detectors shaped in pixel ($100 \times 150\ \mu m^2$) for the innermost part and in microstrips ($80 \times 180\ \mu m^2$) for the outer part. They are grouped to cover different directions over a region extending up to $|\eta| = 2.5$ [31]. Between ten and fifteen hits from a single particle, depending on the pseudo-rapidity, should occur in the whole tracker system and a trajectory can be fitted to the associated pattern.

The tracker is tasked with a seemingly paradoxical mission: record the successive positions of a particle without disturbing its dynamics nor its energy, since this is only measured afterwards in the calorimeters. For this to happen, in the tracker the radiation length X_0, the characteristic distance of a material over which a particle loses an average of $1 - \frac{1}{e}$ of its energy through electromagnetic interactions, should be as large as possible. This also limits the quantity of material to be used. Photons are the most affected by this transit with up to half of them converting to e^+e^- pairs. Other undesired effects may include multiple scattering, bremsstrahlung and nuclear interactions. To ensure a satisfying resolution, the alignment of the system requires careful considerations.

2.2.2 The calorimeters

The ECAL measures the energy of electrons (positrons) and photons. A particular phenomenon takes place when a particle of these species enter the active volume: they generate electromagnetic showers. The global effect is an ionisation in the lead tungstate crystals ($PbWO_4$ with a short radiation length of 0.89 cm) that is converted into a scintillation light proportional to the energy of the incident particle. Collecting this light, photodetectors converts its intensity into a proportional electronic signal to be digitalised by an ADC.

The HCAL essentially accomplishes the same task for hadrons. Its structure is made of alternating layers of absorbers (stainless steel at the extremities and brass for the rest) and plastic scintillators. The calorimeter is decomposed into a part inside the magnet coils, the hadron barrel calorimeter covering $|\eta| < 1.3$, and one outside, the outer calorimeter to increase the thickness to twelve interaction lengths λ_I, the distance for a hadron to lose on average 63 % of its energy through strong interactions. Finally, the $1.3 < |\eta| < 3.0$ region is covered by the endcap calorimeter and the $3.0 < |\eta| < 5.2$ one by forward calorimeters made of Cherenkov light detectors with radiation-hard quartz fibers, given the high radiation flux.

2.2.3 The solenoid

The entire fly through the inner detectors occurs in the presence of an intense 3.8 T magnetic field parallel to the beam axis and thus bending the trajectory of charged particles. The field is generated by the superconducting magnet, a complex structure of niobium titanium coils capable of carrying a nominal current of 19,500 A and cooled down by liquid helium. Curving the trajectory offers a way of deducing the momentum of the charged particle from the curvature of its path according to the same relation governing the capabilities of an accelerator, equation 2.1.

2.2.4 The muon system

The inner detectors are encapsulated into a thick muon detection system, interleaving muon chambers and steel return yoke. This is a crucial part as it allows to identify muons and measure their momenta.

Muons behave like heavy electrons, about 200 times more massive. They mostly interact with matter by ionisation which lets them cover large distances, their lifetime being 2.2 μs. Hence their associated detection system covers the outermost part of the machine, with a significant radius of material. The technology consists in a mixture of gases plunged into an electric field that accelerates the ions to form avalanches. Due to this particular configuration, muons are reliably identified in the detector.

2.2.5 The triggering system

The bunch spacing in the machine delivers a crossing frequency of 40 MHz. This is far too quick for the detector to be able to save and transfer data related to every event (which takes about 1 MB) before even taking the PU into account. There is a distinct need to select events that matter to let only those cross the bottleneck of data transfer. This is the task of the triggering system, endowed with a hierarchical structure of filters and flags. Each level operates at an appropriate rate for the frequency of events reaching its entry. A rate reduction of about a 10^6 factor occurs in two levels. The Level-1 (L1) Trigger, with an output frequency of a 100 kHz, uses Field-Programmable Gate Array (FPGA) and bases its decision on rough information from the calorimeters and the muon system. The High Level Trigger (HLT), combining complex information from all parts of the detector, exploits software capabilities processed by a CPU farm to reduce the rate to approximately 400 Hz before storage.

2.3 Reconstructing events at CMS

The setup unveiled in the previous section records the data according to the trigger procedure. The events now need to be reconstructed by identifying the wild variety of particles in the final state. These include electrons, photons, neutral and charged hadrons, and muons. Neutrinos are only inferred by missing transverse energy as they do not interact with the detector.

Concerning the tracks of charged particles, the reconstruction uses hits in the pixel and strip trackers. A Kalman filter adapted into the Combinatorial Track Finder software fits tracks to the hits by proceeding iteratively. For the first iteration, very tight criteria are imposed to seed and reconstruct tracks, thus allowing a low fake rate but offering only a moderate efficiency. The criteria are then loosen, hence increasing the efficiency, with hits safely assigned at the previous step removed. An iteration involves: seeding a five parameter helix (due to the trajectory in an almost uniform magnetic field) to a limited set of hits, adding other hits to the track, fitting the track candidates to estimate the trajectory parameters and finally some criteria are demanded by quality cuts to remove fake tracks [17].

In this Master thesis, leptons play an especially important part as they constitute the final state under study. The reconstruction procedures unveiled here thus focus on electrons and muons (their respective anti-partners are treated the same way, only the charge flipping sign). They are reconstructed by a dedicated algorithm named Particle Flow (PF) by combining information from all subdetectors. Taus, being very massive and decaying quickly with a 2.9 10^{-13} s lifetime, would require a significantly different methodology as they are more difficult to identify and reconstruct. They produce neutrinos and have a leptonic decay in about a third of the cases. Hadronic channels form the rest and include jets which showers have very specific features, such as small track multiplicity (they are made up of a low number of charged particle, usually between one and three) and are more collimated and isolated compared to those generated by gluons and quarks. Overall, taus suffer from a lower reconstruction efficiency and higher backgrounds (they are easier to imitate from jets). They are thus to be left out of this analysis.

2.3.1 Electrons

The procedure to reconstruct electrons is in fact similar to that required for photons, as suggested by their common calorimeter. They indeed both generate showers of a relatively small width in the ECAL. Matching the location of the deposited energy with a trajectory from the tracking part usually allows to disentangle electrons from photons, the latter leaving no trace. However, as mentioned in section 2.2.1, photons can convert to a dielectron pair in the tracker. Electrons of course are also free to radiate a photon, the famous bremsstrahlung. The very nature of these processes leads to a certain distribution of energy in ϕ-direction around the main helix of the trajectory.

Conditions on the shower shape and isolation (requiring a track to be the sole inhabitant of a cone within a given opening), to mention only some of them, make up a selection process to distinguish the candidate electron from jets. Sequential cuts or multivariate analysis (MVA, a name commonly employed in Particle Physics to describe machine learning procedures [33]) are implicated in this decision.

As will be shown in the following chapters, there is a need to know the charge of the lepton with good precision. The sign of the curvature of the track in the magnetic field delivers that information, though with a high misidentification of 10 % for electrons at large $|\eta|$, in the endcaps. To tackle this, two other estimations of the charge are established, one using a rougher version of the track (before ultimate fitting) and the other one defining the charge sign as the sign of the difference, in the ϕ-direction, of the vectors joining the beam spot to the supercluster position (defined in the following paragraph) and to the first hit of the electron reconstructed track. Selecting the charge indicated by a majority voting between the three methods reduces the misidentification rate to 1.5 % [17] for electron produced in Z boson decays.

Momentum is derived from the measurement of the track parameters or from the supercluster parameters, for low and high energy candidates respectively. The supercluster is a union of ECAL cluster pieces, each containing a minimum amount of energy (for electrons, 1 GeV), that is located at the energy weighted mean of cluster positions.

2.3.2 Muons

Muons are identified with the tracks from the tracker and from those in the muon system. A global track is fitted by a Kalman filter acting on hits from both of these systems. This is very efficient if the muon has left several hits in different muon stations and has a low p_T, as the efficiency can reach 99 %. Muon momentum can only be derived from track information, since it is very unlikely for such a particle to leave a significant proportion of its energy within the calorimeters. This limits the measurement precision at high momentum (> 100 GeV) compared to electrons.

2.3.3 Jets

A jets is a beam of a large number of neutral and charged particles arising from the hadronisation of quarks or gluons. They are reconstructed by dedicated algorithms to identify their content using the different parts of the detector. Energy clusters not associated to any tracks are assigned to neutral hadrons and charged hadrons are matched a track and an energy deposit. A particular scenarios occurs when a jet originates from the hadronisation of a b-quark, as these can be distinguished from the other types due to the existence of a secondary vertex a few hundreds μm away from the primary one. In CMS, a discriminator based on a boosted decision tree (BDT) is therefore used to identify these b-jets. The input variables to this BDT contain information related to secondary vertices, the presence of leptons in the jet, and so on [34].

2.3.4 Missing transverse energy

Missing transverse energy (MET), the negative vectorial sum of all PF particles in the event, is typically important for particles weakly interacting with the detector, as their energy is not measured. MET therefore provides an indication on the presence of these particles, such as neutrinos. Momentum imbalance in the transverse plane leads to an estimation of this value, which requires a fine measurement of all the elements the detector is sensitive to. It can be inferred by different methods accounting for bias from pile-up.

2.3.5 Vertex reconstruction

The procedure to reconstruct the vertices of a given event starts with selecting tracks and grouping those that seem to emanate from the same points. These origins are then obtained by a fitting procedure returning the vertex positions [35].

2.3.6 Event selection in this analysis

This analysis exploits ee and $\mu\mu$ final states in the detector. In order to reject jets misidentified as leptons, several conditions are applied on the reconstructed candidates. These conditions are based on various discriminating variables related, for example, to the quality of the muon track or to the shower shape in the calorimeters for electrons. The charges of the leptons are also demanded to be well reconstructed and to have opposite values. Since one does not expect any activity around the leptons in the signal considered, some isolation requirements are also applied. The final lepton selection follows the one used in [36]. Each lepton is also demanded to pass an acceptance cuts corresponding to $p_T > 25$ GeV, per trigger reason, and $|\eta| < 2.5$, due to the detector geometry. The former requirement is loosen to authorise a trailing $p_T > 15$ GeV and a leading $p_T > 25$ GeV in the reconstructed level. Jets in the analysis are submitted to two conditions: $|\eta| < 2.4$ and $p_T > 40$ GeV.

Summary of the Chapter

In this chapter, the LHC operation at CERN and the CMS experiment have both been introduced with a focus on the minimal information required to carry out an analysis of the proton-proton collisions data collected during Run II, covering operations from 2016 to 2018. Some details on the accelerator complex, such as the nominal values of the machine and some important parameters, mainly the luminosity and the pile-up, have been defined. The CMS detector structure and the event reconstruction procedure have also been explored.

CHAPTER 3

Models and Simulations

This chapter introduces the final aspects of the procedure to analyse the physical information hidden in the CMS data. To uncover the photon-induced lepton pair creation (PI), the signal, one needs to accurately select the right same-flavour high energy lepton pairs. Indeed, other background processes will also contribute to these events. They can be separated into a reducible part, only mimicking the final state in part or passing by misidentification, and an irreducible part, having the same final state. Among the reducible backgrounds, there are, to mention a few, diboson productions (such as the WW*,* WZ*,* ZZ *with leptonic decay of the gauge boson) and* $t\bar{t}$ *following leptonic decay channels (leading to real leptons + other things). Among the irreducible one, the Drell-Yan (DY) dominates.*

A common methodology in Particle Physics is to simulate all processes of importance with respect to a certain final state. Monte Carlo (MC) simulations are generated to reproduce both the physical events (generated level) and the measurement process (reconstructed level). These matter greatly to correct for acceptance and efficiency effects and are explored in section 3.1.

In the context of proton-proton collision, another layer of simulation is required since the projectiles have the daunting characteristic of not being fundamental particles but an agglomerate of "partons", a complicated entanglement of quarks, gluons and photons. Sections 3.2 and 3.3 open up the proton content in terms of the parton distribution functions. Finally, sections 3.4.2 and 3.4.3 start the analysis by presenting the generated level simulations of the hard scattering processes and the files for the reconstructed level ones.

3.1 Simulating events

Most of the proton-proton interactions are elastic scattering and inelastic scattering with limited momentum transfer. Interesting physics occurs rather when large portions of the momenta are transferred in what is called a "hard" scattering. In these cases, the interaction is modelled in two steps: the hard interaction between any two partons that can be computed by perturbative calculations and the rest of the interaction. However, complexity increases when considering higher order diagrams, such as initial state radiation (ISR) or final state radiation (FSR) of gluons.

Monte Carlo simulations perform a variety of tasks. At the generated level, they produce a large number of individual random events in a fashion prescribed by theoretical concerns. They also simulate the interactions of the various final state particles with the detector to bring the generated events to the reconstructed level. The reconstructed simulations can then follow the same treatment as the data collected. A typical proton-proton interaction is simulated with the following steps [17]:

- A hard scattering process: partons of each proton collide. The exact momenta of these constituents should be returned by the parton distribution functions (section 3.2). The matrix element formalism can be employed at leading order (LO) or next-to-leading order (NLO) or even NNLO, ... depending on the generator employed.
- Around the hard scattering event, parton showering in the initial and final states can occur through hadronisation and radiation of quarks and gluons. This can be included into the matrix element computation but with serious limitations, particularly in case of a large number of particles in the final state.
- The underlying events: the parts of the protons not concerned with the central event can interact or decay.
- Hadronisation: low energy quarks and gluons (below the typical 1 GeV) cannot be assumed to be free and they have to recombine into colourless hadrons.
- Particles with a short lifetime, such as taus, have their decay simulated.
- Pile-up is added to reproduce the multiple interactions per bunch crossing in the output data. Mainly, soft inelastic collisions are added to the interesting hard scattering event. This procedure is quite difficult to calibrate, particularly in terms of expected number of vertices. The problem is avoided by reweighing the MC datasets to match the PU distributions observed in the data.

Various MC generators are available such as Pythia [37] MadGraph and Powheg. Pythia has the advantage of being a general purpose program describing both the hard process and the parton showering. It is however limited to LO computation and treats radiative corrections in the parton showering. This is not well suited for high momentum transfer and thus limits its use. MadGraph on the other hand computes radiative corrections in the matrix element and therefore is more accurate in high energy situations. Finally, Powheg is specialised in NLO calculation, therefore offering a more accurate estimate of inclusive cross sections. The events are then submitted to a dedicated toolkit called GEANT4 to simulate the passage of particles through the different CMS subdetectors and bring the simulation to the reconstructed level. The CMSSW library is a C++ dedicated software for the CMS experiment that contains knowledge on the detector and methods for physical analysis.

3.2 The parton distribution function formalism

At the LHC, the parton distribution functions (PDFs) represent the probability density to find a given parton p as a function of the fraction x of the total longitudinal proton momentum in the proton-proton centre of mass frame. These distributions depend on the 4-momentum transfer squared Q^2 and are labelled $f_p(x_p, Q^2)$. This dependence is not surprising as quantum fluctuations authorise an unrestricted number of gluons and quarks to room the nucleus at high energies. The three valence quarks picture (two ups, one down) has to be corrected to account for the possibility of having a sea of quarks and gluons.

The distributions themselves are derived by a phenomenological approach that matches experimental results, essentially from the study of lepton-hadron collisions. The measurement occurs at a given Q^2 and can be extrapolated to other energies through evolution equations (such as the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi equation, DGLAP [38]). This procedure is subject to large possible uncertainties being propagated. Different sets of PDF have been proposed by various collaborations. An example is proposed in figure 3.1 where the bumps at $x = 0.1$ for the up and down quarks indicate their valence status. For lower x, gluons start to dominate.

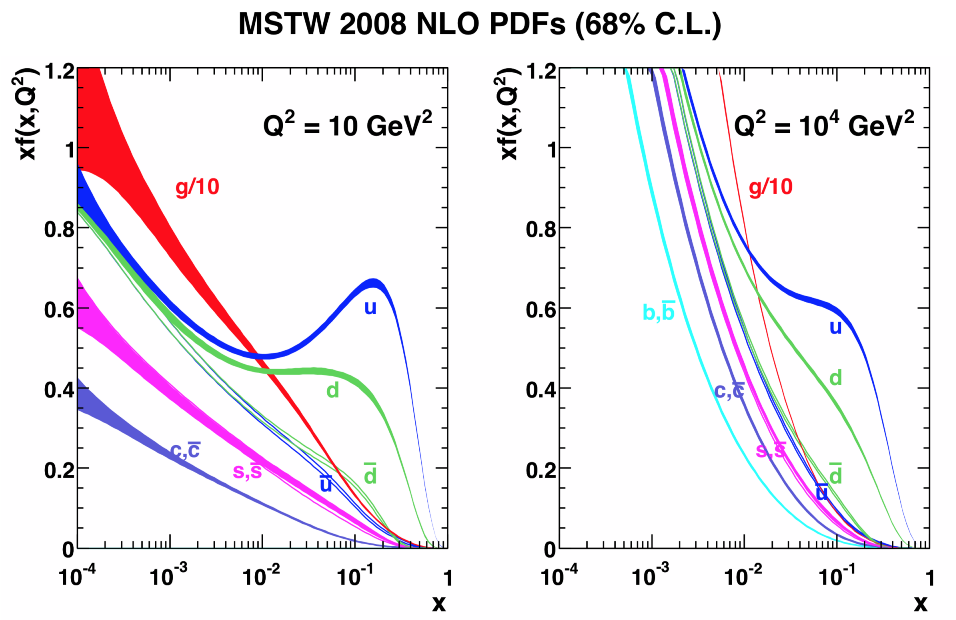


Figure 3.1 – The NLO PDF at the two different scales of $Q^2 = 10$ GeV^2 and $Q^2 = 10^4$ GeV^2 with 1 σ uncertainty bands (68 % confidence level, CL) from [38].

The centre of mass energy $\sqrt{\hat{s}}$ of the two partons involved in the hard scattering process is related to that of the two protons $\sqrt{s}$ via:

$$\sqrt{\hat{s}} = \sqrt{x_A x_B \cdot s} \tag{3.1}$$

where x_A and x_B are the fractions of momentum carried by the partons of, respectively, proton A and B that intervene in the hard scattering process. Note that, given the distributions, quarks (that can be valence ones) are far more likely to have a large momentum than antiquarks (which can only belong to the sea). Hence, a quark-antiquark interaction is on average boosted in the direction of the quark in the pp centre of mass frame.

Knowing a hard scattering cross section, one can thus compute the cross section for the process to happen from two protons colliding. Writing the cross section for partons A and B to form X as $\sigma_{AB\to X}$ and that for the total proton-proton to X cross section as $\sigma_{pp\to X}$, one gets [39]:

$$\sigma_{pp\to X} = \sum_{A,B=\{q,\bar{q},g\}} \int_0^1 dx_1 dx_2 \; f_A(x_1,\mu_f) f_B(x_2,\mu_f) \; \sigma_{AB\to X}(x_1,x_2,\mu_f,\alpha_s(\mu_R)) \tag{3.2}$$

where μ_f is the factorisation scale, the energy (Q^2) to which the PDFs where extrapolated, α_s is the strong coupling constant renormalised at the chosen scale μ_R, and the sum is over all possible parton configurations.

3.3 The photon PDF of the proton

Quantum fluctuations also allow for photons to contribute to the zoo of constituents sharing the proton momentum and a fast moving charged particle (or neutral one composed of charged elements) can have its electromagnetic field interpreted in terms of a photon distribution. The photon contribution is however much smaller since it is suppressed by a power of the electromagnetic coupling α (note that this also impacts the DGLAP formalism and splitting functions to account for QED corrections).

LUXqed is a model-independent evaluation of that photon parton distribution function (PPDF) inside the proton. The same approach as in the case of the regular PDF was roughly followed with elastic form factor fitted in electron or positron scattering off the photon field of the proton [40, 41]: $e + p \to e + X$. The idea for evaluating this process is to consider a new channel, beyond the Standard Model, with a flavour changing photon-lepton vertex. Obtaining a high precision PPDF $f_{\gamma/p}$ over a large range of energy proves quite difficult to tackle. Thanks to constraints set by the LUXqed calculations, the overall uncertainty of the PPDF furnished by a data-driven determination was reduced by more than an order of magnitude in the now state-of-the-art set NNPDF3.1luxQED [42]. This overall improvement in accuracy suggested that the PI process could become large at high invariant masses or p_T and constitute a serious source of theoretical uncertainty at high energy, motivating the recent studies of this process.

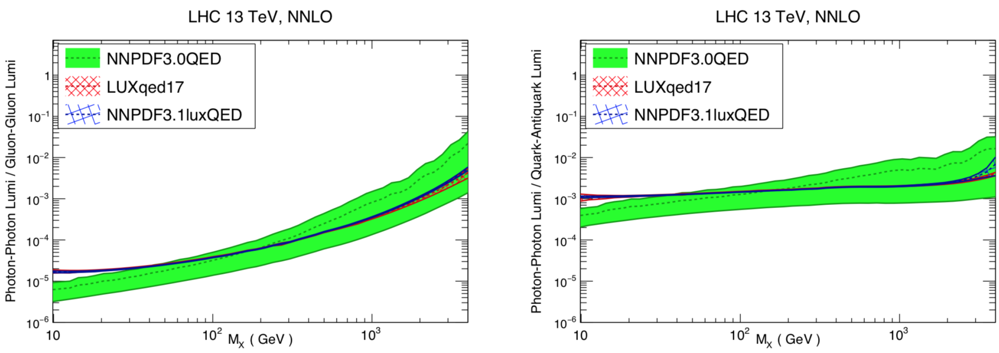


Figure 3.2 – The $\mathcal{L}_{\gamma\gamma}/\mathcal{L}_{gg}$ (left) and $\mathcal{L}_{\gamma\gamma}/\mathcal{L}_{q\bar{q}}$ (right) ratios for different PPDF [42].

There are multiple sources of uncertainties to the PPDF such as the elastic component and the inelastic resonance. These are all added at the last step of the fitting procedure. The resulting luminosity for the photon content $\mathcal{L}_{\gamma\gamma}$ obtained for the NNPDF3.1luxQED and for the previous NNPDF3.0QED and LUXqed17 are compared in figure 3.2 as ratios of the photon part on the gluons ($\mathcal{L}_{gg}$) and the quark-antiquark ($\mathcal{L}_{q\bar{q}}$) parts. The uncertainty on the result suggests a decisive improvement, though accuracy reduces at higher invariant masses. Indeed, figure 3.3 clearly shows this increasing uncertainty when reaching masses of above 2.5 TeV.

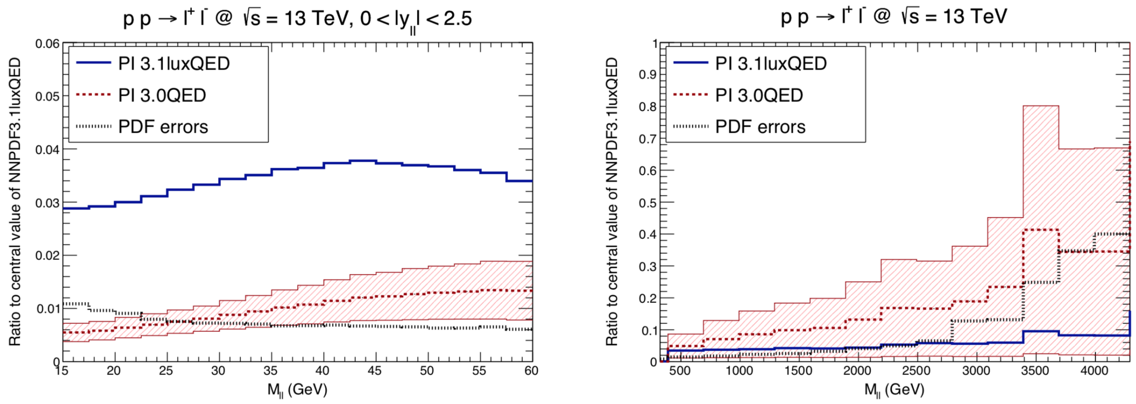


Figure 3.3 – Ratio of photon-initiated to DY neutral current lepton pair production as a function of the invariant mass of the pair $M_{l\bar{l}}$, for masses in the region [15, 60] GeV (left) and > 400 GeV (right). The NNPDF3.0QED uncertainty band is in red and the PDF errors give the uncertainties in NNPDF3.1luxQED [42].

3.4 Simulating the DY and PI

This section opens up with a discussion on how to best assess the forward-backward asymmetry, unveiled in section 1.5, by polling it through appropriate Monte Carlo simulations and an exotic new frame. The second part focuses on the generated level as it offers a good insight into the theoretical aspects of the hard process. The penultimate and final sections introduce the fully reconstructed simulations, that were carried out by the ULB group, and their normalisation.

3.4.1 Assessing the forward-backward asymmetry: the Collins-Soper frame

Foward-backward asymmetry is an important tool in the search for physics beyond the Standard Model. New neutral gauge bosons Z' and compositeness (a hypothetical new four-fermion effective interaction) could indeed lead to deviations from the standard expectations of this asymmetry [27]. At the experimental level, the forward-backward asymmetry can easily be displayed by showing the $\cos\theta$ distribution (defined in section 2.2) and its probable asymmetry in comparing the π bin to the $-\pi$.

There is however an important caveat in this discussion. The angle θ is defined in the $q\bar{q}$ frame in which the quark interacting has a known direction. It is however confined to the proton and only that direction and that of the out-coming leptons momenta are accessible. Hence, an exact measurement of θ is not possible if at least one of the initial parton has a non-null transverse momentum since the quark direction is not directly measurable. The way around this problem is to work with a privileged coordinate system in the study of the angular distribution of lepton pairs from dihadron interactions: the Collins-Soper frame (CS) [43].

Calling $\vec{p}_A$ and $\vec{p}_B$ the proton momenta in laboratory frame and $\vec{p'}_A$ and $\vec{p'}_B$ in the centre of mass of the dilepton pair, one defines the CS frame, in the centre of mass of the dilepton pair, with:

- the z-axis bisects $\vec{p'}_A$ and $-\vec{p'}_B$,
- the x-axis, orthogonal to z, is spanned by $\vec{p'}_A$ and $\vec{p'}_B$ with its direction such that the projections of these vectors along x are negative,
- the y-axis is set so the whole system is dextrogyral.

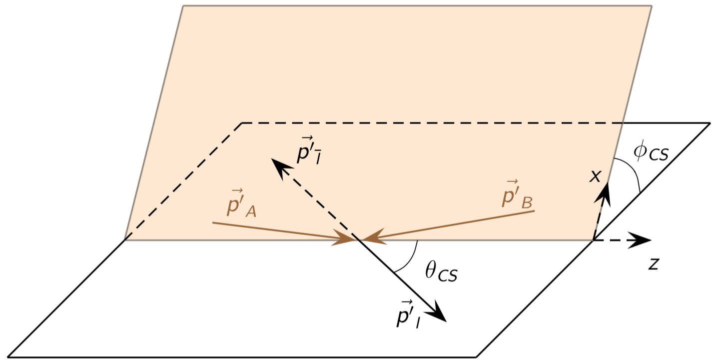


Figure 3.4 – The Collins-Soper frame. The protons momenta are in the orange plane and the leptons are emitted back-to-back in the white plane [15].

The frame is displayed in figure 3.4. In this configuration, $\cos\theta_{CS}$ can be expressed as derived in appendix B:

$$\cos\theta_{CS} = 2\frac{E_{\bar{l}}\, p_{z,l} - E_l\, p_{z,\bar{l}}}{M_{l\bar{l}}\,\sqrt{M_{l\bar{l}}^2 + P_{T,l\bar{l}}^2}}. \tag{3.3}$$

This frame offers a good approximation of the angle θ, between the negative lepton and one of the parton momenta, through the angle θ_{CS} if the transverse boost of the lepton pair is small compared to its longitudinal boost. In this context however, the case of $q\bar{q}$ interaction, the angle θ should be restrain to the one between the negatively charged lepton and the quark, not the antiquark. This can be taken into account by requiring hadron A to offer the quark parton. Hence, taking z to point in the same direction as A, the sign of the cosine can be tuned to match this requirement:

$$\cos\theta_{CS} = 2\frac{E_{\bar{l}}\, p_{z,l} - E_l\, p_{z,\bar{l}}}{M_{l\bar{l}}\,\sqrt{M_{l\bar{l}}^2 + P_{T,l\bar{l}}^2}}\frac{p_{z,q}}{|p_{z,q}|}, \tag{3.4}$$

where the last term corrects the sign depending on the direction of the quark. The final subtlety is the impossibility to access the quark direction, given the nature of hadrons. This is where the PDF matters: one indeed knows that, in most cases, proton-proton collisions have quarks carrying a larger momentum than antiquarks, since the former are valence ones and the latter originate from the sea. Hence the $q\bar{q}$ pair is most oftenly boosted in the quark direction and the hadron A is then taken to be the one which direction, in the laboratory frame, is the same as the longitudinal boost of the dilepton pair. This comes down to estimating $\frac{p_{z,q}}{|p_{z,q}|} \approx \frac{p_{z,l\bar{l}}}{|p_{z,l\bar{l}}|}$ to get:

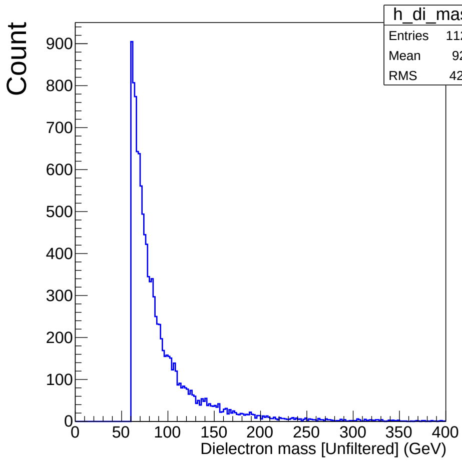
$$\cos\theta_{CS} \approx 2\frac{p_{z,l\bar{l}}}{|p_{z,l\bar{l}}|}\frac{E_{\bar{l}}\, p_{z,l} - E_l\, p_{z,\bar{l}}}{M_{l\bar{l}}\,\sqrt{M_{l\bar{l}}^2 + P_{T,l\bar{l}}^2}}. \tag{3.5}$$

Note that this approximation slightly dilutes the asymmetry since, however unlikely, it may sometimes be possible for the antiquark to have a larger $|p_z|$ than the quark, hence causing an error of π on the angle. To insist on its centre of mass frame status, the angle is sometimes fitted with a "*", though this convention will not always be adopted and the reference frame considered should be clear from context (the laboratory frame version is barely used).

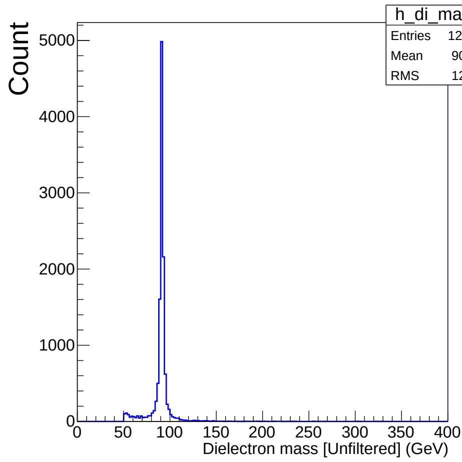
3.4.2 The generated level

The first step in the analysis consists in a generated level study: directly simulating the hard scattering process and searching for any specific differences in kinematics of the final state between the irreducible background (DY) and the signal (PI), before comparing to the data at the reconstructed level. The simulations were run solely for electrons and positrons (no significant differences at the generated level are expected with muons or even taus).

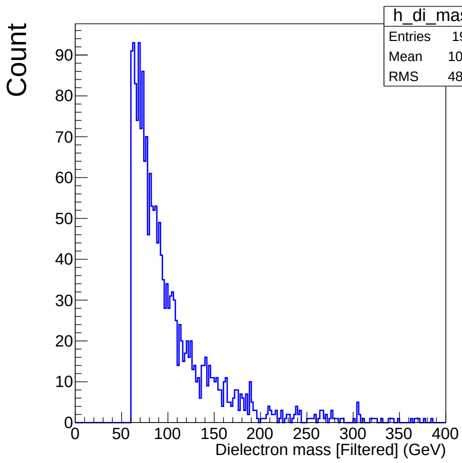
This first analysis exploits simulations run by the ULB group. The DY is simulated to NLO using MadGraph and the PI to LO with the PPDF *LUXQED17PDF4LHC15* on Pythia 8, each for two different bins of mass. Accessed with the help of a ROOT command file personally updated, the files were analysed with interesting distributions stored into histograms and various variables polled to search for interesting behaviours. The first one, the invariant mass, is presented in figure 3.5 for both pre and post acceptance cuts, defined in section 2.3.6.



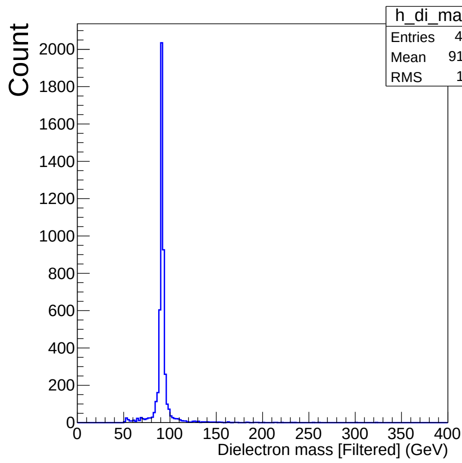
(a) PI: $M_{l\bar{l}} > 60$ GeV.



(b) DY: $M_{l\bar{l}} > 50$ GeV.



(c) PI: $M_{l\bar{l}} > 60$ GeV.



(d) DY: $M_{l\bar{l}} > 50$ GeV.

Figure 3.5 – Invariant mass of the dilepton pair at the generated level before (top) and after (bottom) acceptance cuts.

For the DY, a very noticeable peak rises around the Z boson mass, about 91 GeV, as expected. The acceptance cuts have a larger impact on the PI than the DY as the former reduces to about a

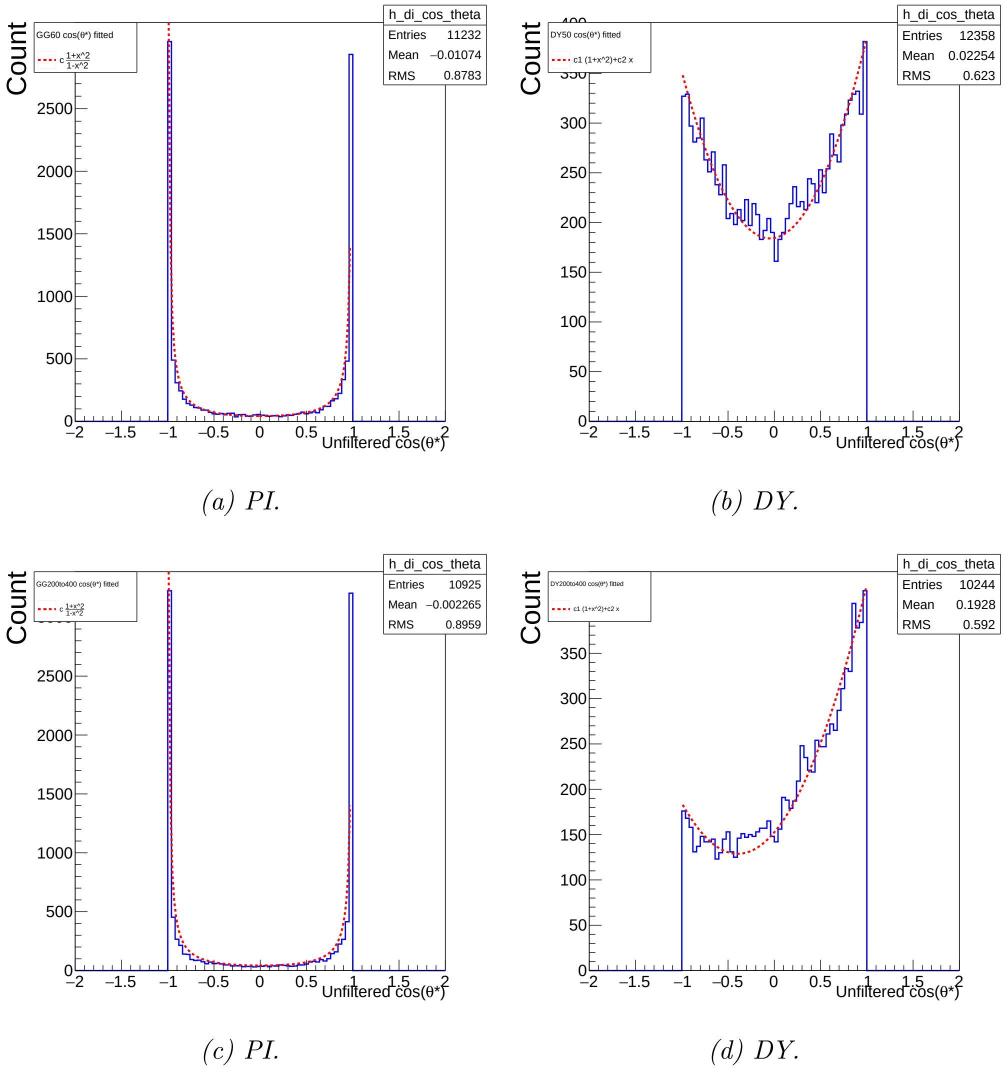
17 % of its initial population while the latter diminishes to only 39 %. The overall behaviour of the distributions is however not dramatically impacted by the cuts.

The $\cos\theta_{CS}$ offers a good opportunity to control the angular behaviour of the cross sections derived in the first chapter as well as a powerful test of the most promising feature: the forward-backward asymmetry. Indeed, fitting both distributions with the following equations (inspired by equation 1.10 for the DY and equation 1.23 for the PI respectively):

$$c_1\ (1+\cos^2\theta) + c_2\ \cos\theta, \tag{3.6}$$

$$c\ \frac{1+\cos^2\theta}{1-\cos^2\theta}, \tag{3.7}$$

where the first one, for the DY, is in fact valid for any spin-1 mediator. Fitting these to the simulations, one obtains figure 3.6, where the quality of the fit is self-explanatory.

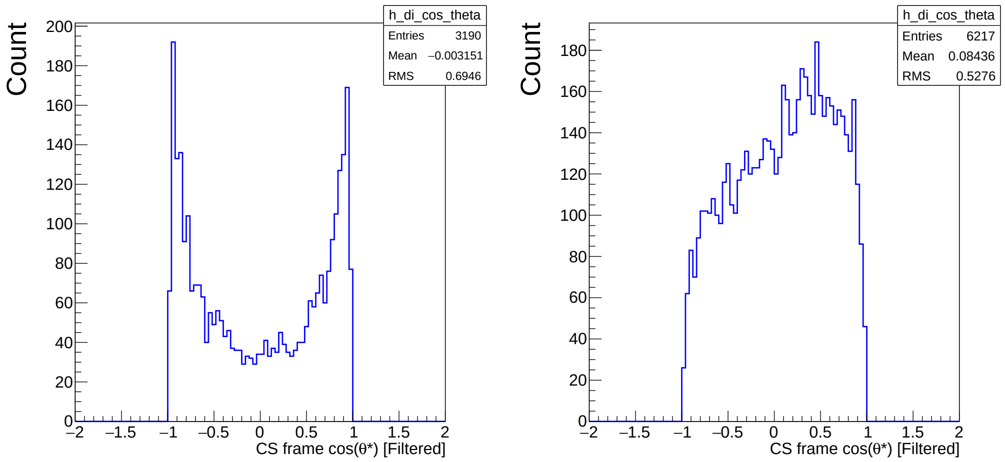


(a) PI. *(b) DY.*

(c) PI. *(d) DY.*

Figure 3.6 – The $\cos\theta_{CS}$ distributions (blue) and fits (dotted red), $M_{l\bar{l}} > 60$ GeV (top) and $M_{l\bar{l}} \in [200, 400]$ GeV (bottom) before acceptance cuts.

The forward-backward asymmetry is easy to spot: the PI is perfectly symmetric under a $\pi \to -\pi$ flip of the θ_{CS} angle while the DY is not. Interestingly, the PI is far more concentrated on the extremities of the distribution while the DY is flatter. The effect of the asymmetry in the DY is much more pronounced in the mass bin above the Z boson resonance, as expected from the study of the cross section [44]. Unfortunately, this distinct behaviour is slightly shattered when taking into account the acceptance cuts and the π error when the $\bar{q}$ has a larger $|p_z|$ than the quark, as indicated in figure 3.7.



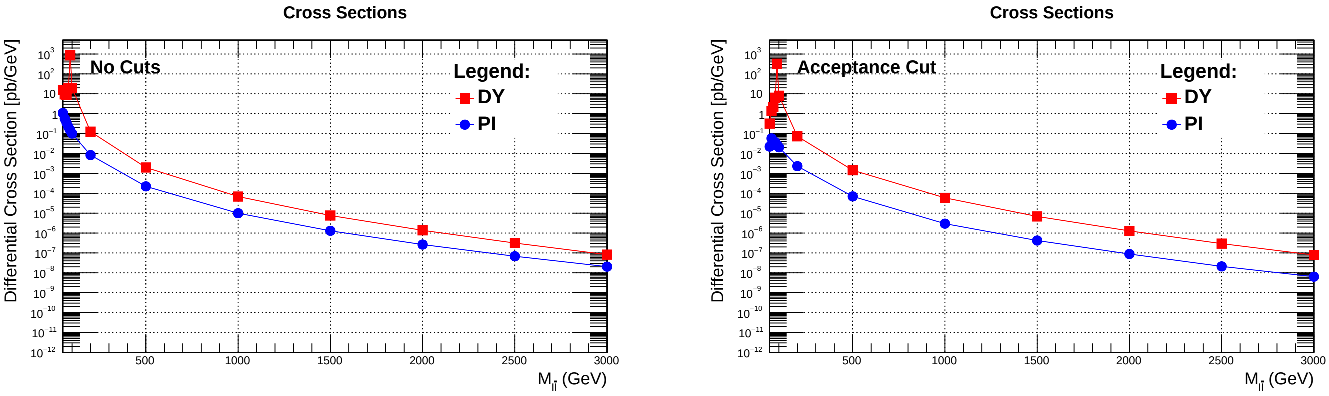
(a) PI. *(b) DY.*

Figure 3.7 – The $\cos\theta_{CS}$ *distributions with* $M_{l\bar{l}} \in$ *[200, 400] GeV and with acceptance cuts.*

The acceptance cuts have a significant impact on the extremities of the distributions. Indeed, the $\cos\theta$ and pseudo-rapidity (hence rapidity given the ultra-relativistic regime) are not uncorrelated. A large $|\cos\theta|$ implies the leptons are emitted with angles close to 0 and π (they are indeed back-to-back in the centre of mass frame of the dilepton pair, where this variable is considered) and they are thus very forward. Now, the acceptance cuts remove events with $|\eta|$ larger than 2.5, in the laboratory frame, and this condition is likely to happen in a very forward event with a large rapidity $y_{l\bar{l}}$ (given the large $|p_z|$), leading the electron to fall out of the detection window. Other plots derived from these simulations are proposed in appendix C. In particular, figure C.6 indicates electrons are preferentially emitted with a larger rapidity than the positrons in the case of the DY, another sign of the forward-backward asymmetry.

The cross sections themselves can be computed to leading order as a function of the mass with Pythia. In order to compare how both the PI and DY processes evolve with energy, a privately produced Pythia simulation was written using the set *LHAPDF6:LUXqed17_plus_PDF4LHC15_nnlo_100* and the default PDF set for the PI and DY respectively. This simulation considers all pairs of leptons for the final state. The differential cross section is in fact polled by asking the simulation to compute the cross sections in a succession of mass intervals of 1 GeV starting at the values 50, 60, 70, 80, 90, 100, 200, 500, 1000, 1500, 2000, 2500, and 3000 GeV (thus the sets [50, 51] GeV, [60, 61] GeV, ..., [3000, 3001] GeV). For each computation, 2000 events were demanded and protons were collided at a centre of mass energy of 13 TeV. The simulation also stores into a TTree (a type of ROOT file storing information, leaves, into branches of a single file, colourfully called a tree) the leptons, quarks and photons p_T, p_z, η, ϕ and θ as well as their Pythia ID (an integer number assigned to each particle species to identify it).

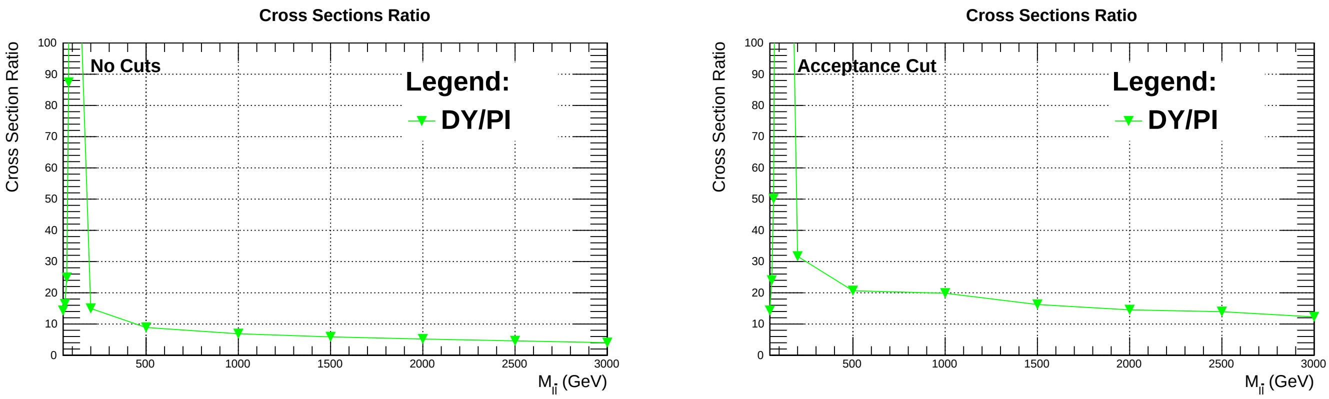
An acceptance count computes the fraction of events passing certain control checks as well as the acceptance cuts (for leptons: $p_T > 25$ GeV and $|\eta| < 2.5$). This results into a set of text files containing the cross sections, their errors and acceptance ratios as well as a ROOT file containing the TTree with the variables aforementioned. The computation was carried out on the *m-machine* server of the IIHE research department while the output files were analysed locally with ROOT macros to produce the following set of plots.



(a) *Without acceptance cuts.* (b) *With acceptance cuts.*

Figure 3.8 – The cross sections for the PI (blue) and DY (red) as a function of $M_{l\bar{l}}$.

Figure 3.8 presents the DY and PI cross sections as a function of $M_{l\bar{l}}$, without and with acceptance cuts. The peak around 91 GeV of the DY distribution corresponds again to the Z boson. At this mass, the DY clearly dominates the PI. However, as displayed in figure 3.9, the ratio DY over PI decreases with $M_{l\bar{l}}$, indicating a resurgence of the PI.



(a) *Without acceptance cuts.* (b) *With acceptance cuts.*

Figure 3.9 – The cross sections ratios DY/PI as a function of $M_{l\bar{l}}$.

Clearly, applying the acceptance cuts increases the ratio DY/PI, thus favouring the Drell-Yan over the signal. The acceptance cuts for the DY and PI are displayed in figure 3.10. Note how the acceptance starts by strongly reducing both distributions (this is due to the p_T cut that is more likely to be met for events with a larger invariant mass) but then tends to 0.95 for the DY and 0.3 for the PI at masses of a few TeV.

The interesting question now is to consider what type of cuts could be applied to the processes to favour the photon-induced. Remembering the discussion of the previous section, one can be tempted

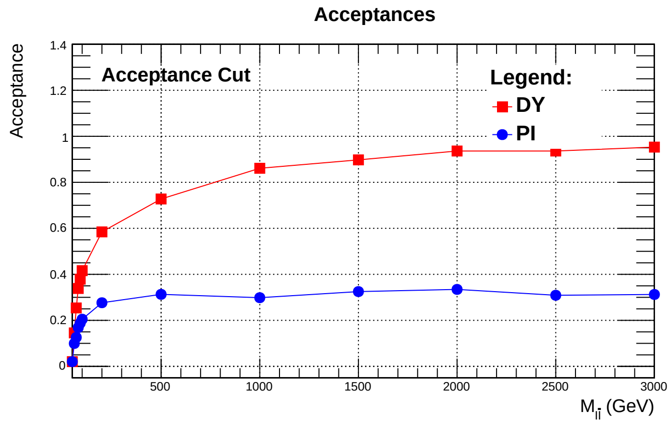


Figure 3.10 – The acceptance cut ratios for the DY (red) and PI (blue).

to cut on $\cos\theta_{CS}$ and exploiting the forward-backward asymmetry. Indeed, the DY is proportionally more tilted towards $\theta_{CS} = 0$ than $\theta_{CS} = \pi$. For example, one could demand three bins in $\cos\theta_{CS}$: below -0.8, in [-0.8, 0.8] or above 0.8. Joined to that, one could separate each distribution into two pieces demanding a cut on the rapidity with $|y_{l\bar{l}}|$ below or above 0.75. Initially, three bins in rapidity were demanded with the largest one covering events with a rapidity exceeding 1.5. However, very few such events can pass the acceptance cuts when this condition is joined with the top bin in $\cos\theta_{CS}$ as these events often have a lepton falling out of the acceptance. To summarise, the cuts applied are:

<table>
<tr><td>$\cos\theta_{CS}$</td><td colspan="2">$<$ -0.8</td><td colspan="2">[-0.8, 0.8]</td><td>$>$ 0.8</td></tr>
<tr><td>$|y_{l\bar{l}}|$</td><td colspan="3">$<$ 0.75</td><td colspan="2">$>$ 0.75</td></tr>
</table>

They lead to the set of acceptance ratio distributions displayed in figure 3.11. Evidently, asking a $\cos\theta_{CS}$ in [-1, -0.8] strongly favours the PI. Enough so that the acceptance gets better for this process than for the DY. This comes at a heavy price as most events are rejected, with only between 5 to 10 % accepted in the low rapidity bin and even less in the high one.

One can clearly observe how a high rapidity mixed with a large value of $|\cos\theta_{CS}|$ diminishes drastically the acceptance, due to the aforementioned effect. Indeed, θ_{CS} is, in the centre of mass of the lepton pair frame, connected to the expression of the pseudo-rapidity η, in the laboratory frame, which is limited by the geometrical acceptance. As this last variable, in the ultra-relativistic case, correctly approximates the rapidity y, the connection between both variables justifies the rarity of events in the bins where the $|\eta| < 2.5$ cut has the largest impact: $|y_{l\bar{l}}| > 0.75$ with the extremal $\cos\theta$ bins.

The set of cuts proposed thus indicates it is possible to favour the PI by an appropriate use of relevant variables. They however are not sufficient to let the signal dominates the DY as indicated by figure C.17 of appendix C.

3.4.3 The reconstructed level

At the reconstructed level, the simulations follow the same steps as the generated level plus a layer of detector reconstruction added on top and furnished by the GEANT4 program. The files were produced by the ULB group with different generators employed for the various backgrounds and signal. The

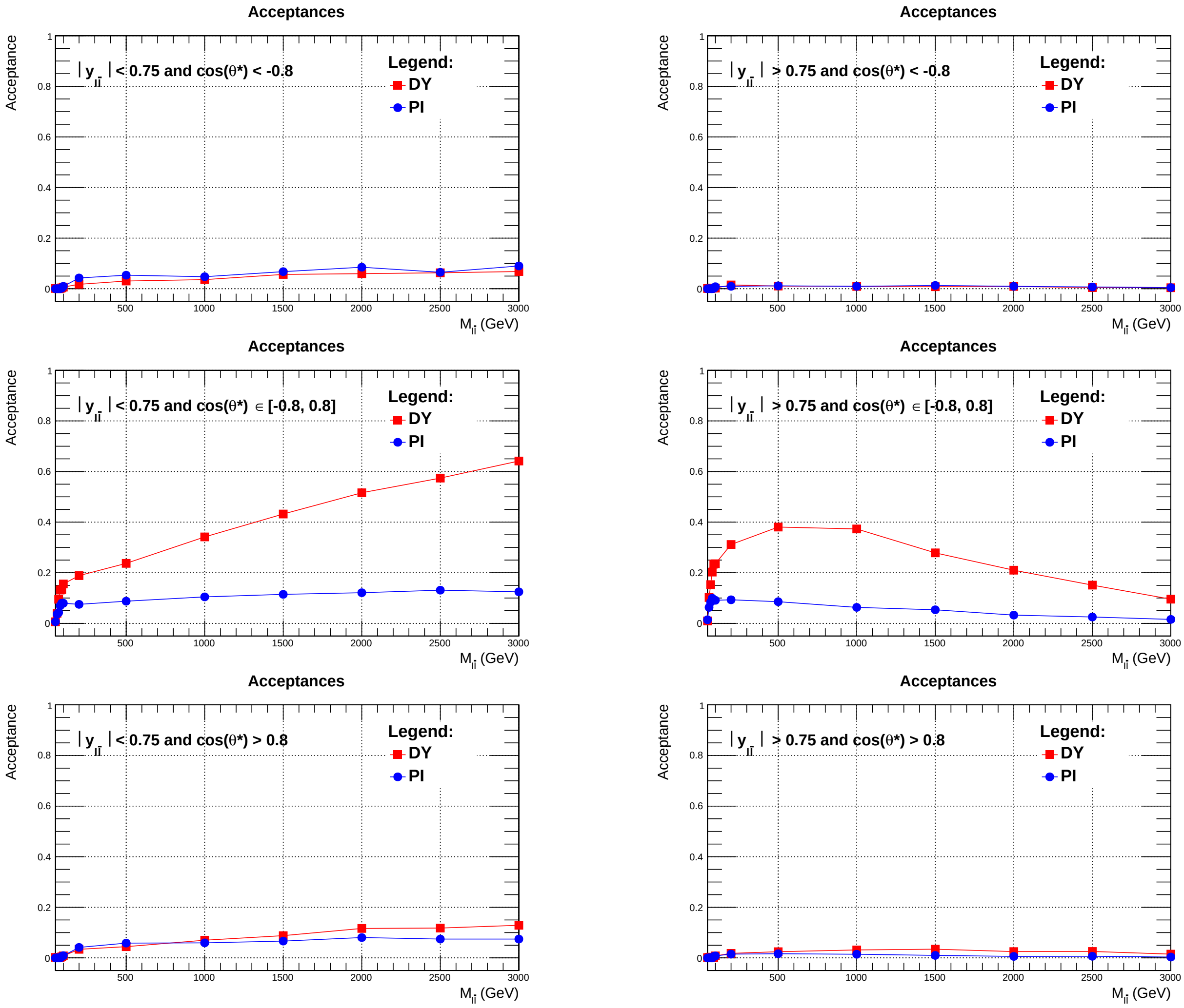


Figure 3.11 – Acceptance distributions as a function of the dilepton invariant mass $M_{l\bar{l}}$ for the different sets of cuts.

cross sections are computed to either NLO or NNLO from theoretical models and are used to reweigh simulations that are NLO for all backgrounds. The signal, the PI, is LO. Another process that could occur from a diphoton initial states is, for example, the $\gamma\gamma \rightarrow WW$, which cross section in fact dominates the PI at higher masses [45, 46]. However, restricting the final state to dilepton events incurs an additional factor of 0.1 squared to this process due to the branching ratio of the decay $W \rightarrow l + \nu$. Hence this process will not be considered here given this cumbersome attenuation.

Parton showering and hadronisation are described using the Pythia 8.2 generator [37] with the CUETP8M1 (for data collected in 2016) and CP5 (for 2017 and 2018) tunes for the underlying event [47, 48, 49]. The PI comes from privately produced samples using only 2018 conditions and detector simulations. Ideally, a different simulation should have been produced for each year. This is however not as dramatic as one could fear, since most of the data comes from 2018 (the luminosity being much higher than in 2016 and 2017) and since the 2017 data taking campaign was run under quite similar conditions to 2018. There are however significant differences in settings with 2016, both at the level of the tune and the pixel detector, the latter having been replaced after that year, and the PI simulation may be over optimistic for 2016. This will be shown to impact an important variable in the following steps: the number of charged hadrons (nCH) connected to the vertex fit.

All other backgrounds that lead to a dilepton pair in the analysis are simulated with a proper configuration for each year. For the main ones:

- The DY: NLO simulation with MadGraph [50, 51].
- The $t\bar{t}$: NLO simulation with Powheg.
- The WW, WZ and ZZ: NLO simulation with Powheg [52].
- The $t\bar{t} + W/Z$: NLO simulation with MadGraph.
- The W + jets: NLO simulation with MadGraph.
- The single $t + W$: NLO simulation with Powheg.

The data is separated into three distinct sets of files for 2016, 2017, and 2018 with integrated luminosities of 36, 41, and 60 fb^{-1}, hence a total of 137 fb^{-1}. All files are stored on the *m-machine* server of the IIHE department and an analysis program coded in an object oriented way in C++ is used to go through these various ROOT files and generate the desired histograms and text files.

3.4.4 The reconstructed level: Data/MC normalisation

The selection procedure to isolate interesting events is introduced in section 2.3.6. Several corrections to the simulations should be applied to properly describe the data. Mainly:

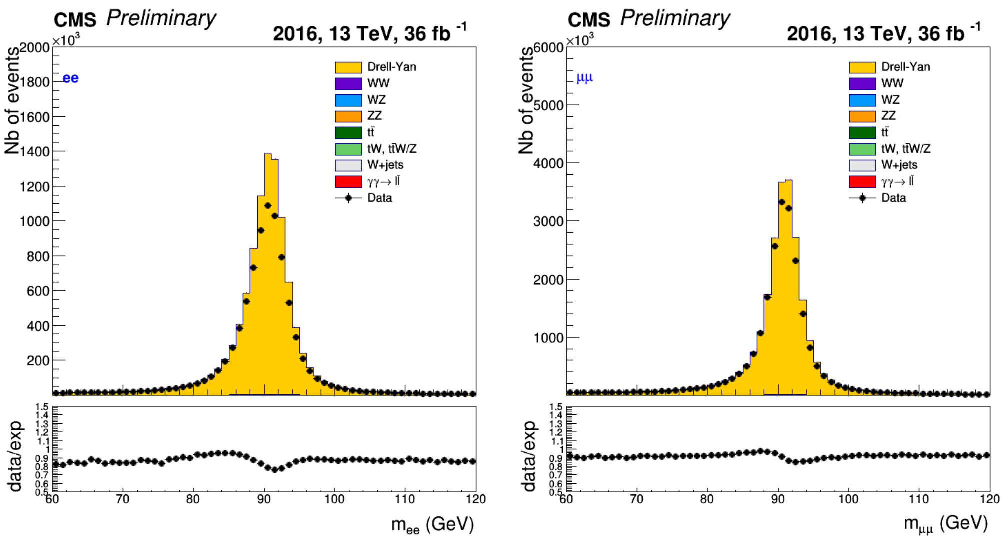
- The distribution of the number of extra proton-proton interactions is made similar to data.
- A "prefiring correction", describing a triggering inefficiency (at L1) observed in 2016 and 2017 related to time shifted ECAL deposits at high pseudorapidity.
- A scale factor related to the lepton selection and triggering efficiency.

The former two are applied using weights centrally provided by the CMS collaboration while the latter was personally derived for this work. Figure 3.12 displays the resulting invariant mass spectra for 2016 before applying the normalisation scale factors, with data and simulations normalised to the luminosity. The bottom part of the figure displays the ratio of the data over the sum of simulation expectations for each bin. The agreement between predictions and theory is tested by checking for a flat ratio overall compatible with 1 inside errors.

A special program was thus written to count the events in the 60 to 120 GeV invariant mass spectrum region of the dilepton pair for ee, $\mu\mu$ and $ee + \mu\mu$ in the data and simulations. A constant weight was derived for each year so that the integral of all expectations over that mass window equals that of the data. The obtained weights are displayed in table 3.1.

Year	$ee + \mu\mu$	ee	$\mu\mu$
2016	0.866757	0.803477	0.894007
2017	0.883595	0.761664	0.945355
2018	0.882376	0.768371	0.940456
All	0.878600	0.774872	0.929102

Table 3.1 – The Z-mass peak matching weights for all years and lepton configurations.



(a) ee. *(b) $\mu\mu$.*

Figure 3.12 – The invariant mass spectra before normalisation scale factors (2016 data).

These scale factors can be interpreted as the square of the data/MC lepton efficiency ratio. They are assumed to be energy and η independent, a hypothesis that will be verified in the next chapter. The mass region centred around the Z boson peak is a natural choice, since there the DY dominates and the physics is well understood.

Summary of the Chapter

The important parton distribution functions unveiling the content of the proton were introduced in this chapter, as well as the basic methodology to analyse the data of the CMS experiment. Monte Carlo simulations are explored to identify optimal cuts to dissociate the backgrounds from the signal. An introductory discussion opened the study by going through the generated level distributions. It showed the possibility to favour the PI over the DY. The forward-backward asymmetry, defined in the first chapter, was also shown to be consistent with the expected angular behaviour derived from the simulations. In order to observe this, the Collins-Soper frame was introduced to approximate the angle between the negatively charged lepton and the quark, because the direction of the latter is impossible to measure due to colour-balance requirement in hadrons. Finally, the reconstructed files were presented and the normalisation scaling factors derived.

CHAPTER 4

Selecting the PI

If a picture is worth a thousand words, surely a plot as figure 4.1 is worth a Master thesis. This histogram represents the invariant mass distribution of dielectron and dimuon pairs recorded by the CMS detector (black dots) and expected from the various scaled SM backgrounds (stacked histograms) with the photon-induced (PI) signal in red, for a range starting at 50 GeV and up to 3 TeV.

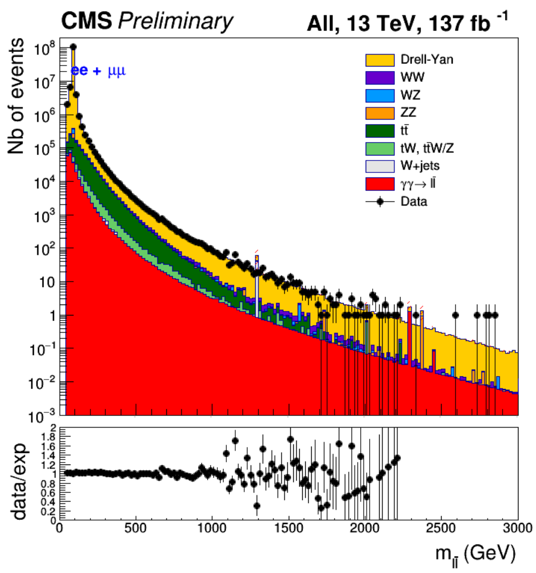


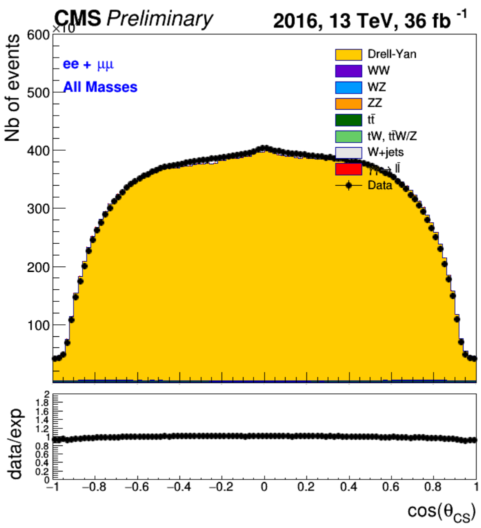
Figure 4.1 – The invariant mass spectrum of the $ee/\mu\mu$ dilepton pairs for all years of Run II.

Black bars indicate the statistical incertitude of the data and red dashed area that of the simulations, due to limitations in number of events generated. Evidently, for low masses, the agreement is indeed verified in this case, thanks to the scaling introduced in section 3.4.4. Regarding the information unveiled by the histogram, a few things are interesting to mention, such as the presence of the Z boson resonance around the 91 GeV mass and that the Drell-Yan dominates the other backgrounds and the signal. The PI does start to play a significant role when the mass increases above a few TeV.

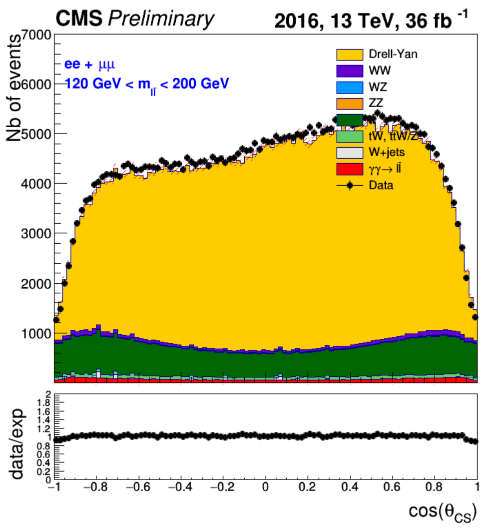
This chapter aspires to two things. The first one is to introduce an optimal set of cuts to isolate the PI signal from other backgrounds. It then moves on to a discussion on the quality of the simulations of various parameters and, in particular, how to correct the simulations to more closely match the data. Most of the analysis is constructed using the 2016 dataset. The entire dataset is then exploited, boosting the luminosity from 36 to 137 fb^{-1}.

4.1 The $\cos\theta$ distribution before background cleaning

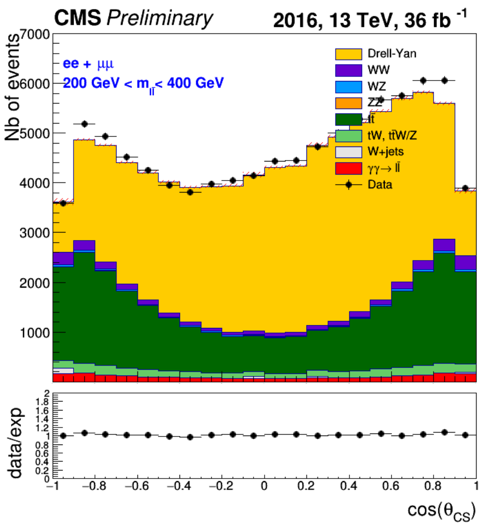
Isolating the photon-induced lepton pair production from other backgrounds is not an easy task. It is indeed vastly dominated by the other distributions at low masses and has an irreducible background: the Drell-Yan. From the discussion at generated level of section 3.4.2, the forward-backward asymmetry exhibited by this last process could be used to disentangle them.



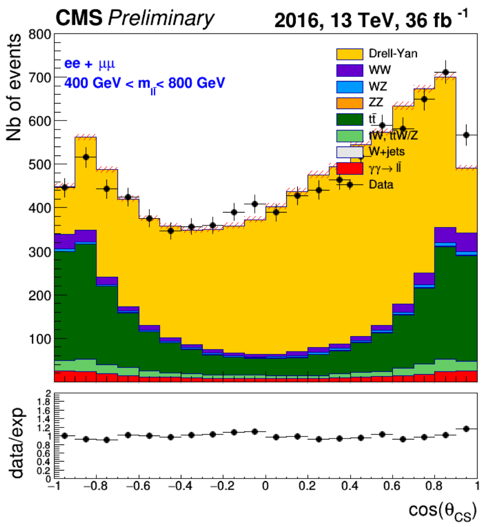
(a) All masses ($M_{l\bar{l}} > 50$ GeV).



(b) Mass bin: [120, 200] GeV.



(c) Mass bin: [200, 400] GeV.



(d) Mass bin: [400, 800] GeV.

Figure 4.2 – The $\cos\theta_{CS}$ distributions for $ee/\mu\mu$ dilepton pair events (2016 data) for different mass bins.

The main idea to pursue the separation is to go back to the topology of the events and to search for variables distinguishing the PI from other processes. Figure 4.2 displays the $\cos\theta_{CS}$ distribution for $ee/\mu\mu$ dilepton pair events (2016 data) for different mass bins. The stability of the Data/MC ratio as a function of $\cos\theta$ for various mass bins validates the hypothesis of flat normalisation scaling factors discussed in section 3.4.4. From now on, these normalisation weights are automatically included.

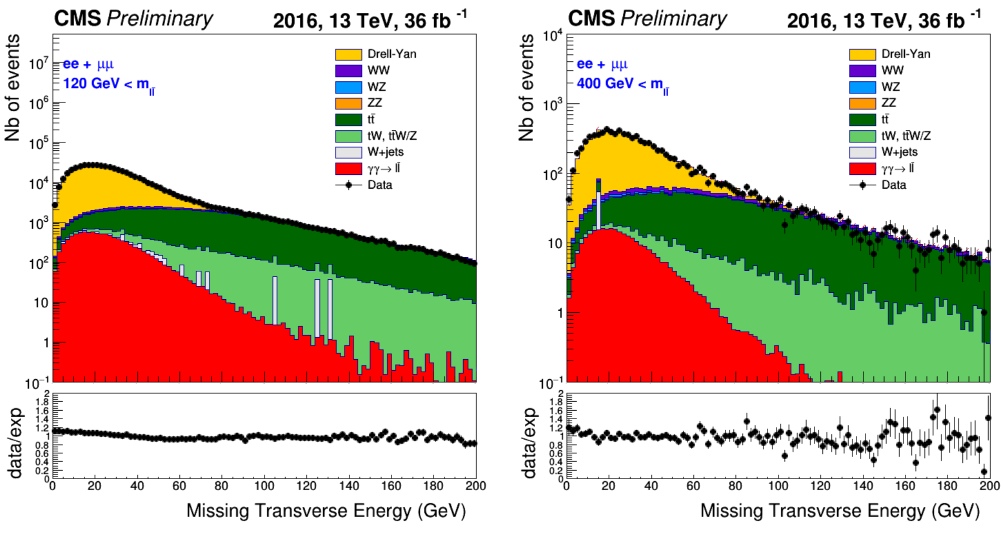
The forward-backward asymmetry of the DY becomes more visible at higher invariant masses, as discussed in previous sections. Note how the $t\bar{t}$ and WW backgrounds mimic the PI $\cos\theta_{CS}$ distribution.

4.2 Background rejection criteria for the $t\bar{t}$ and WW

In the next two sections, a missing transverse energy (MET, defined in section 2.3.4) and a number of jet cuts are investigated to suppress the $t\bar{t}$ and WW backgrounds. In both these processes, neutrinos are produced, hence some MET is expected. Additionally for the $t\bar{t}$ background, extra jets are emitted. These two characteristics are not expected to be present in the cases of the DY and PI and thus offer a natural choice of variables to investigate to suppress the backgrounds.

4.2.1 Missing transverse energy and its significance

Let us first explore the case of missing transverse energy, with figure 4.3 displaying that information for different mass bins. Note the good agreement between expectations and data.

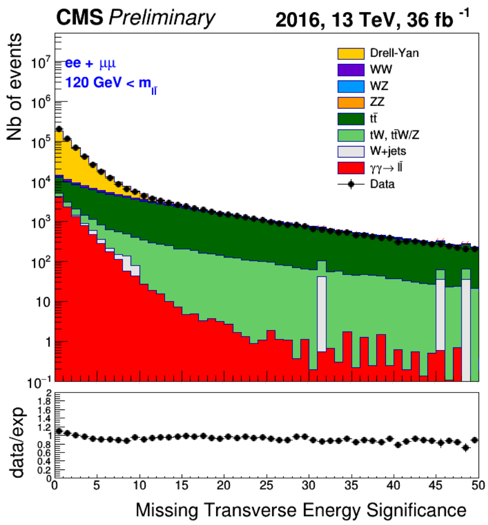


(a) Mass bin: above 120 GeV. *(b) Mass bin: above 400 GeV.*

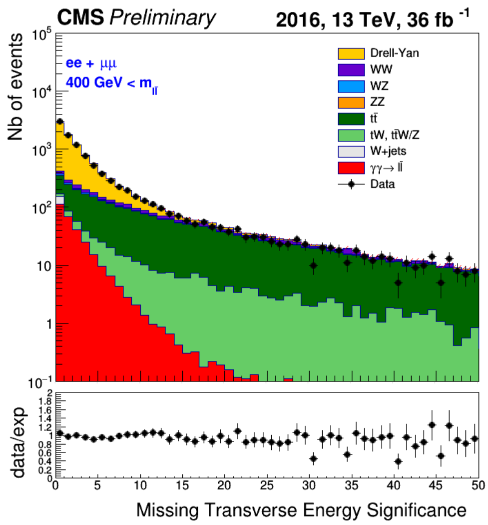
Figure 4.3 – The missing transverse energy distributions for different mass bins (2016 data).

For the $t\bar{t}$ distribution, the missing transverse energy does peak at higher values than the PI, as indicated by its increasing contribution. The WW also displays this tendency, though reading it is made harder by the log-scale. Except in the rare case of light leptons pairs emanating from the decay of a tauonic pair, no MET is expected in the DY and PI processes. The non-null measured MET is therefore mainly related to the finite resolution of the CMS detector, mostly due to the neutral particles energy resolution and the pile-up contribution.

While the transverse momentum of charged particles can accurately be measured using both the curvature of their track and the calorimeters, neutral particles momentum measurement relies only on the calorimeter information which suffers from limited resolution at low p_T. At large PU, such low p_T particles are produced in large number resulting in a sizeable degradation of the MET resolution. In order to mitigate this effect, an alternative variable to the MET can be exploited: the "MET significance" (metSig) [53]. It provides an estimate of how likely the MET measured in a given event comes purely from finite resolution effects. The resulting metSig is less corrupted by the effect of pile-up and still denounces the presence of neutrinos in the final state. Observe figure 4.4, displaying the equivalent information to figure 4.3 with the metSig.



(a) Mass bin: above 120 GeV.



(b) Mass bin: above 400 GeV.

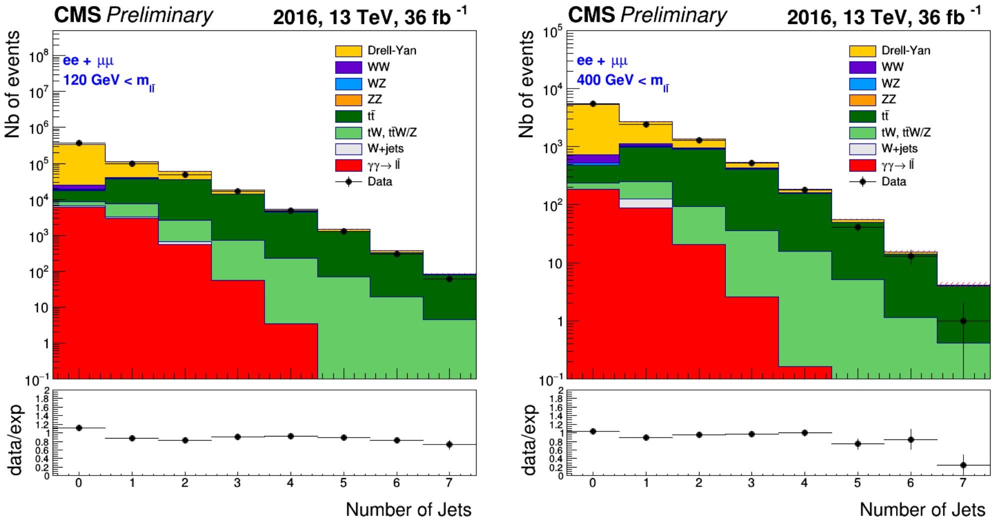
Figure 4.4 – The missing transverse energy significance distributions for different mass bins (2016 data).

While the differences between figure 4.3 and 4.4 are striking, particularly concerning the behaviour of the DY and PI, the interpretation stays the same. The key advantage of the updated variable resides in the way the DY and PI concentrate and stick to low values of the metSig. It offers a natural choice for cutting the distributions while keeping a large fraction of these processes. Indeed, restricting to events with a metSig inferior to ten seems like an appropriate cut to favour both the DY and PI at every mass bin considered. Its impact on the $t\bar{t}$ promises to be decisive.

Note that the discussion above is made simpler by the decision to limit the analysis to electronic and muonic leptons. Taus would indeed, in both their hadronic and leptonic decays, involve neutrinos (e.g. $\tau \to \nu_\tau + W \to \nu_\tau + e\,\nu_e$ or $\tau \to \nu_\tau + W \to \nu_\tau + \mu\,\nu_\mu$) in the final state and thus the emergence of both MET and metSig. Adapting the analysis to include these heavy cousins of the leptonic family would therefore demand another strategy to isolate both the PI and DY from the rest of the backgrounds. For the same reason, this cut also provides an attenuation of the $\gamma\gamma \to WW$ that was neglected in the simulations, as discussed in section 3.4.3.

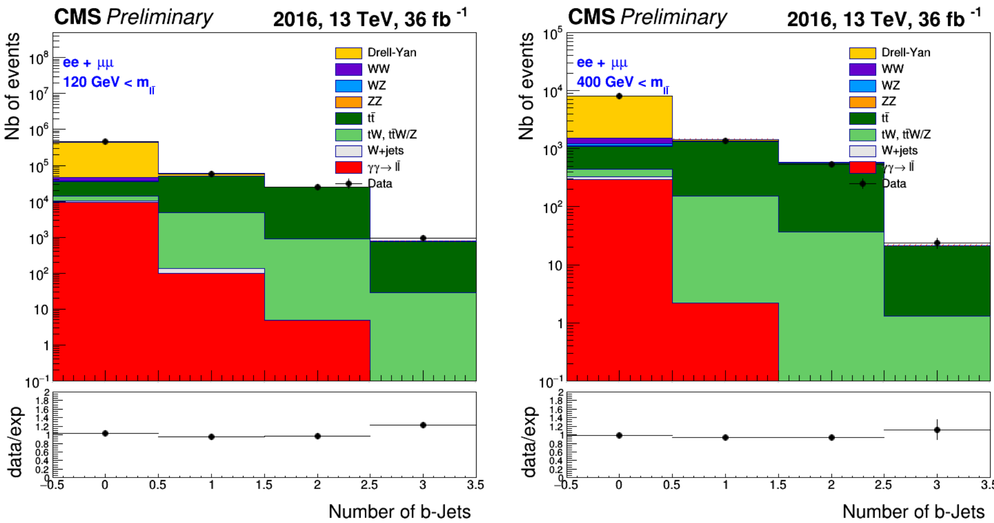
4.2.2 Number of jets and number of b-jets

For the DY and PI processes, jets are not commonly expected to be present in the final state but may occur as part of the underlying event, pile-up or NLO subprocesses. In opposition, the $t\bar{t}$ background involves quarks in the final state hence jets are expected. Figure 4.5 and 4.6 respectively display the number of jets and b-jets (the latter has a specific treatment, as discussed in section 2.3.3).



(a) Mass bin: above 120 GeV. *(b) Mass bin: above 400 GeV.*

Figure 4.5 – Distributions of the number of jets for different mass bins (2016 data).



(a) Mass bin: above 120 GeV. *(b) Mass bin: above 400 GeV.*

Figure 4.6 – Distributions of the number of b-jets for different mass bins (2016 data).

As expected, the $t\bar{t}$ peaks at larger number of jets while both the DY and PI concentrate around no jets. Note however that having a single jet is not so uncommon in the case of the two "signals" though having a b-jet is rarer, given the approximate 10^2 factor of difference between 0 and 1 b-jet for the PI and DY distributions. A cut tolerating as much as one jet and not a single b-jet would clearly favour the DY and PI over the $t\bar{t}$.

4.2.3 Other possible cuts

Other variables offer the possibility to disentangle the DY and PI from the rest of the backgrounds. To mention only a few, cuts on the lepton pair p_T, the equivalent but somewhat normalised $\frac{p_T}{m_{l\bar{l}}}$, the difference in azimuthal angle of the leptons $\Delta\phi$, and the equivalent acoplanarity $1-\frac{|\Delta\phi|}{\pi}$ (measuring the degree of deviation from coplanarity) have all been investigated. Indeed, for the first variable, one expects the DY and PI to peak around 0 while the other processes do not generate back-to-back leptons and can thus have larger values of the pair p_T. For the last two indicators, once again the back-to-back nature of the leptons generated by the signals favours them to have an acoplanarity tending to nought, since $|\Delta\phi| \sim \pi$. These behaviours are indeed observed, as suggested in figures D.1 and D.2 of appendix D.1. The already obtained discriminating power will however be shown to be sufficient and exploiting an additional cuts based on these variables would reject a significant part of the PI, hence they are left aside.

4.2.4 Impact on the $\cos\theta$ distribution

Following the discussion on the various variables to distinguish the DY and PI from the $t\bar{t}$ and WW, table 4.1 summarises the different cuts suggested (these will be referred as “Extra Filters” in histograms).

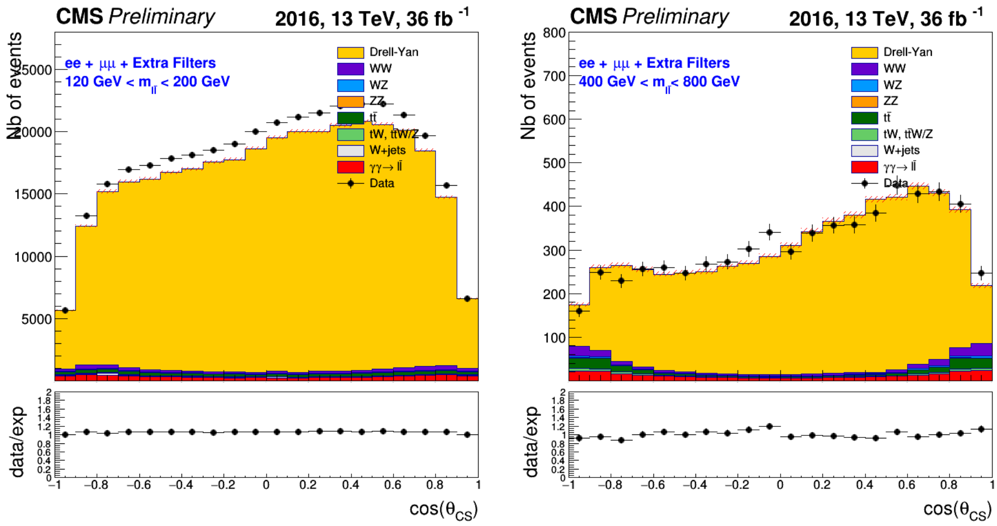
Variable	Criterion
metSig	$\leqslant 10$
number of jets	$\leqslant 1$
number of b-jets	$= 0$

Table 4.1 – The filtering cuts to reduce the $t\bar{t}$ and WW.

The effect of the cuts is represented in figure 4.7, showing the modifications incurred by the $\cos\theta_{CS}$ distribution in two mass bins (corresponding to the unfiltered figures 4.2b and 4.2d). The intense reduction in $t\bar{t}$ and WW is a satisfying outcome of the filtering procedure, though the latter was overall less targeted.

4.3 DY suppression using vertex information

Additional cuts are needed to disentangle the PI from its irreducible background: the DY. It turns out there is a last detail about the hard scattering process that was not exploited: the impact of the interaction on the protons themselves. In a DY interaction, each proton lost a quarkonic parton and are now out of colour-balance, under threat from the stirring grip of QCD. The two protons remnants must necessarily combine with each other to cancel out the colour field created. As a result, many additional low p_T particles are produced and extra tracks are expected to be observed at the interaction vertex. This is not the case for the PI process, where colour-neutral photons are extracted from the protons that thus do not necessarily break down. They can either remain intact, get excited or explode. If both protons remain intact (elastic scattering), no additional activity is expected at the vertex. This is in fact also true if one or both protons perish: due to the absence of colour exchanges between the protons, the decay products are expected to be very forward and therefore outside the tracker acceptance.

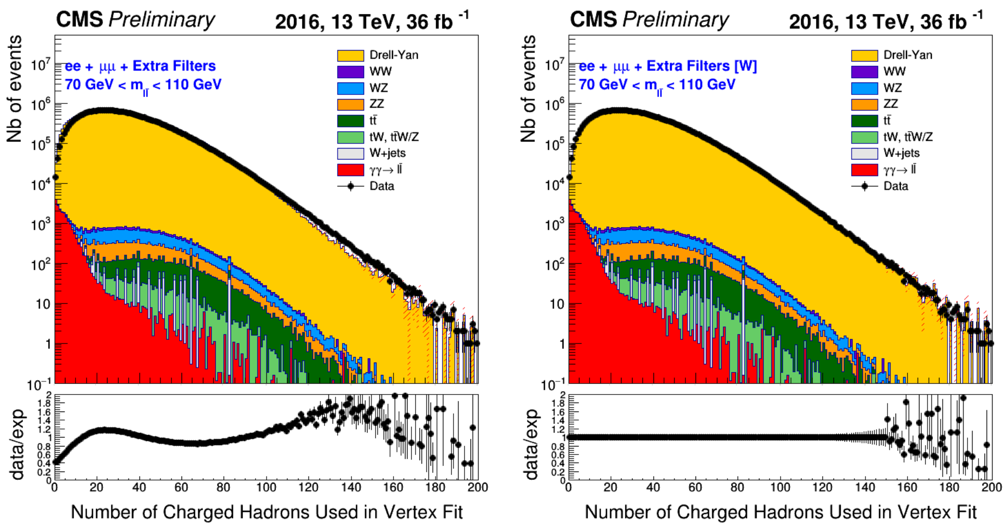


(a) *Mass bin [120, 200] GeV.* (b) *Mass bin [400, 800] GeV.*

Figure 4.7 – The filtered $\cos\theta_{CS}$ *distributions for ee/μμ dilepton pair events (2016 data) for different mass bins.*

4.3.1 The number of charged hadrons associated to the vertex

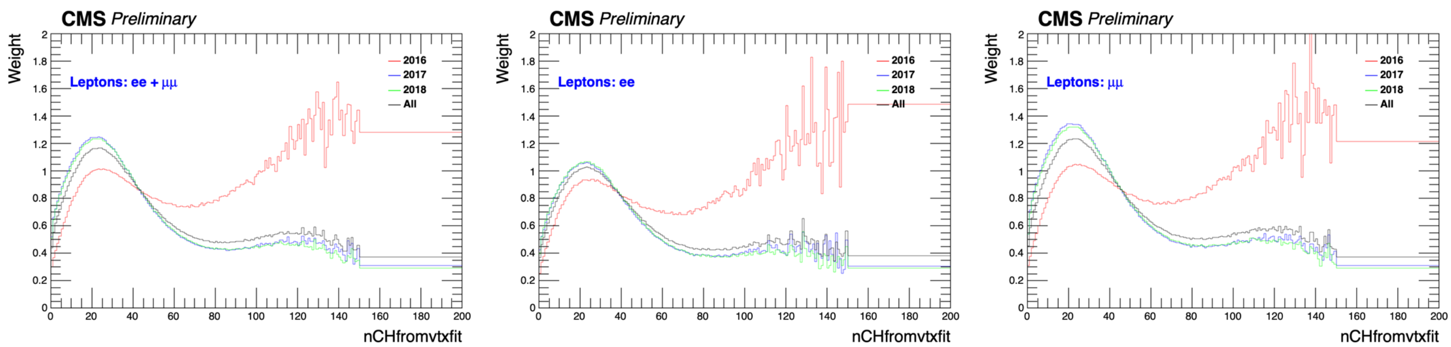
An experimentally accessible variable that can help distinguishing the two processes is therefore the number of charged hadrons (nCH). This variable should exhibit a very different pattern between colour-neutral processes like the PI and those awakening chromodynamical effects. It is however easily corrupted by pile-up events that wrongfully allocate hadrons to the vertex of the hard scattering process. This undesired imprecision can be corrected by restraining the hadron counting to those matched to the right vertex, as derived in section 2.3.5. The thus modified variable is named "number of charged hadron from the vertex fit" and will be abbreviated indiscriminately by nCH or nCHfromvtxFit in the next parts. Figure 4.8a unveils its distributions after filtering for the [70, 110] GeV mass bin.



(a) *Before the nCH weights.* (b) *After the nCH weights.*

Figure 4.8 – The filtered number of charged hadrons from the vertex fit distribution for the [70, 110] GeV mass bin (2016 data).

This variable is difficult to simulate as indicated by the significant deviations between data and expectations in the region of interest (low nCH). Several experimental and theoretical reasons could explain this, such as the non perfect description by the simulations of the underlying event or of the tracking and vertexing efficiency and fake rate. Since a good description of the nCH distribution in the DY is crucial to this analysis, correction factors are derived to make the data and simulations agree in the Z peak region [70, 110] GeV, where the PI signal contamination is small (figure 4.8b). The extracted weights are then applied to all mass bins. They are computed separately for each lepton flavour ($ee/\mu\mu$) and each year of data taking, with a single weight derived to cover the region exceeding 150 charged hadrons. They are shown in figure 4.9.



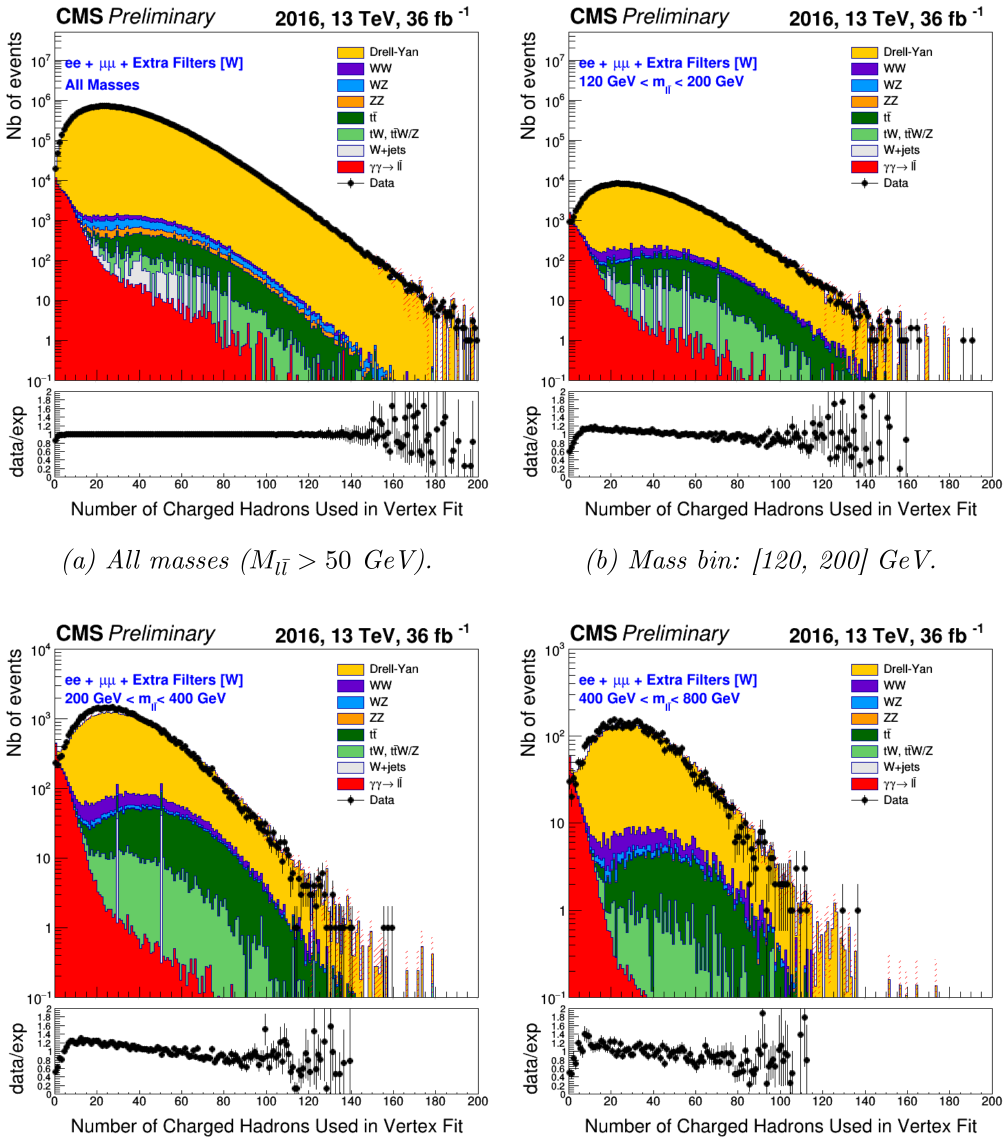
(a) $ee/\mu\mu$. *(b) ee.* *(c) $\mu\mu$.*

Figure 4.9 – The number of charged hadrons correction weights for all years and different lepton flavours.

While 2017/2018 weights are relatively similar, 2016 weights are however significantly different. This is thought to be due to two effects. Firstly, an important upgrade of the pixel detector took place between 2016 and 2017, with the addition of an extra pixel layer. Secondly, as discussed in section 3.4.3, 2016 and 2017/2018 simulations use a different tune to describe the underlying event. The different weights observed between the ee and $\mu\mu$ channels however is still to be understood. A possible explanation could come from the different impact of electron and muon tracks in the vertex reconstruction and fit. Applying the prescribed weights modifies the histograms to those displayed in figure 4.10, where the "[W]" witnesses the nCH scaling procedure. The simulation of the DY, in the region where it dominates, appropriately describes the data as indicated by the satisfying data to expectation ratio at large number of charged hadrons. Simulations in the region of most significance for the PI present however a deficit with respect to the data.

4.3.2 Impact of the pile-up and underlying event

Although the PI does peak distinctively at nought charged hadrons, an extended tail is surprisingly observable. There is a factor 100 between twenty and nought charged hadrons in the nCH distribution at all masses, as shown in figure 4.10a. What could be an additional source of charged hadrons for the PI, differing from the hard scattering process at the main vertex? Two candidates generating these hadrons are investigated: the colour string between proton remnants and the pile-up. Little is indeed known concerning the calibration required for modelling the colour interactions of the protons in the subtle PI process. This effect may lead to such a large number of hadrons that any imprecision in the underlying event is drastically felt in the nCH distributions.



(a) All masses ($M_{l\bar{l}} > 50$ GeV).

(b) Mass bin: [120, 200] GeV.

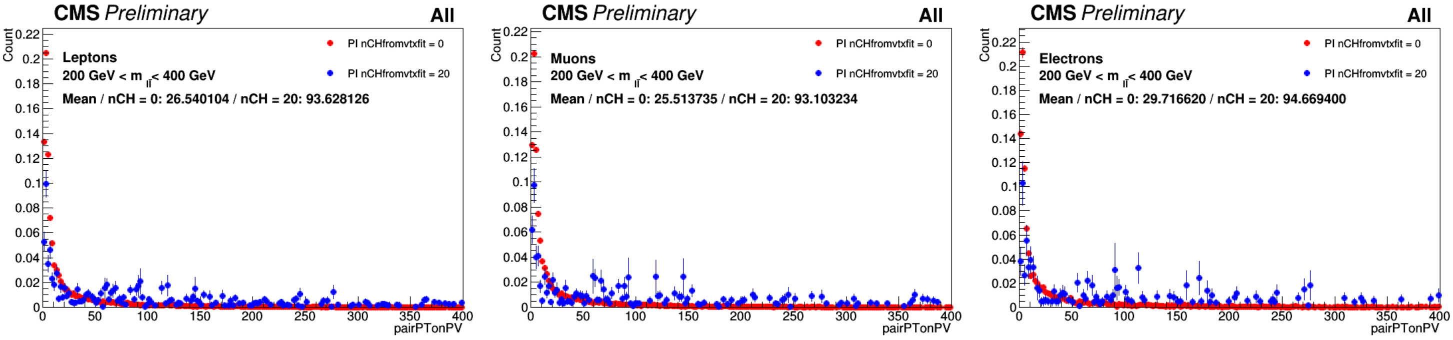
(c) Mass bin: [200, 400] GeV.

(d) Mass bin: [400, 800] GeV.

Figure 4.10 – The nCH-weighed filtered number of charged hadrons from the vertex fit distributions for different mass bins (2016 data).

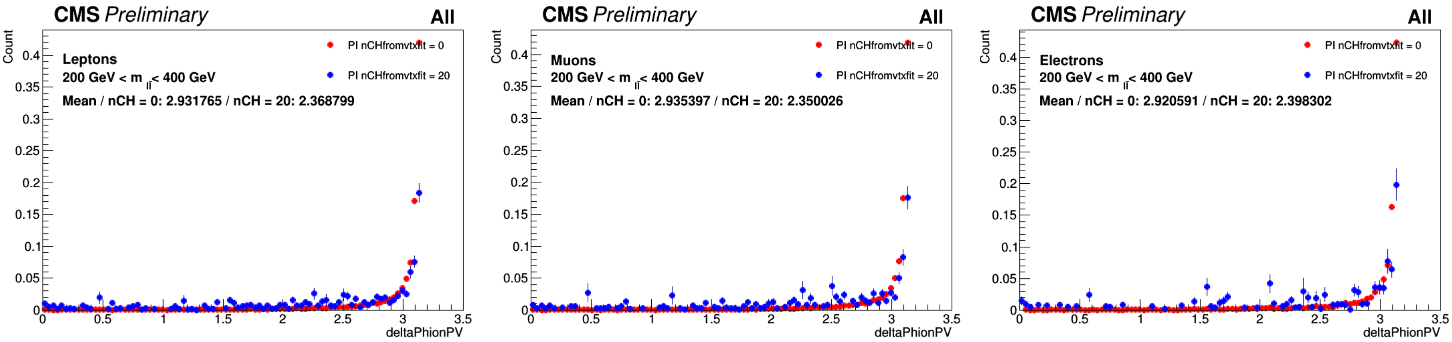
The role of the underlying event in the nCH generation can be explored by looking at the kinematical variables of the lepton pair. Figure 4.11 displays the lepton pair p_T for nought and twenty charged hadrons. The difference in mean between the nought and twenty charged hadrons indicates a kinematical effect is indeed generated in the events leading to a higher number of hadrons, where the p_T of the pair is larger. Figure 4.12 also points out this effect by peaking into the $\Delta\phi$ distribution between the two leptons. The same conclusion arises: $\Delta\phi$ is closer to π for a low value of nCH. This does suggest a role of the colour charge in the population of nCH obtained at high value of the variable.

Another effect, the pile-up, is also expected to play a significant role. The PU contributes in two ways to the nCH population. A trace belonging to a pile-up vertex can be wrongfully associated to the main one, hence increasing the number of hadrons in the fit. In large PU configuration, many vertices are expected in the event, hence a large number of reconstructed vertices (nPV) with a certain probability for these pile-up vertices to be merged with the right one. The former effect would slightly



(a) ee/μμ. *(b) μμ.* *(c) ee.*

Figure 4.11 – The lepton pair p_T for nought (red) or twenty (blue) charged hadrons in vertex fit (Run II data) in the [200, 400] GeV mass bin.

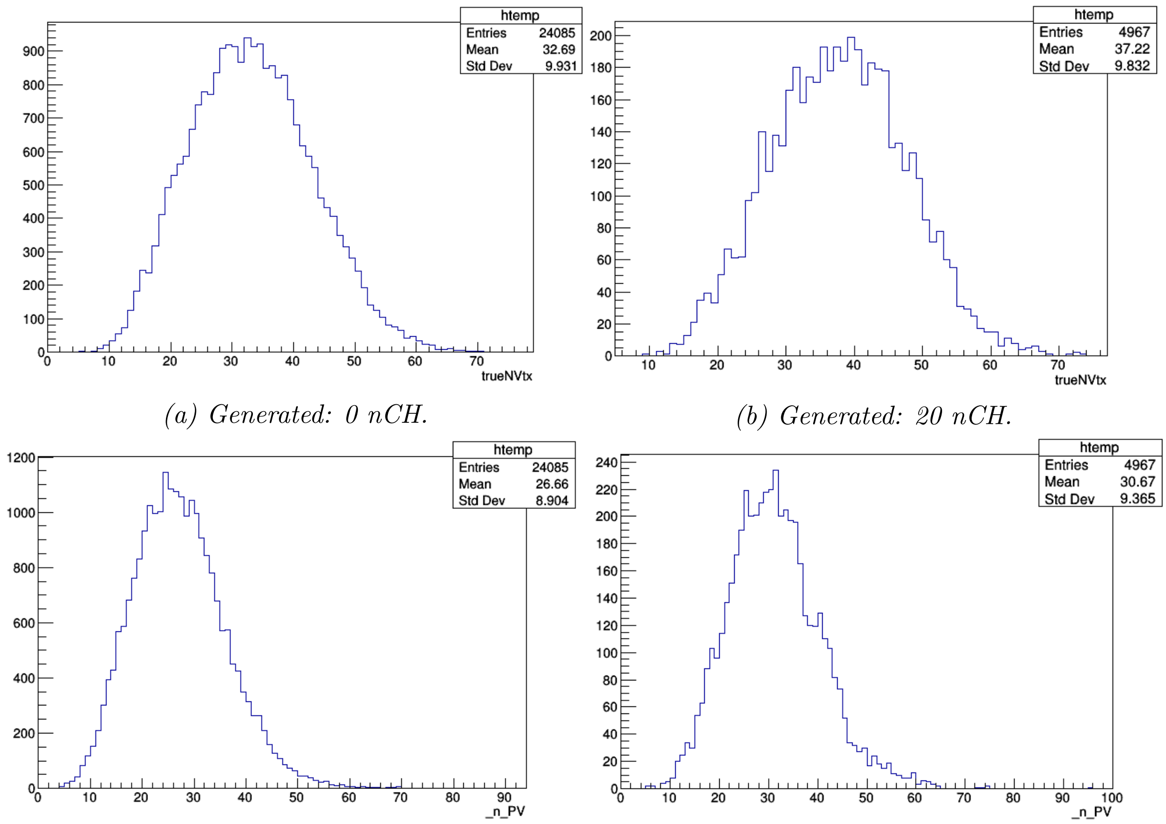


(a) ee/μμ. *(b) μμ.* *(c) ee.*

Figure 4.12 – The $\Delta\phi$ between leptons for nought (red) or twenty (blue) charged hadrons in vertex fit (Run II data) in the [200, 400] GeV mass bin.

increase the number of hadrons while the latter process could explain why some large nCH values can be obtained. The plots in figure 4.10 indeed suggest the existence of two different slopes in the decrease of the nCH distribution for the PI. A fast drop at low values, probably explained by single traces being wrongfully fitted as well as the colour effect, and a slowly decreasing plateau at high nCH values that could find its origin in wrongfully merged vertices. This last effect could be verified by comparing the origin points of the lepton pair to the vertex they are matched to, as a mismatch between the two points could be due to the reconstructed vertex actually resulting from the merging of two slightly offset real ones. This indicator is unfortunately not available in the simulations exploited here and the only available evidence consist in figure 4.13, showing the number of generated and reconstructed vertices for the PI at two different number of charged hadrons.

Two remarks are notable on these plots. Firstly, the number of vertices of both kinds increases when considering larger nCH values, so the effect of the pile-up is apparent. Secondly, when going from the generated to the reconstructed level, a similar fraction of vertices are lost by either not being reconstructed or being merged. This offers some ground for the merging vertex impact on the nCH distribution theory. Another argument for the pile-up effect is suggested by figure 4.14, indicating the nCH distributions for different reconstructed vertex bins.



(a) Generated: 0 nCH. *(b) Generated: 20 nCH.*

(c) Reconstructed: 0 nCH. *(d) Reconstructed: 20 nCH.*

Figure 4.13 – The number of generated vertices (above, trueNVtx) and reconstructed ones (below, _n_PV) for nought (left) and twenty (right) charged hadrons in the PI simulation of the mass bin [200, 400] GeV.

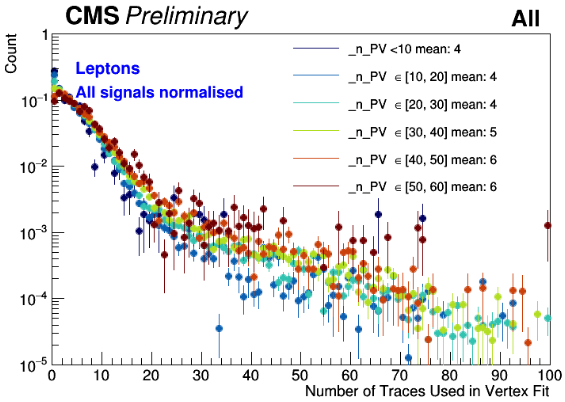
Evidently, the average nCH increases with the number of reconstructed vertices considered, another signature of the pile-up. Electrons have on average more hadrons contributing to the fit, a natural consequence of the difficulties to reconstruct their trajectory compared to muons, hence a lower fit quality and more fake traces being matched. For high number of reconstructed vertices bins, the nCH distributions of the PI peak at a value of one or two, not zero anymore. Hadrons are clearly wrongfully being assigned to the main vertex as this is more likely to happen when many vertices are present. In addition, there is a tendency for the plateau to be higher for the many-reconstructed vertices scenarios, a sign of the merging vertex hypothesis.

The individual impact of each contribution is difficult to establish based on the available variables and a larger scope study could explore the origin of the extra hadrons by searching the MC event output files (that would offer better insights into the role of colour) and by switching off the pile-up perturbation in the simulations to offer some quantitative conclusions.

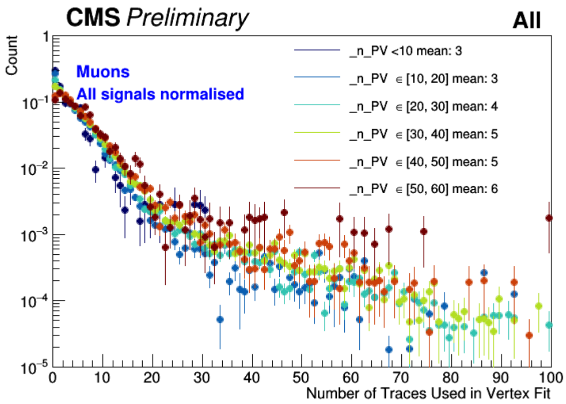
4.4 The $\cos\theta$ distribution at low number of extra traces

The impact of a nCH < 5 and a nCH < 2 cuts on the $\cos\theta_{CS}$ is displayed in figure 4.15. Note that although there is an evident mismatch between data and expectations for certain mass bins, there is already a notable improvement compared to the unweighed ones, shown in appendix D.2.

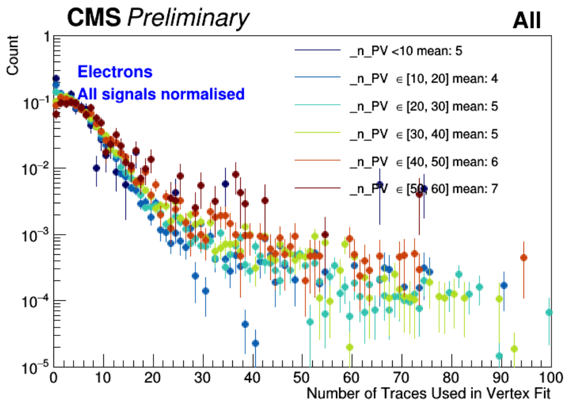
The backgrounds have been satisfyingly removed from the picture with the PI now dominating several distributions, the DY surviving significantly only in the Z peak region. These distributions can



(a) ee/μμ.



(b) μμ.



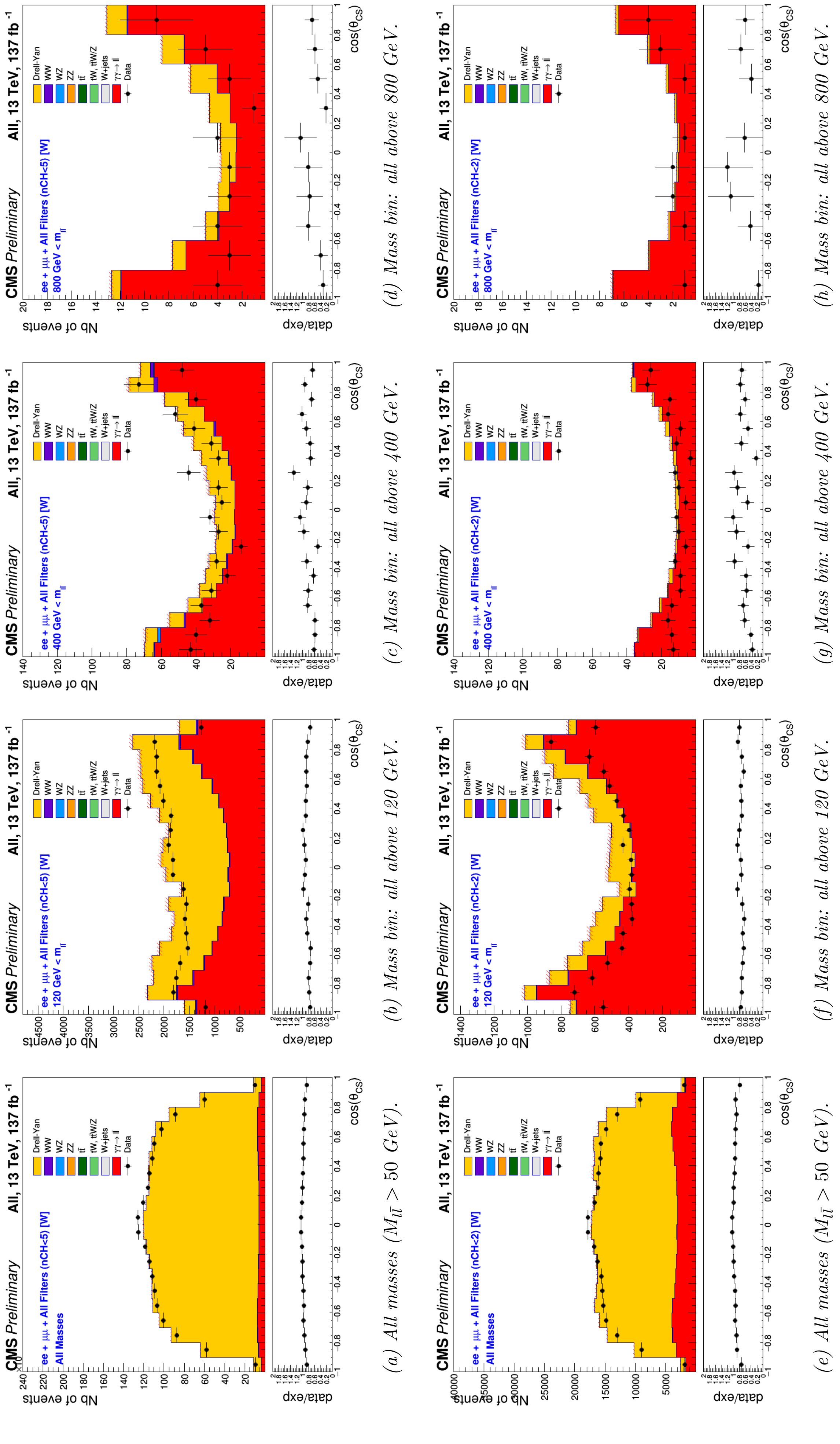
(c) ee.

Figure 4.14 – The (normalised) nCH distributions of the PI for different bins of number of reconstructed vertices (Run II data) in the [200, 400] GeV mass bin.

be fitted thanks to the difference in behaviour of the PI and DY $\cos\theta_{CS}$ distributions. The PI selection improves by a more demanding limitation of nCH < 2, as shown in the second column of figure 4.15.

Summary of the Chapter

This chapter presents a set of cuts and weights to isolate the PI from backgrounds at the reconstructed level. The significance of the missing transverse energy and the number of jets and b-jets in the event are criteria used to suppress the $t\bar{t}$ and the WW. The number of charged hadrons in the vertex fit can then be used to favour the PI signal over the DY. The chapter concludes with a discussion on the quality of the nCH distributions, and the probable PU and underlying event origins of larger than expected values for the signal. Finally, the resulting $\cos\theta_{CS}$ distributions after all cuts are displayed.



(a) All masses ($M_{l\bar{l}} > 50$ GeV).

(b) Mass bin: all above 120 GeV.

(c) Mass bin: all above 400 GeV.

(d) Mass bin: all above 800 GeV.

(e) All masses ($M_{l\bar{l}} > 50$ GeV).

(f) Mass bin: all above 120 GeV.

(g) Mass bin: all above 400 GeV.

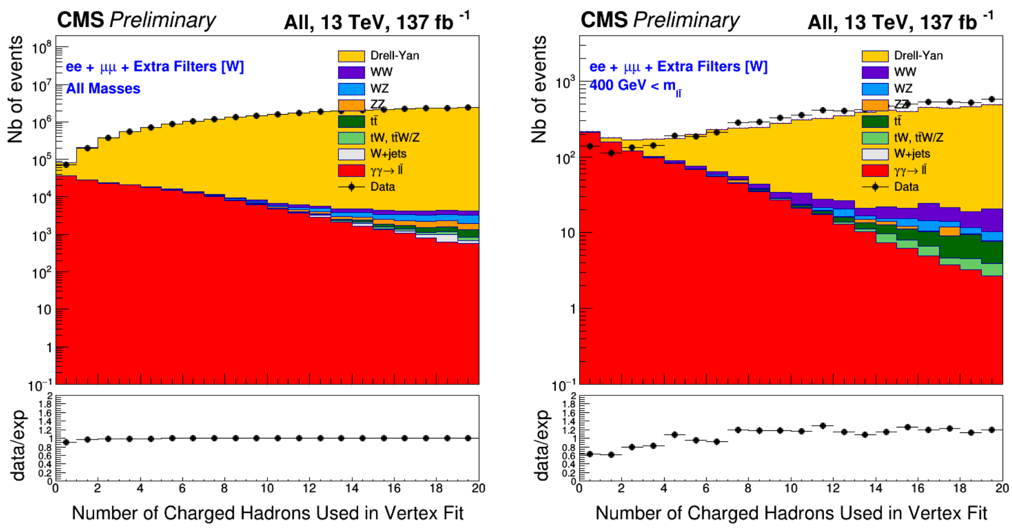
(h) Mass bin: all above 800 GeV.

Figure 4.15 – The nCH-weighed filtered $\cos\theta_{CS}$ distributions with $nCH < 5$ (a to d) and $nCH < 2$ (e to h) (Run II data).

CHAPTER 5

Study of the PI

To summarise the current situation, the PI signal has been proven to be favoured in a region demanding a cut on missing transverse energy, no b-jet and a limited number of jets. To even more disentangle the process from the backgrounds, the number of charged hadrons appearing in the vertex fit, as shown in figure 5.1, suggested a further step by taking a low value of the variable, section 4.4. Indeed, the PI peaks near nought charged hadron given the relative stability of the protons in the underlying event, the larger picture forming the stage for the hard scattering.



(a) All masses. *(b) Mass bin: all above 400 GeV.*

Figure 5.1 – The nCH-weighed filtered number of charged hadrons from the vertex fit distributions for two different mass bins (Run II data).

This chapter starts with a procedure to adjust the relative contributions of the DY and PI using a fit of the filtered distributions to the data. It then exploits the derived corrections to adjust the cross sections. To conclude, an introduction to a BSM analysis is suggested by scanning the invariant mass spectrum in the derived region favouring the PI.

5.1 Fitting the histograms: a walk to the cross section corrections

Having isolated two regions with a large fraction of photon-induced lepton pair production, sometimes even in dominating status, a fitting procedure can be applied to the histograms to extract the resulting PI and DY contributions. The strategy for the fit relies on having access to samples where the topologies of both processes are distinguishable. The simulations are perceived as two pieces covering the bins, a sort of basis, that can be arranged by a certain scaling to explain most of the data. The optimal scalings are the result of a maximum likelihood estimation (MLE). The likelihood of a histogram is classically defined as:

$$\mathcal{L} = \prod_{i=1}^{n} p_i(N_{\text{obs}_i}; N_{\text{exp}_i}) \tag{5.1}$$

where n is the number of bins, N_{exp_i} is the expected number of events in bin i and is set to $N_{\text{exp}_i} = \mu_i$, N_{obs_i} is the number of observed events in bin i, and p_i follows a Poisson law of parameter μ_i:

$$p_i(N_{\text{obs}_i}; \mu_i) = \frac{e^{-\mu_i}\mu_i^{N_{\text{obs}_i}}}{N_{\text{obs}_i}!} \tag{5.2}$$

Knowing the expected value, one can formulate the probability to get a certain distribution over the bins of the histogram. The methodology pursued here is the reverse of this statement. Knowing the distribution of the data over the bins, what expected value constitutes the most likely scenario? Writing the expectation as:

$$\mu_i = \alpha_{PI}\ \mu_{PI,i} + \alpha_{DY}\ \mu_{DY,i} + K_{BG}, \tag{5.3}$$

the PI, DY and the leftover backgrounds are parametrically set to constitute the expectation. The parameters α_{PI} and α_{DY} are global (bin-independent) weights to be returned by the fitting procedure. In the current analysis, the other backgrounds are either ignored, $K_{BG} = 0$, or merged with the DY when relevant (in which case α_{DY} is understood to stand for α_{BG}). The fitting procedure consists in finding the parameters maximising the likelihood or, given the numerical setting of the problem, minimising $-\log\mathcal{L}$.

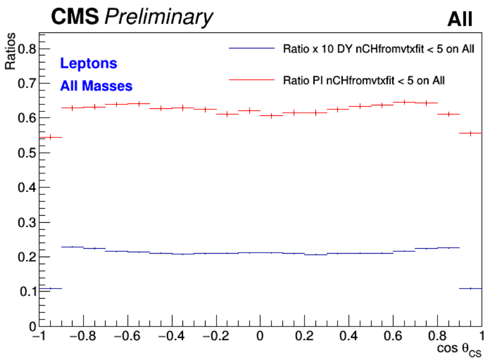
The problem is computationally solved thanks to the MINUIT package of the ROOT distribution [54, 55]. MINUIT consists in a tool designed to find the minimum value of multi-parameter functions. More specifically, the distribution includes methods to study the function being analysed around its minimum, such as incertitude estimation [56] and error control in statistical studies [57]. The specific designs of its methods are centred around χ^2 and log-likelihood to compute, among many options, the best-fit estimators. This is exactly the problem at hand and a dedicated program was written to implement the method for both the $\cos\theta_{CS}$ and nCH distributions. The former is well understood from both the theoretical and experimental vantage point while the latter is far more audacious considering the difficulty to simulate appropriately the variable.

5.1.1 The $\cos\theta_{CS}$ fit

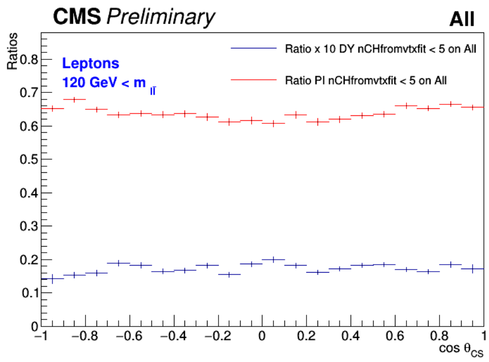
Two methods are employed to fit the $\cos\theta_{CS}$ distributions. One is designed for the version including a cut on the number of charged hadrons and the other one directly applies to the resulting distributions after the filters of table 4.1. Since the statistics in the DY simulation is reduced for the low nCH region, the DY distribution with a nCH cut can be replaced by the one without any after rescaling the latter by a bin-independent weight corresponding to:

$$\frac{DY\ with\ no\ nCH\ cut}{DY\ with\ the\ nCH\ cut}, \tag{5.4}$$

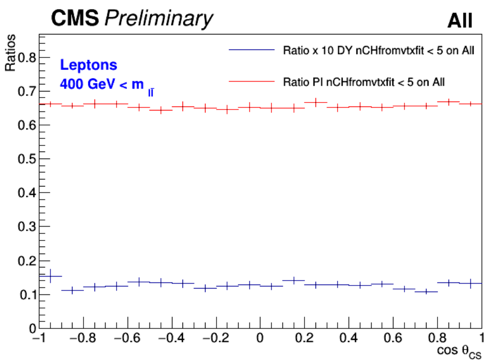
so that the same overall intensity is conserved. This is a valid translation because the $\cos\theta_{CS}$ distributions are mostly shape-independent regarding the nCH variable: whichever nCH binning is considered, the $\cos\theta_{CS}$ should not be impacted. This statement is demonstrated in figure 5.2, displaying the ratio of the number of events with nCH < 5 over the total, as a function of $\cos\theta_{CS}$. A flat ratio versus $\cos\theta_{CS}$ is indeed seen for the DY and the PI, except in the first mass bin.



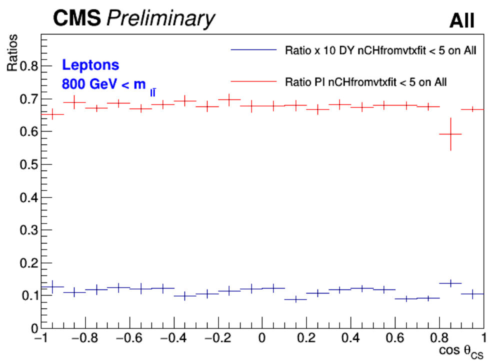
(a) All masses.



(b) Mass bin: above 120 GeV.



(c) Mass bin: above 400 GeV.

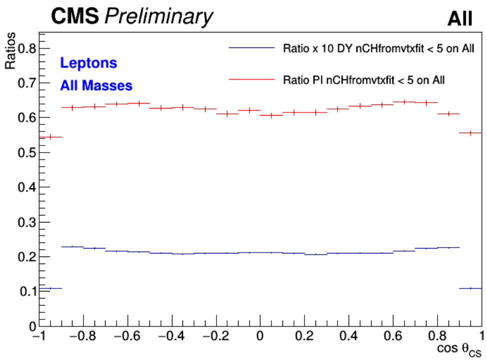


(d) Mass bin: above 800 GeV.

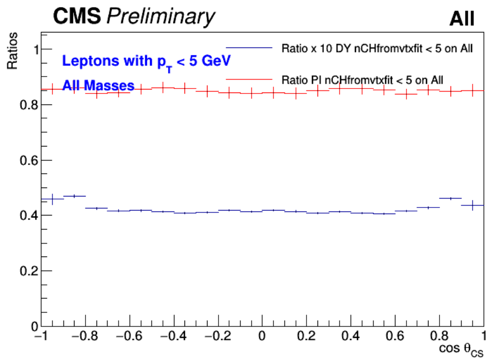
Figure 5.2 – The ratios nCH < 5 on no cut of $\cos\theta_{CS}$ distributions for the DY (multiplied by 10) and PI at various mass bins(Run II data).

The exception for the mass bin including all masses is found to originate from the Z boson mass region. It is understood to be a dynamical effect from the cuts used in the analysis. Indeed, only the bins bordering the distribution seem to fall short of the plateau. In the mass region under consideration, the lepton pair has on average an invariant mass of $M_{l\bar{l}} \approx 90$ GeV (the Z boson resonance). The bins where the hypothesis seems to fail correspond to an absolute value of $|\cos\theta_{CS}| \approx 0.9$, hence a $|\sin\theta_{CS}| \approx 0.44$. Leptons in these regions have a typical $p_T \approx \frac{M_{l\bar{l}}}{2}\sin\theta \approx 19.6$ GeV. Now, among the various triggers and cuts demanded in the analysis, the leading lepton in the ee and $\mu\mu$ cases is required to possess a minimal p_T of 25 GeV. This is a demanding cut for leptons in the bordering region of the $\cos\theta_{CS}$ in the Z boson peak mass bin and many fall short of this requirement (hence the bell shape of the $\cos\theta_{CS}$ distributions).

How does this effect depend on nCH? In large number of charged hadrons events, it is likely a jet (from ISR/NLO QCD) may have occurred in the interaction. This unbalance in momentum boosts the lepton pair in a certain direction. Hence, for example, an initially almost null p_T pair with both initial leptons having p_T of 19.6 GeV can be replaced by a boosted pair with lepton p_T of 15 and 25 GeV. This boosted pair would manage to pass the filters (demanding at least 15 GeV for the trailing lepton and 25 GeV for the leading one), while the original configuration would fail. This is indeed the case as proven by figure 5.3, showing the same ratio with and without a dilepton pair $p_T < 5$ GeV cut and the effect is therefore understood to be dynamical and favouring the bordering $\cos\theta_{CS}$ in cases of larger nCH population, hence the decreasing ratio at the extremities.



(a) No p_T cut.



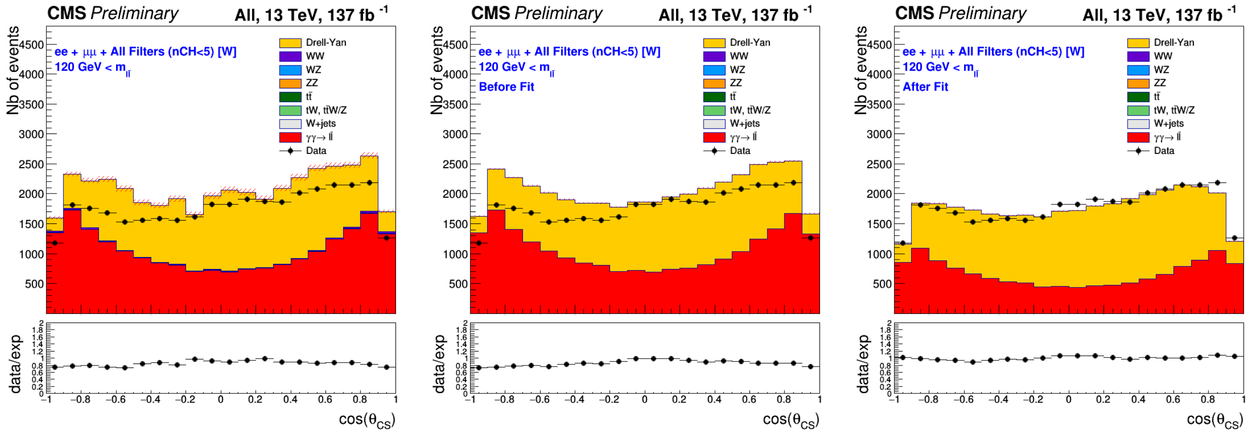
(b) $p_T < 5$ GeV.

Figure 5.3 – The ratios nCH < 5 on no cut (left) and idem with a lepton pair $p_T < 5$ GeV cut (right) of the $\cos\theta_{CS}$ distribution for the DY (multiplied by 10) and PI at all masses (Run II data).

Fortunately, this complication of the shape-independence hypothesis of the $\cos\theta_{CS}$ in the Z boson mass region is not dramatic. In any cases, the fit will perform poorly in this vastly DY dominated region as the topological differences with the PI is rendered moot by the shear intensity of the Drell-Yan. The extrapolation therefore does prove interesting in the most relevant mass bins as the thus translated DY has more statistics and is hence less subject to statistical fluctuations. Figure 5.4 displays the $\cos\theta_{CS}$ distributions before and after the fit in the above 120 GeV mass region with nCH < 5, all years combined and all lepton flavours. Clearly, the DY distribution is smoother after translation and the statistical errors of the MC are reduced.

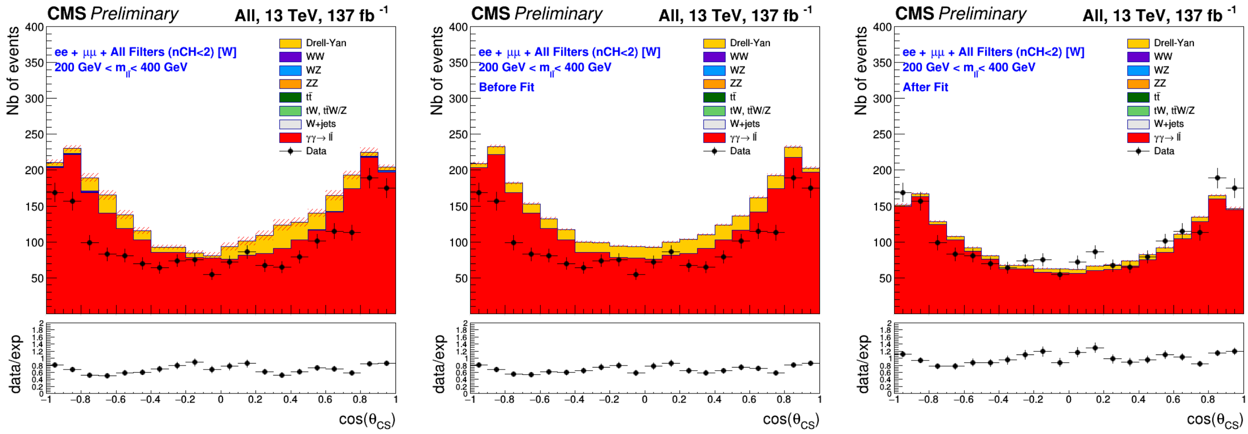
Figure 5.5 shows the same information for the [200, 400] GeV mass region with nCH < 2. Note how in figure 5.5c most of the data can directly be understood as emanating from the PI process, an important direct observation of a process that is still not fully understood in hadron colliders. More complete fit results for each parameter are displayed in appendix figure D.6 (the plotting format is introduced in section 5.1.2).

Finally, the $\cos\theta_{CS}$ distributions with no charged hadron cuts can also directly be fitted. In their case, the remaining backgrounds are joined with the DY so that a common weight correction is given to all backgrounds. Figure 5.6 displays, for example, the $\cos\theta_{CS}$ fitted histograms in the [400, 800] GeV mass region. Interestingly, and as opposed to all other results detailed later, the PI fitted normalisation here is found to be greater than 1. This result should however be taken with caution as the fit is completely driven by the two extreme bins, where other backgrounds than the DY are still relevant.



(a) Original distributions. *(b) Before fit [DY translated].* *(c) After fit [DY translated].*

Figure 5.4 – The nCH-weighed $\cos\theta_{CS}$ distributions in the above 120 GeV mass region with nCH < 5 (Run II data).



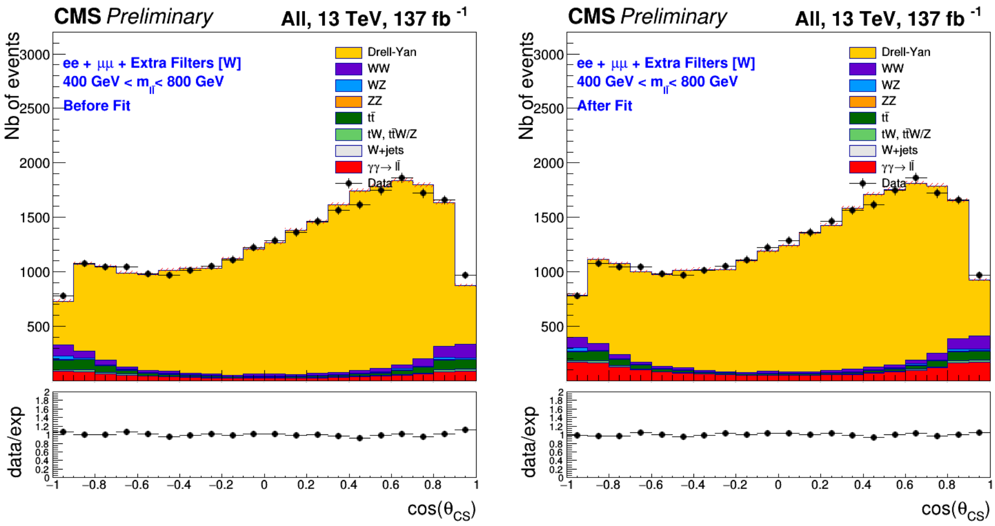
(a) Original distributions. *(b) Before fit [DY translated].* *(c) After fit [DY translated].*

Figure 5.5 – The nCH-weighed $\cos\theta_{CS}$ distributions in the [200, 400] GeV mass region with nCH < 2 (Run II data).

5.1.2 The nCH fit

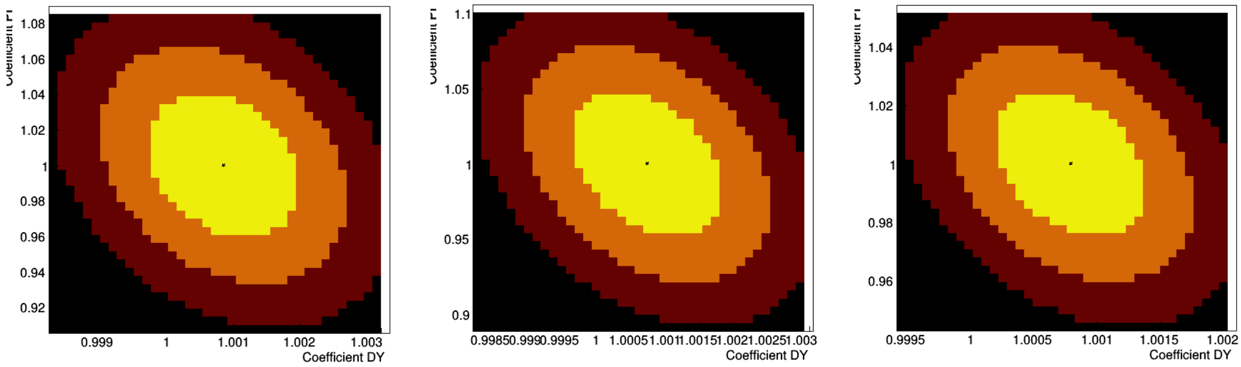
The nCH distributions offer another channel to fit the PI and DY to the data though the quality of the simulations is not assured due to the incertitude regarding the MC. It also gives a natural control set for the precision of the fitting method. Indeed, per-bin weights were derived to rescale the DY in the [70, 110] GeV mass region so that the DY and PI perfectly explain the data in that bin. Hence, fitting them in the [0, 10] nCH region should output weights of 1. Figure 5.7 shows this by indicating the values of the two parameters obtained from the MLE computation with a star and a colour-code revealing the confidence interval of the prediction: 1 σ (68.27 % confidence level) is indicated in yellow, 2 σ (95.45 %) in orange, and 3 σ (99.73 %) in maroon.

Clearly, the fitted values obtained are 1 for both the DY and PI. The DY is much more constrained than the PI given its dominating status. Note how the fitted intervals concentrate on 1 with the years, an effect due to the evolution of available data matching the increase in luminosity along the years. The fitting procedure is thus validated by the obtained results. An example of a fitting chain is displayed



(a) Before fit. *(b) After fit.*

Figure 5.6 – The nCH-weighed $\cos\theta_{CS}$ distributions in the [400, 800] GeV mass region (Run II data).



(a) 2016 data. *(b) 2017 data.* *(c) Run II data.*

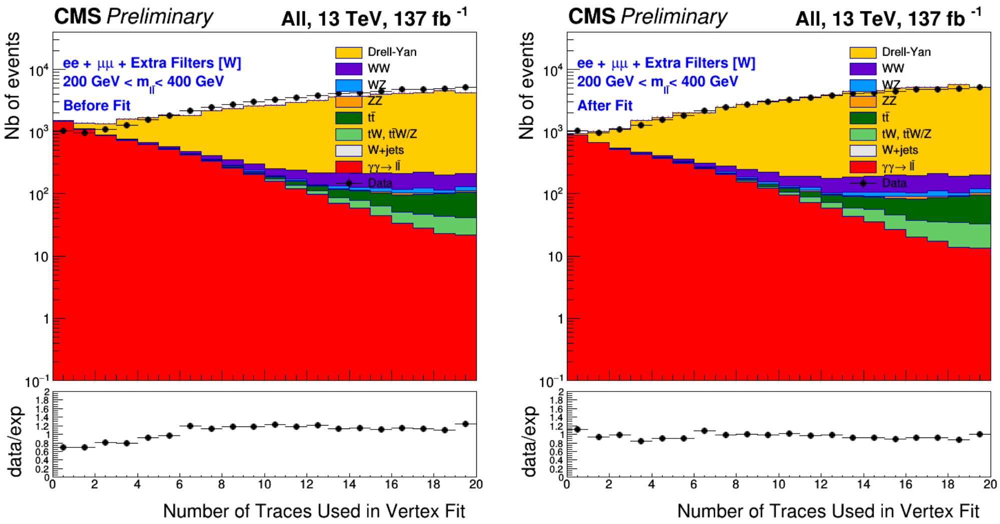
Figure 5.7 – The parameter fitted in the control region [70, 110] GeV of the nCH distributions for different years.

in figure 5.8, targeting the [200, 400] GeV mass region with all years combined and all leptonic final states.

Note that the fitting procedure also works here thanks to the distinct topological behaviours of the PI, being a steep decreasing slope from the start, and the DY, progressively increasing with nCH. The PI is essentially fixed by the requirement of the first bin and the DY accommodates the rest of the distribution. Appendix figure D.6 displays some additional fit results on the whole Run II for both parameters and different mass bins.

5.1.3 Global results

In order to assess the stability of the method, three different fits are performed and results are studied separately for each year and lepton flavour, as well as combined. The first (second) fit uses the $\cos\theta_{CS}$ distribution after a tight (loose) selection requiring less than 2 (5) nCH at the vertex. The third one is performed directly on the nCH distribution. Results are respectively shown in figure 5.9a, 5.9b and 5.10, where the fitted signal normalisations are shown for various bins of mass defined in table



(a) Before fit. *(b) After fit.*

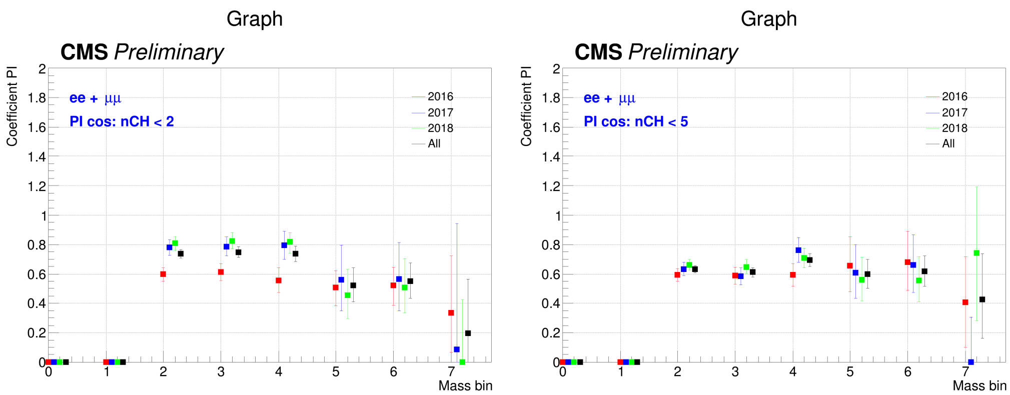
Figure 5.8 – The nCH-weighed nCH distributions in the [200, 400] GeV mass region (Run II data).

5.1. The displayed errors correspond to a 1 σ uncertainty.

Mass Bin Index	Mass Interval
0	All masses
1	[60, 120] GeV
2	Above 120 GeV
3	[120, 200] GeV

Indicator	Mass Interval
4	[200, 400] GeV
5	Above 400 GeV
6	[400, 800] GeV
7	Above 800 GeV

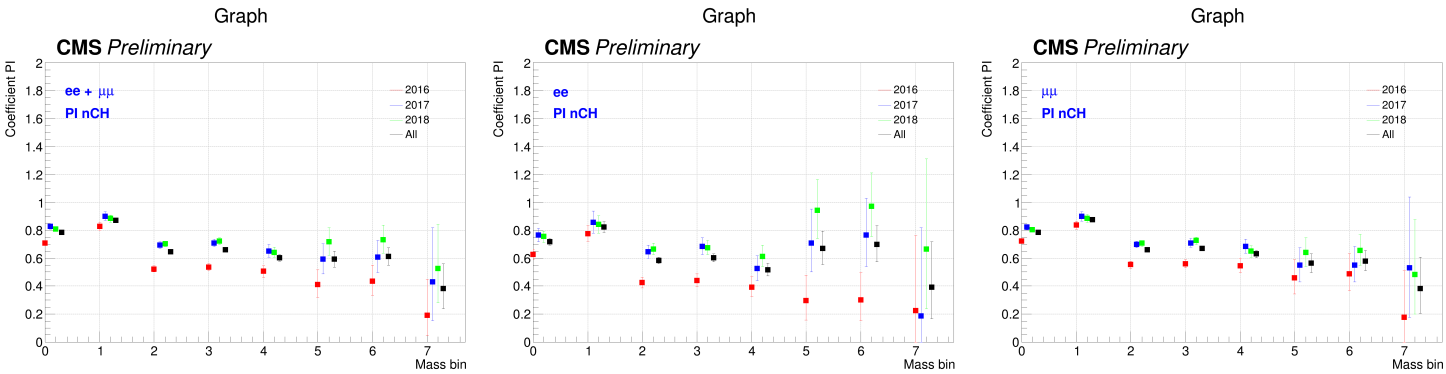
Table 5.1 – The different mass intervals and their indicators.



(a) nCH < 2. *(b) nCH < 5.*

Figure 5.9 – The PI fitting weights for the $\cos\theta_{CS}$ distributions for all mass bins (except 0 and 1), all years.

The fits to the nCH filtered $\cos\theta_{CS}$ distributions enjoy a reassuring stability, all suggesting an average correction of 0.6 for the PI simulations. Additional results are available in appendix D.3 with a fit on the $2 \leq$ nCH < 5 (figure D.5a) and one on the $\cos\theta_{CS}$ with no nCH cut (figure D.5b). Also in the appendix, figure D.7 splits the fit results for ee and $\mu\mu$ separately. These weights are in agreement with the nCH fit results that also suggest an average correcting weight of 0.6 for the mass bin [200, 800] GeV. The year 2016 globally appears to be lower than 2017/2018, maybe due to the pixel detector change.



(a) ee + μμ. *(b) ee.* *(c) μμ.*

Figure 5.10 – The PI fitting weights for the nCH distributions for all years and different lepton configurations.

To conclude the analysis, table 5.2 summarises the set of weights derived for the various mass windows and fitting procedures for the Run II data. Note that the weights for ee and $\mu\mu$ are available in table D.1.

	$ee/\mu\mu$		
Mass Interval	$\cos\theta_{CS}$ [nCH $<$ 2]	$\cos\theta_{CS}$ [nCH $<$ 5]	nCH
$[60, 120]$ GeV	*excluded*	*excluded*	0.87± 0.01
$[120, 200]$ GeV	0.75 ± 0.03	0.61 ± 0.03	0.66 ± 0.01
$[200, 400]$ GeV	0.73 ± 0.05	0.60 ± 0.04	0.60 ± 0.02
$[400, 800]$ GeV	0.55 ± 0.13	0.61 ± 0.11	0.61 ± 0.06
Above 800 GeV	0.19 ± 0.37	0.42 ± 0.31	0.38 ± 0.18

Table 5.2 – The different mass intervals and the corrections suggested by the various fitting procedures (Run II data).

The set of fitted weights extracted from the MLE procedure are summarised by "set 1" for those derived by the nCH $<$ 2 distribution, "set 2" for nCH $<$ 5 and "set 3" for the nCH distributions. Note that the correcting weights seem to agree with a result in [46] displayed in figure 5.11. They define the survival factor $S_{\gamma\gamma}$ as a correction which distance to one quantifies the overestimation incurred by neglecting finite size effects in the proton. This is roughly equivalent to the fitting factors derived in this analysis and the corrections do reasonably agree. They validate the less-than-one corrections obtained for every reasonable situations here, a result also obtained in [23, 58].

Finally, table 5.3 displays the PI cross sections and their corrections. Note that this final point should be endowed with a discussion on systematics and statistical errors that could be the object of further work. Systematics can be roughly estimated by looking at the spread in corrections weights obtained by the three methods, reported in table 5.2, being in the order of 0.1.

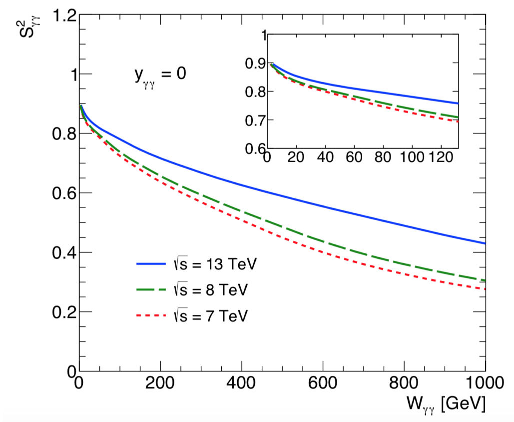


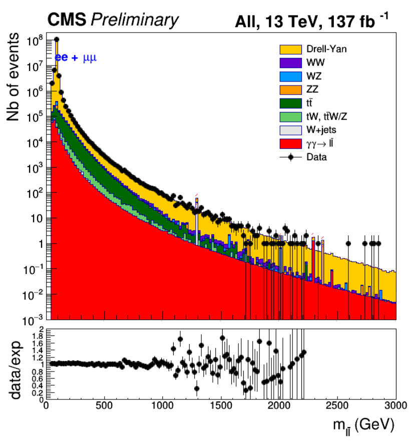
Figure 5.11 – The survival factor $S_{\gamma\gamma}$ at zero rapidity as a function of the photon-photon centre of mass energy $W_{\gamma\gamma}$ [46].

Mass Region	Cross Section [pb]	Weight (± stat. ± syst.)	Weighed Cross Section [pb]
$[60, 120]$ GeV	11.87	$0.87 \pm 0.01 \pm 0.1$	$10.33 \pm 0.12 \pm 1.19$
$[120, 200]$ GeV	1.82	$0.61 \pm 0.03 \pm 0.1$	$1.11 \pm 0.05 \pm 0.18$
$[200, 400]$ GeV	$5.04\ 10^{-1}$	$0.60 \pm 0.04 \pm 0.1$	$(3.03 \pm 0.20 \pm 0.50) \times 10^{-1}$
$[400, 800]$ GeV	$6.22\ 10^{-2}$	$0.60 \pm 0.11 \pm 0.1$	$(3.73 \pm 0.68 \pm 0.62) \times 10^{-2}$
> 800 GeV	$5.89\ 10^{-3}$	$0.42 \pm 0.31 \pm 0.1$	$(2.47 \pm 1.83 \pm 0.59) \times 10^{-3}$

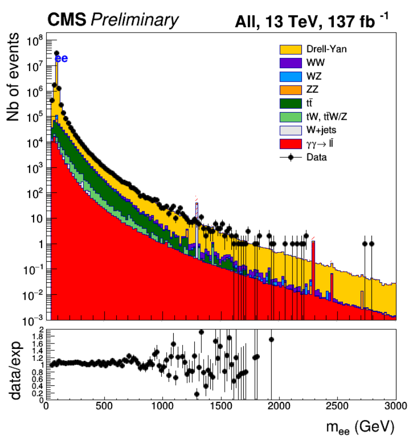
Table 5.3 – The PI cross sections at different masses and their corrections, using set 2.

5.2 Introduction to a BSM search for deviation in high invariant mass

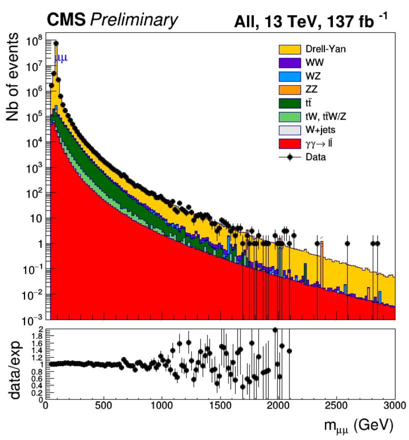
Now that an effective isolation of the PI from other backgrounds has been obtained, one can explore the dilepton invariant mass spectrum to search for any deviation. Figure 5.12 displays the mass spectrum of the entire Run II with reweighing of the MC for Z-mass peak matching.



(a) $ee + \mu\mu$.



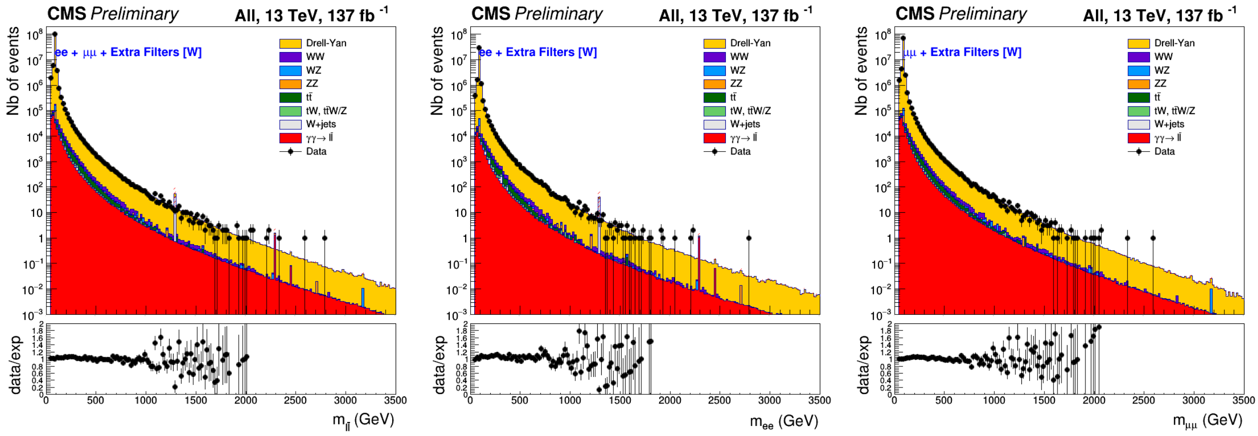
(b) ee.



(c) $\mu\mu$.

Figure 5.12 – The invariant mass spectrum for all years combined and different lepton configurations.

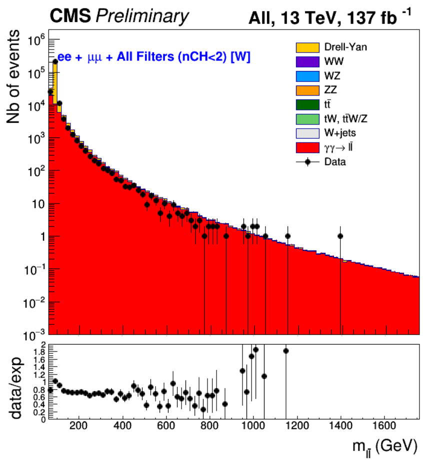
Note that the ratio of the number of dimuon to dielectron events is three to one, due to better identification efficiencies for the former. Applying the filters to remove most of the $t\bar{t}$ and some of the WW leads to figure 5.13.

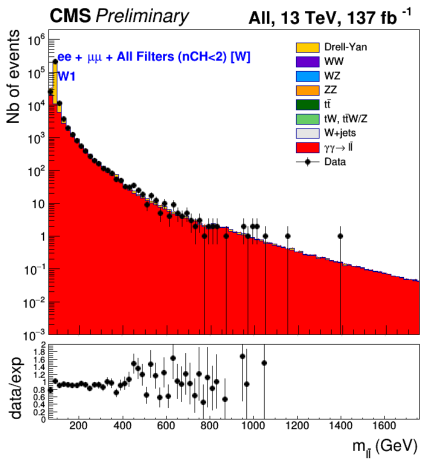


(a) ee + μμ. *(b) ee.* *(c) μμ.*

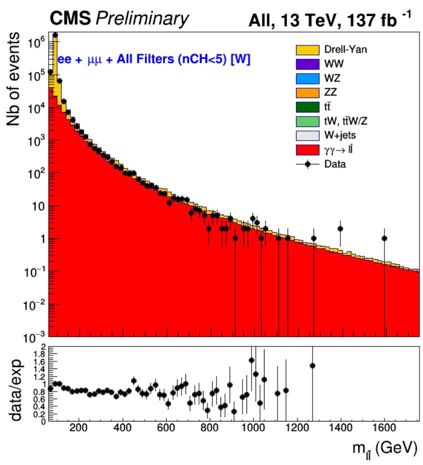
Figure 5.13 – The filtered invariant mass spectrum for all years combined and different lepton configurations.

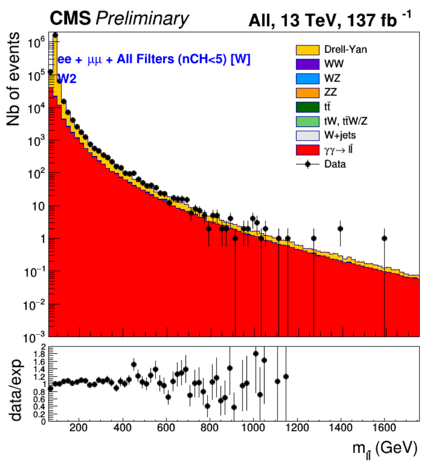
The agreement data/MC is still satisfying. Finally, applying the nCH cuts modifies the spectrum to figure 5.14. The right plots represent the situation after applying the MLE weights from the above section.





(a) No fit weights [nCH < 2]. *(b) Weighed by Set 1 [nCH < 2].*





(c) No fit weights [nCH < 5]. *(d) Weighed by Set 2 [nCH < 5].*

Figure 5.14 – The nCH-weighed filtered invariant mass spectrum with nCH cut (Run II data).

The derived weights from the previous procedure felicitously correct the simulations to match the data at all mass considered. The agreement data/MC is indeed corrected from a slowly decreasing ratio to a flat plateau at one. With the incertitude accounted for, no significant exotic deviation is observed from the expectations at all mass considered, though this would require further steps to be quantified rigorously. Note that the expectations for the nCH < 2 distribution performed worse than for the nCH < 5 version, before MLE weights. This clearly emanates from the difficulties to model the distributions at low charged hadrons number and reminds of the crucial importance this variable plays in the selection of the PI.

To conclude the analysis, table 5.4 indicates the evolution of the number of events after the different cuts both for the data and for the DY and PI, the latter being weighed appropriately at each step. The ratio 3/1 for muons to electrons is apparent from reading the table. The set of cuts managed to bring the PI from a mere 0.18 % of the data to about 20 %. Note also that the vast majority of the surviving DY events are in the Z boson mass peak region. In more massive bins, the share of the data explained by the PI can be drastically larger than in the overall region, reaching even 80 % in a specific sector, as shown in table D.2.

$ee/\mu\mu$					
Cut Type	Data	DY	PI	DY/Data	PI/Data
Original	$1.22\ 10^8$	$1.20\ 10^8$	$2.21\ 10^5$	98.7 %	0.18 %
Filtered	$1.15\ 10^8$	$1.14\ 10^8$	$2.04\ 10^5$	99.4 %	0.18 %
+ nCH < 5	$1.90\ 10^6$	$1.81\ 10^6$	$1.09\ 10^5$	95.0 %	5.71 %
+ nCH < 2	$2.73\ 10^5$	$2.22\ 10^5$	$0.56\ 10^5$	81.4 %	20.6 %

$\mu\mu$					
Cut Type	Data	DY	PI	DY/Data	PI/Data
Original	$8.69\ 10^7$	$8.57\ 10^7$	$1.71\ 10^5$	98.7 %	0.20 %
Filtered	$8.19\ 10^7$	$8.15\ 10^7$	$1.59\ 10^5$	99.5 %	0.19 %
+ nCH < 5	$1.42\ 10^6$	$1.35\ 10^6$	$8.83\ 10^4$	94.7 %	6.19 %
+ nCH < 2	$2.18\ 10^5$	$1.75\ 10^5$	$4.73\ 10^4$	80.4 %	21.7 %

ee					
Cut Type	Data	DY	PI	DY/Data	PI/Data
Original	$3.50\ 10^7$	$3.46\ 10^7$	$5.19\ 10^4$	98.6 %	0.15 %
Filtered	$3.29\ 10^7$	$3.28\ 10^7$	$4.70\ 10^4$	99.5 %	0.14 %
+ nCH < 5	$4.75\ 10^5$	$4.57\ 10^5$	$2.14\ 10^4$	96.1 %	4.51 %
+ nCH < 2	$5.48\ 10^4$	$4.64\ 10^4$	$9.27\ 10^3$	81.4 %	16.9 %

Table 5.4 – The number of events in the data and in the weighed simulations (the latter being further weighfed by set 2 for the nCH cuts).

Summary of the Chapter

In this chapter, the $\cos\theta_{CS}$ and nCH distributions are fitted using a maximum likelihood approach to obtain scaling weights for each process, year and lepton configuration. These corrections, derived for different mass bins, can be employed to update the cross sections. To conclude the analysis, an introduction towards a BSM search is proposed by displaying the dilepton invariant mass spectrum after isolation. No significant deviations are observed within statistical uncertainties and systematics are left for further work.

Conclusion and Outlooks

The analysis presented here is enshrined into a large scale effort to uncover traces of new physics. The limitations of the Standard Model are legion and the plethora of theoretical extensions to englobe these missing elements introduce new elementary particles. Searching for these exotic signals is a major focus of the CMS experiment at CERN. Among the many studies, dilepton final states are particularly privileged and scanned at large invariant masses to find evidence of new physics. No significant deviations from the Standard Model have yet been observed and limits are being placed on the production cross section of BSM processes.

The analysis addressed several aspects. The photon-induced (PI) lepton pair production, starting from two photonic partons of the protons, can be isolated in ee and $\mu\mu$ final states by applying a set of cuts. Demanding a low missing transverse energy significance, no b-jets and at most a single jet proved sufficient to diminish the $t\bar{t}$ and WW background contributions. The irreducible background of the PI signal, the Drell-Yan (DY), can be extinguished by demanding a low number of charged hadrons in the vertex fit. These conditions exploit differences in the physical process at the event level.

The selected signal in the PI simulation is then adjusted to the data and mass-dependent corrections to the theoretical cross section are extracted. This is possible thanks to the different expected behaviours of the PI and DY for both the $\cos\theta$ and nCH distributions. The corrections obtained are in the neighbourhood of 0.6 for the [200, 800] GeV invariant mass bin. They are consistent with recent results from the literature suggesting the PI cross section is overestimated by neglecting finite size effects in the proton.

The analysis concluded with an opening towards a BSM physics search in the PI enriched region. Further work is necessary to properly account for systematic uncertainties arising at all steps. A deeper understanding of the distribution of the number of charged hadrons fitted at the vertex in photon-induced lepton production also requires additional considerations as this variable drastically impact the signal region. The physical interest of this study not only resides in the search for a deviation in this specific channel but also in the better understanding of a SM process that could help constrain incertitude in other searches by improving our knowledge of photon-initiated processes in pp collision, a still unfamiliar affair in the world of hadron colliders.

Bibliography

[1] W.N. Cottingham and D.A. Greenwood. *An introduction to the Standard Model of Particle Physics*. Cambridge University Press, second edition, 2001. (Cited on page 4)

[2] M.D. Schwartz. *Quantum field theory and the Standard Model*. Cambridge University Press, 2014. (Cited on pages 4, 8, and 9)

[3] M.E. Peskin and D.V. Shroeder. *An introduction to quantum field theory*. Perseus Books, 1995. (Cited on pages 4 and 11)

[4] F. Englert and R. Brout. Broken symmetry and the mass of gauge vector mesons. *Phys. Rev. Lett.*, 13:321–323, August 1964. (Cited on page 4)

[5] P. W. Higgs. Broken symmetries and the masses of gauge bosons. *Phys. Rev. Lett.*, 13:508–509, October 1964. (Cited on page 4)

[6] S.P. Martin. A Supersymmetry primer. pages 1–98, 1997. [Adv. Ser. Direct. High Energy Phys.18,1(1998)]. (Cited on page 6)

[7] H. Georgi and S. L. Glashow. Unity of All Elementary Particle Forces. *Phys. Rev. Lett.*, 32:438–441, 1974. (Cited on page 6)

[8] J. R. Ellis. The Superstring: Theory of Everything, or of Nothing? *Nature*, 323:595–598, 1986. (Cited on page 6)

[9] P. Langacker. The physics of heavy Z' gauge bosons. *Rev. Mod. Phys.*, 81:1199–1228, 2009. (Cited on page 6)

[10] A. Leike. The Phenomenology of extra neutral gauge bosons. *Phys. Rept.*, 317:143–250, 1999. (Cited on page 6)

[11] H. S. Bawa. Recent results from Beyond Standard Model (BSM) searches at LHC. *Nucl. Part. Phys. Proc.*, 267-269:277–286, 2015. (Cited on page 6)

[12] S. Sekmen. Beyond the Standard Model physics at the high luminosity LHC. *PoS*, ICHEP2018:283, 2018. (Cited on pages 6 and 15)

[13] The ATLAS collaboration. Search for high-mass dilepton resonances using 139 fb^{-1} of *pp* collision data collected at $\sqrt{s} = 13$ TeV with the ATLAS detector. 2019. (Cited on page 7)

[14] The CMS Collaboration. Search for high-mass resonances in dilepton final states in proton-proton collisions at $\sqrt{s} = 13$ TeV. *JHEP*, 06:120, 2018. (Cited on page 7)

[15] L. Thomas. *Search for new heavy narrow resonances decaying into a dielectron pair with the CMS detector*. PhD thesis, Brussels U., 2014. (Cited on pages 7, 27, A. 8, and B. 1)

[16] G. Fasanella. *Search for new physics in dielectron and diphoton final states at CMS*. PhD thesis, Brussels U., 2017. (Cited on page 7)

[17] C. Caillol. *Scalar boson decays to tau leptons : in the Standard Model and beyond*. PhD thesis, Brussels U., 2016. (Cited on pages 7, 19, 20, and 23)

[18] The CMS collaboration. Measurement of the differential Drell-Yan cross section in proton-proton collisions at 13 TeV. 2016. (Cited on page 7)

[19] The ATLAS Collaboration. Measurement of the high-mass Drell–Yan differential cross-section in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector. *Phys. Lett. B*, 725:223 – 242, 2013. (Cited on page 7)

[20] The CMS Collaboration. Search for new physics in top quark production in dilepton final states in proton-proton collisions at $\sqrt{s} = 13$ TeV. *Internal Report*, 2019. (Cited on page 7)

[21] D. Bourilkov. Photon-induced background for dilepton searches and measurements in *pp* collisions at 13 TeV. *Internal Report*, 2016. (Cited on page 7)

[22] The CMS Collaboration. Search for exclusive or semi-exclusive $\gamma\gamma$ production and observation of exclusive and semi-exclusive e^+e^-. *JHEP*, 80, 2012. (Cited on page 7)

[23] The ATLAS Collaboration. Measurement of exclusive $\gamma\gamma \rightarrow l^+l^-$ production in proton–proton collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector. *Phys. Lett.*, B749:242–261, 2015. (Cited on pages 7 and 56)

[24] The CMS Collaboration. Exclusive $\gamma\gamma \rightarrow \mu^+\mu^-$ production in proton-proton collisions at $\sqrt{s} = 7$ TeV. *JHEP*, 2012. (Cited on page 7)

[25] S.D. Drell and T.-M. Yan. The Drell-Yan process. *Phys. Rev. Lett.*, 25, 1970. (Cited on page 7)

[26] I. R. Kenyon. The Drell-Yan process. *Rept. Prog. Phys.*, 45:1261, 1982. (Cited on page 7)

[27] J. L. Rosner. Forward-backward asymmetries in hadronically produced lepton pairs. *Phys. Lett. D*, 54, 1996. (Cited on pages 10 and 26)

[28] H. Murayama. Final exam solutions for the course Standard Model I 229 A. http://hitoshi.berkeley.edu/229A.html, 2019. (Cited on page 12)

[29] L. Evans and P. Bryant. LHC Machine. *JINST*, 3:S08001, 2008. (Cited on pages 13 and 14)

[30] P. Germain, D. Dekkers, and D. Manglunki. Introduction aux accélérateurs de particules. University Lecture ULB, 2019. (Cited on page 14)

[31] The CMS Collaboration. The CMS experiment at the CERN LHC. *JINST*, 3(08), 2008. (Cited on pages 15, 16, and 18)

[32] J. Butler et al. CMS Phase II upgrade scope document. CERN-LHCC-2015-019. LHCC-G-165, 2015. (Cited on page 16)

[33] P. Baldi, P. Sadowski, and D. Whiteson. Searching for Exotic Particles in High-Energy Physics with Deep Learning. *Nature Commun.*, 5:4308, 2014. (Cited on page 20)

[34] The CMS Collaboration. Identification of heavy-flavour jets with the CMS detector in *pp* collisions at 13 TeV. *JINST*, 13(05):P05011, 2018. (Cited on page 21)

[35] The CMS Collaboration. Description and performance of track and primary-vertex reconstruction with the CMS tracker. *JINST*, 9(10):P10009, 2014. (Cited on page 21)

[36] The CMS Collaboration. Search for Supersymmetry in events with at least three electrons or muons, jets, and missing transverse momentum in proton-proton collisions at $\sqrt{s} = 13$ TeV. *JHEP*, 02:067, 2018. (Cited on page 21)

[37] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands. An Introduction to PYTHIA 8.2. *Comput. Phys. Commun.*, 191:159–177, 2015. (Cited on pages 23 and 33)

[38] A.D. Martin, W.J. Stirling, and R.S. Thorne et al. Parton distributions for the LHC. *Eur.Phys.J.C*, 63:189–285, 2009. (Cited on page 24)

[39] L. Favart. Physique auprès des collisionneurs. University Lecture ULB, 2018-2019. (Cited on page 24)

[40] A.V. Manohar, P. Nason, G.P. Salam, and G. Zanderighi. The photon content of the proton. *JHEP*, 12:046, 2017. (Cited on page 25)

[41] A.V. Manohar, P. Nason, G.P. Salam, and G. Zanderighi. How bright is the proton? A precise determination of the photon parton distribution function. *Phys. Rev. Lett.*, 117, 2016. (Cited on page 25)

[42] The NNPDF Collaboration: V. Bertone, S. Carrazza, N.P. Hartland, and J. Rojo. Illuminating the photon content of the proton within a global PDF analysis. *SciPost Phys.*, 5:8, 2018. (Cited on pages 25 and 26)

[43] J.C. Collins and D.E. Soper. Angular distribution of dileptons in high-energy hadron collisions. *Phys.Rev.*, D16:2219, 1977. (Cited on page 26)

[44] E. Accomando, J. Fiaschi, F. Hautmann, S. Moretti, and C.H. Shepherd-Themistocleous. Photon-initiated production of a dilepton final state at the LHC: Cross section versus forward-backward asymmetry studies. *Phys. Rev. D*, 95, 2017. (Cited on page 30)

[45] The CMS Collaboration. Evidence for exclusive $\gamma\gamma \to W^+W^-$ production and constraints on anomalous quartic gauge couplings at $\sqrt{s} = 8$ TeV. *JHEP*, 8, 2016. (Cited on page 33)

[46] M. Dyndal and L. Schoeffel. The role of finite-size effects on the spectrum of equivalent photons in proton–proton collisions at the LHC. *Phys. Lett.*, B741:66–70, 2015. (Cited on pages 33, 56, and 57)

[47] P. Skands, S. Carrazza, and J. Rojo. Tuning PYTHIA 8.1: the Monash 2013 tune. *Eur. Phys. J. C*, 74:3024, 2014. (Cited on page 33)

[48] V. Khachatryan et al. Event generator tunes obtained from underlying event and multiparton scattering measurements. *Eur. Phys. J. C*, 76:155, 2016. (Cited on page 33)

[49] The CMS Collaboration. Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements. Technical Report CMS-PAS-GEN-17-001, CERN, Geneva, 2018. (Cited on page 33)

[50] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *JHEP*, 07:079, 2014. (Cited on page 34)

[51] R. Frederix and S. Frixione. Merging meets matching in MC@NLO. *JHEP*, 12:061, 2012. (Cited on page 34)

[52] P. Nason and G. Zanderighi. W^+W^-, WZ and ZZ production in the POWHEG BOX V2. *Eur. Phys. J. C*, 74:2702, 2014. (Cited on page 34)

[53] The CMS Collaboration. Performance of missing transverse momentum in pp collisions at $\sqrt{s} = 13$ TeV using the CMS detector. 2018. (Cited on page 39)

[54] F. James. MINUIT function minimization and error analysis. CERN Program Library Long Writeup D506, 1994. (Cited on page 50)

[55] F. James and M. Winkler. MINUIT user's guide, 2004. (Cited on page 50)

[56] D.J. Koskinen. Lecture 7: Parameter estimation and uncertainty. Advanced Methods in Applied Statistics, 2016. (Cited on page 50)

[57] F. James. MINUIT the interpretation of errors, 2004. (Cited on page 50)

[58] The CMS Collaboration. Study of exclusive two-photon production of W^+W^- pairs in pp collisions at 7 TeV, and constraints on anomalous quartic couplings in CMS. *JHEP*, 07:116, 2013. (Cited on page 56)

APPENDIX A

Cross sections of the DY and PI

A.1 Introduction

This appendix presents the calculation of the matrix element required to express the differential cross section of the DY process. Note that the discussion is in fact already simplified as another SM diagram should give a contribution: the production of the scalar boson H (from electroweak symmetry breaking). The strength of that term however has a coupling of fermions to H proportional to the ratio of the fermion mass on the vacuum expectation of the scalar field v, measured to be around 246 GeV in the SM. Hence, an amplitude of about $\frac{m_q\, m_l}{246^2(\mathrm{GeV}^2)}$ is expected for the scalar boson production channel. The quark content of a proton is a delicate piece of information that is described in section 3.2, with equation 3.1 particularly limiting the availability of heavy quark species. The valence quarks, the main constituents of the proton, are the up and down quarks that are thus produced in larger amount than their heavier counterparts. The heaviest lepton, the tau, is endowed a 1.8 GeV mass while the heaviest quark, the top, has a mass of 173 GeV (the second one, the b quark, has a mass of about 4.6 GeV). Only a top-antitop ($t\bar{t}$) initial state with a tau-antitau final state could compete with the two other processes. But producing such an initial state demands the use of a top quark in the proton, something improbable from proton content distribution making this process irrelevant. It is thus neglected in the following derivation.

A.2 The Feynman rules

The Feynman rules and their associated diagrams are a pictorial representation of the perturbation expansion of the time-ordered products. There are different ways of deriving them, either starting from a quantised Lagrangian formulation of time evolution or from time-dependent perturbation theory from the full interacting Hamiltonian expanded around the free one. The Feynman rules are shown in figure A.1 in the Feynman-'t Hooft gauge (the gauge parameter ξ is set to 1 which, for example, simplifies the photon propagator to $i\Pi_{\mu\nu} = \frac{-i\eta_{\mu\nu}}{p^2}$, effectively dropping a $i(1-\xi)\frac{p_\mu p_\nu}{p^2}$ term).

Note the values of the charges Q and the third component of the weak isospin I_W^3 are shown on table A.1.

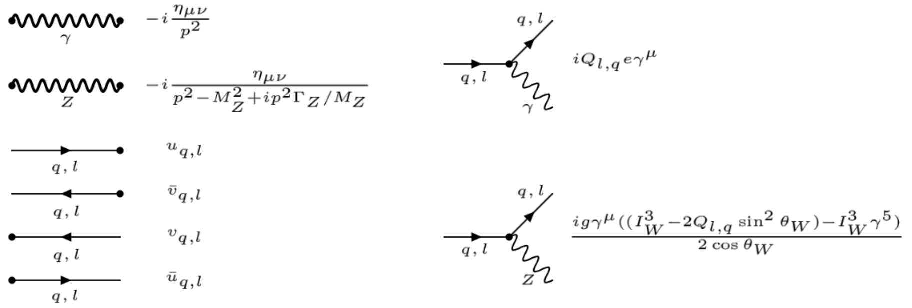


Figure A.1 – The Feynman rules for DY matrix element computation. M_Z and Γ_Z are respectively the Z boson mass and total decay width.

	Q	I_W^3
ν_e, ν_μ, ν_τ	0	$\frac{1}{2}$
e, μ, τ	-1	$-\frac{1}{2}$
u, c, t	$\frac{2}{3}$	$\frac{1}{2}$
d, s, b	$-\frac{1}{3}$	$-\frac{1}{2}$

Table A.1 – Charges Q and third components of the weak isospin I_W^3 in the SM.

These relations include Dirac spinors (bispinor) $u_{q,l}$ ($v_{q,l}$) associated with the fermion (antifermion), either a lepton l or a quark q. They are plane wave solutions of the equations of motions for relativistic spin 1/2 fields satisfying:

$$\sum_{s=1}^{2} u_{s_\alpha}(p)\bar{u}_{s_\beta}(p) = (\not{p} + m\mathbb{1})_{\alpha\beta}, \tag{A.1}$$

$$\sum_{s=1}^{2} v_{s_\alpha}(p)\bar{v}_{s_\beta}(p) = (\not{p} - m\mathbb{1})_{\alpha\beta}, \tag{A.2}$$

where the writing convention adopted here places a bar over spinors to signify $\bar{\psi} = \psi^\dagger\gamma^0$ and a slash across the spinor if it is contracted with the γ^μ matrices ($\not{\psi} = \psi_\mu\gamma^\mu$). $\mathbb{1}$ is the 2×2 identity matrix. The γ^μ matrices are a set of 4×4 matrices satisfying the Dirac algebra:

$$\{\gamma^\mu, \gamma^\nu\} = 2\eta^{\mu\nu}\, I, \tag{A.3}$$

where I is the 4×4 identity matrix. An additional γ^5 matrix, also anti-commuting with the γ^μ, is introduced to be able to project out the left- or right-handed Weyl spinors from a Dirac spinor ($\Psi_D = \Psi_L + \Psi_R$). They are defined, in Weyl representation, as:

$$\gamma^\mu = \begin{pmatrix} 0 & \sigma^\mu \\ \bar{\sigma}^\mu & 0 \end{pmatrix}, \tag{A.4}$$

$$\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3 = \begin{pmatrix} -\mathbb{1} & \\ & \mathbb{1} \end{pmatrix}, \tag{A.5}$$

where $\sigma^\mu = (\mathbb{1}, \vec{\sigma})$, $\bar{\sigma}^\mu = (\mathbb{1}, -\vec{\sigma})$ and $\vec{\sigma}$ is the vector of Pauli matrices $(\sigma^1, \sigma^2, \sigma^3)$. Some interesting properties of these matrices that will prove useful in the computation are listed below:

$$(\gamma^0)^2 = I, \tag{A.6}$$

$$(\gamma^5)^2 = I, \tag{A.7}$$

$$(\gamma^\mu)^\dagger = \gamma^0\gamma^\mu\gamma^0, \tag{A.8}$$

$$(\gamma^5)^\dagger = \gamma^5. \tag{A.9}$$

In a Feynman diagram, external spinors are observable and thus forced to be on-shell (by the Lehmann-Symanzik-Zimmermann (LSZ) reduction formula, relating S-matrix elements for asymptotic momentum eigenstates to quantum fields). This implies external spinors satisfy their equations of motion:

$$(\not{p} - m)u_s(p) = \bar{u}_s(p)(\not{p} - m) = 0 \tag{A.10}$$

$$(\not{p} + m)v_s(p) = \bar{v}_s(p)(\not{p} + m) = 0 \tag{A.11}$$

Finally, many steps of the computation will involve taking the trace of γ matrices. The following relations are easily verified:

$$\mathrm{Tr}[AB...C] = \mathrm{Tr}[B...CA], \tag{A.12}$$

$$\mathrm{Tr}[\eta^{\mu\nu}\mathbb{1}] = \eta^{\mu\nu}\mathrm{Tr}[\mathbb{1}] = 4\eta^{\mu\nu}, \tag{A.13}$$

$$\mathrm{Tr}[\gamma^\mu\gamma^\nu] = 4\eta^{\mu\nu}, \tag{A.14}$$

$$\mathrm{Tr}[\gamma^\mu\gamma^\nu\gamma^\alpha] = 0, \tag{A.15}$$

$$\mathrm{Tr}[\gamma^\alpha\gamma^\mu\gamma^\beta\gamma^\nu] = 4(\eta^{\alpha\mu}\eta^{\beta\nu} - \eta^{\alpha\beta}\eta^{\mu\nu} + \eta^{\alpha\nu}\eta^{\mu\beta}), \tag{A.16}$$

$$\mathrm{Tr}[\gamma^\alpha\gamma^\beta\gamma^\mu\gamma^\nu\gamma^5] = -4i\epsilon^{\alpha\beta\mu\nu}. \tag{A.17}$$

where $\epsilon^{\alpha\beta\mu\nu}$ is a totally antisymmetric tensor such that $\epsilon^{0123} = 1$ and any permutation of adjacent indices flips the sign. In particular, the trace of an odd number of γ matrices is zero!

A.3 All steps of the Drell-Yan cross section calculation

One can now simply read off each Feynman diagram to get the matrix element at leading order.

For figure A.2, the term is:

$$i\mathcal{M}_\gamma = \frac{iQ_lQ_qe^2}{E_{CM}^2}\bar{u}(p_l, s_l)\gamma^\mu v(p_{\bar{l}}, s_{\bar{l}})\bar{v}(p_{\bar{q}}, s_{\bar{q}})\gamma_\mu u(p_q, s_q). \tag{A.18}$$

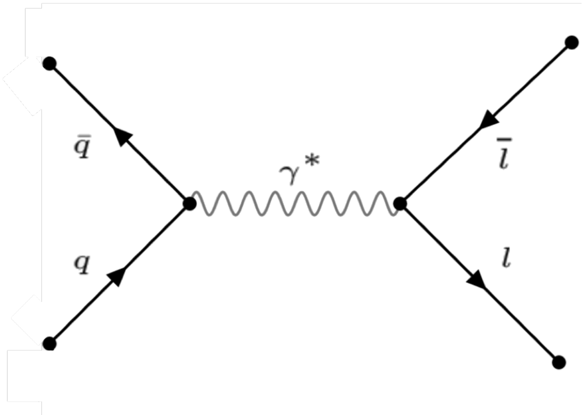


Figure A.2 – The frame to describe the Drell-Yan process.

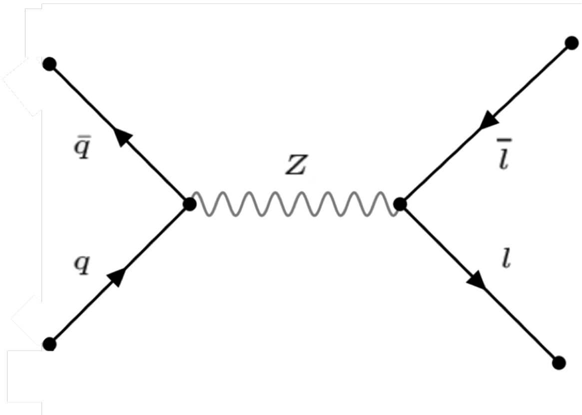


Figure A.3 – The frame to describe the Drell-Yan process.

While for figure A.3, the term is:

$$i\mathcal{M}_Z = \frac{ig^2}{(2\cos\theta_W)^2}\frac{\bar{u}(p_l,s_l)\gamma^\mu(g_{V_l}+g_{A_l}\gamma^5)v(p_{\bar{l}},s_{\bar{l}})\bar{v}(p_{\bar{q}},s_{\bar{q}})\gamma_\mu(g_{V_q}+g_{A_q}\gamma^5)u(p_q,s_q)}{E_{CM}^2-M_Z+iE_{CM}^2\Gamma_Z/M_Z}. \tag{A.19}$$

where

$$e = g\sin\theta_W, \tag{A.20}$$

$$g_{V_{l,q}} = I^3_{Wl,q} - 2Q_{l,q}\sin^2\theta_W, \tag{A.21}$$

$$g_{A_{l,q}} = -I^3_{Wl,q}. \tag{A.22}$$

To simplify the next steps, spinors will be written $\psi(p_i,s_i)\equiv\psi_i$ and one introduces a new parameter

$$\mathcal{R} = \frac{1}{Q_lQ_q\sin^2 2\theta_W}\frac{E_{CM}^2}{E_{CM}^2-M_Z^2+iE_{CM}^2\Gamma_Z/M_Z} \tag{A.23}$$

to rewrite equation A.18 and A.19 as

$$\mathcal{M}_\gamma = \frac{Q_l Q_q e^2}{E_{CM}^2} \bar{u}_l \gamma^\mu v_{\bar{l}} \bar{v}_{\bar{q}} \gamma_\mu u_q \tag{A.24}$$

$$\mathcal{M}_Z = \frac{Q_l Q_q e^2}{E_{CM}^2} \mathcal{R} \bar{u}_l \gamma^\mu (g_{V_l} + g_{A_l} \gamma^5) v_{\bar{l}} \bar{v}_{\bar{q}} \gamma_\mu (g_{V_q} + g_{A_q} \gamma^5) u_q \tag{A.25}$$

The total cross section is thus proportional to the sum squared of the different amplitudes with terms $|\mathcal{M}_\gamma|^2$, $|\mathcal{M}_Z|^2$ and $\mathcal{M}_\gamma \mathcal{M}_Z^* + c.c.$ that can be computed separately and corresponds respectively to a photon propagator, a Z boson one and an interference.

- $|\mathcal{M}_\gamma|^2$:

Using equation A.8, the conjugate of $\mathcal{M}_\gamma$ is simply:

$$\mathcal{M}_\gamma^* = \frac{Q_l Q_q e^2}{E_{CM}^2} [\bar{u}_q \gamma^\mu v_{\bar{q}}][\bar{v}_{\bar{l}} \gamma_\mu u_l], \tag{A.26}$$

where the grouping is there to remind that terms between [ ] are complex numbers for each μ and set of spins. Exploiting this, the average modulus squared matrix element for a photon propagator is:

$$\frac{1}{4} \sum_{s_q, s_{\bar{q}}, s_l, s_{\bar{l}}} \mathcal{M}_\gamma^* \mathcal{M}_\gamma = \frac{Q_l^2 Q_q^2 e^4}{4 E_{CM}^4} \sum_{s_q, s_{\bar{q}}, s_l, s_{\bar{l}}} [\bar{u}_q \gamma^\nu v_{\bar{q}}][\bar{v}_{\bar{l}} \gamma_\nu u_l][\bar{u}_l \gamma^\mu v_{\bar{l}}][\bar{v}_{\bar{q}} \gamma_\mu u_q]. \tag{A.27}$$

Now this expression can be simplified by moving the terms around and making the spinor indices (indicated by greek letters) apparent:

$$\sum_{s_l, s_{\bar{l}}} [\bar{u}_l \gamma^\mu v_{\bar{l}}][\bar{v}_{\bar{l}} \gamma^\nu u_l] = \sum_{s_l, s_{\bar{l}}} \sum_{\alpha,\beta,\epsilon,\delta} \bar{u}_{l_\alpha} \gamma^\mu_{\alpha\beta} v_{\bar{l}_\beta} \bar{v}_{\bar{l}_\epsilon} \gamma^\nu_{\epsilon\delta} u_{l_\delta}. \tag{A.28}$$

The convention adopted here is to sum any two repeated spinor index, with no care given to their position. Using equation A.1 and A.2, this simplifies to:

$$\sum_{\alpha,\beta,\epsilon,\delta} (\not{p}_l + m_l)_{\delta\alpha} (\not{p}_{\bar{l}} - m_{\bar{l}})_{\beta\epsilon} \gamma^\mu_{\alpha\beta} \gamma^\nu_{\epsilon\delta}. \tag{A.29}$$

If one now neglects the masses in this equation (effectively applying the ultra-relativistic limit) and taking the γ out of the $\not{p}$, this can be rewritten into:

$$\sum_{s_l, s_{\bar{l}}} [\bar{u}_l \gamma^\mu v_{\bar{l}}][\bar{v}_{\bar{l}} \gamma^\nu u_l] = \mathrm{Tr}[p_{l_\rho} p_{\bar{l}_\sigma} \gamma^\rho \gamma^\mu \gamma^\sigma \gamma^\nu]. \tag{A.30}$$

The last step in this computation is to take advantage of equation A.16 to reduce this, as well as the quark part that follows exactly the same steps, to finally get:

$$\sum_{s_l, s_{\bar{l}}} \bar{u}_l \gamma^\mu v_{\bar{l}} \bar{v}_{\bar{l}} \gamma^\nu u_l = 4(p_l^\mu p_{\bar{l}}^\nu - p_l^\rho p_{\bar{l}\rho} \eta^{\mu\nu} + p_l^\nu p_{\bar{l}}^\mu), \tag{A.31}$$

$$\sum_{s_q, s_{\bar{q}}} \bar{u}_q \gamma_\mu v_{\bar{q}} \bar{v}_{\bar{q}} \gamma_\nu u_q = 4(p_{q_\mu} p_{\bar{q}_\nu} - p_q^\rho p_{\bar{q}_\rho} \eta_{\mu\nu} + p_{q_\nu} p_{\bar{q}_\mu}). \tag{A.32}$$

Injecting this in equation A.27 crumbles the whole relation to:

$$\frac{1}{4} \sum_{s_q, s_{\bar{q}}, s_l, s_{\bar{l}}} \mathcal{M}_\gamma^* \mathcal{M}_\gamma = 32 \frac{Q_l^2 Q_q^2 e^4}{4 E_{CM}^4} [(p_l.p_q)(p_{\bar{l}}.p_{\bar{q}}) + (p_l.p_{\bar{q}})(p_{\bar{l}}.p_q)], \tag{A.33}$$

where the "." indicates the typical Lorentz indices contraction. Note that by the ultra-relativistic limit (equation 1.3- 1.4), the products of 4-momenta appearing can be simplified as $(p_l.p_q) = (p_{\bar{l}}.p_{\bar{q}}) = \frac{E_{CM}^2}{4}(1 + \cos\theta)$ and $(p_l.p_{\bar{q}}) = (p_{\bar{l}}.p_q) = \frac{E_{CM}^2}{4}(1 - \cos\theta)$. Now, by averaging over the three colour configurations acceptable for the initial state, this can be put into the differential cross section equation 1.5 to finally give:

$$\left(\frac{d\sigma_\gamma}{d\Omega}\right)_{CM} = \frac{Q_l^2 Q_q^2 e^4}{24(4\pi)^2 E_{CM}^2} [(1 + \cos\theta)^2 + (1 - \cos\theta)^2]. \tag{A.34}$$

Note in particular how a flip of $\theta \to \theta + \pi$ leaves the equation invariant, something one could dubbed "forward-backward symmetry". Note also the $\left(\frac{e^2}{4\pi}\right)^2 = \alpha^2$ dependence, as expected from the coupling discussion.

- **$|\mathcal{M}_Z|^2$:**

The procedure is quite similar though the final result will vastly change due to the presence of the $\mathcal{R}^2$ factor and of the γ^5 matrix in the couplings. The conjugate of equation A.25 is:

$$\mathcal{M}_Z^* = \frac{Q_l Q_q e^2}{E_{CM}^2} \mathcal{R}^* [\bar{u}_q \gamma^\mu (g_{V_q} + g_{A_q} \gamma^5) v_{\bar{q}}][\bar{v}_{\bar{l}} \gamma_\mu (g_{V_l} + g_{A_l} \gamma^5) u_l]. \tag{A.35}$$

As in the previous computation, the lepton and quark parts simplify:

$$\sum_{s_l, s_{\bar{l}}} \bar{u}_l \gamma^\mu (g_{V_l} + g_{A_l} \gamma^5) v_{\bar{l}} \bar{v}_{\bar{l}} \gamma^\nu (g_{V_l} + g_{A_l} \gamma^5) u_l = \mathrm{Tr}[p_{l\rho} p_{\bar{l}\sigma} \gamma^\rho \gamma^\mu (g_{V_l} + g_{A_l} \gamma^5) \gamma^\sigma \gamma^\nu (g_{V_l} + g_{A_l} \gamma^5)], \tag{A.36}$$

$$\sum_{s_q, s_{\bar{q}}} \bar{u}_q \gamma_\mu (g_{V_q} + g_{A_q} \gamma^5) v_{\bar{q}} \bar{v}_{\bar{q}} \gamma_\nu (g_{V_q} + g_{A_q} \gamma^5) u_q = \mathrm{Tr}[p_q^\rho p_{\bar{l}}^\sigma \gamma_\rho \gamma_\mu \gamma_\sigma \gamma_\nu (g_{V_q} + g_{A_q} \gamma^5)(g_{V_q} + g_{A_q} \gamma^5)], \tag{A.37}$$

with simply two factors $(g_{V_q} + g_{A_q}\gamma^5)$ and $(g_{V_l} + g_{A_l}\gamma^5)$ appearing on each sides. They however will fundamentally impact the result as they do not commute with the γ matrices. These factors can nonetheless be grouped since $\gamma^5\gamma^\mu\gamma^\nu = \gamma^\mu\gamma^\nu\gamma^5$ by double anticommutation. The $\frac{1}{4}\sum_{s_q, s_{\bar{q}}, s_l, s_{\bar{l}}} \mathcal{M}_Z^* \mathcal{M}_Z$ will thus be fitted the product of two traces:

$$\mathrm{Tr}[p_{l\rho} p_{\bar{l}\sigma} \gamma^\rho \gamma^\mu \gamma^\sigma \gamma^\nu (g_{V_l}^2 + g_{A_l}^2 + 2 g_{V_l} g_{A_l} \gamma^5)] \times \mathrm{Tr}[p_q^\alpha p_{\bar{l}}^\beta \gamma_\alpha \gamma_\mu \gamma_\beta \gamma_\nu \gamma^5)(g_{V_q}^2 + g_{A_q}^2 + 2 g_{V_q} g_{A_q} \gamma^5)]. \tag{A.38}$$

This expression can be separated into three terms: one without γ^5 matrix, one with one and a final one with two of them. The first term, deprived of any γ^5 matrix, will be proportional to the one encountered in the photon case:

$$32\ (g_{V_l}^2 + g_{A_l}^2)(g_{V_q}^2 + g_{A_q}^2)[(p_l.p_q)(p_{\bar{l}}.p_{\bar{q}}) + (p_l.p_{\bar{q}})(p_{\bar{l}}.p_q)]. \tag{A.39}$$

Taking a γ^5 into a trace makes it proportional to the antisymmetric tensor, as indicated by equation A.17. In particular, that trace is antisymmetric for the exchange of μ and ν while the one deprived of γ^5 matrix is symmetric under the same permutation. The product is thus forced to be null. However, a γ^5 in both traces is a non trivial scenario leading, still according to equation A.17, to:

$$4\ (g_{V_l} g_{A_l} g_{V_q} g_{A_q}) p_{l\rho} p_{\bar{l}\sigma} p_q^{\alpha} p_{\bar{q}}^{\beta} (-16) \epsilon^{\rho\sigma\mu\nu} \epsilon_{\alpha\mu\beta\nu}, \tag{A.40}$$

where the contraction of antisymmetric tensor $\epsilon^{\rho\sigma\mu\nu}\epsilon_{\alpha\mu\beta\nu}$ reduces to $2(\delta_\beta^\rho \delta_\alpha^\sigma - \delta_\alpha^\rho \delta_\beta^\sigma)$. Hence, this last product of traces is:

$$128\ (g_{V_l} g_{A_l} g_{V_q} g_{A_q})[(p_l.p_q)(p_{\bar{l}}.p_{\bar{q}}) - (p_l.p_{\bar{q}})(p_{\bar{l}}.p_q)]. \tag{A.41}$$

In conclusion, the modulus squared matrix element averaged over spin configurations sums up to:

$$\begin{aligned}\frac{1}{4}\sum_{s_q,s_{\bar{q}},s_l,s_{\bar{l}}} \mathcal{M}_Z^* \mathcal{M}_Z =& \frac{Q_l^2 Q_q^2 e^4}{4E_{CM}^4} |\mathcal{R}|^2 \{128\ (g_{V_l} g_{A_l} g_{V_q} g_{A_q})[(p_l.p_q)(p_{\bar{l}}.p_{\bar{q}}) - (p_l.p_{\bar{q}})(p_{\bar{l}}.p_q)] \\ &+ 32\ (g_{V_l}^2 + g_{A_l}^2)(g_{V_q}^2 + g_{A_q}^2)[(p_l.p_q)(p_{\bar{l}}.p_{\bar{q}}) + (p_l.p_{\bar{q}})(p_{\bar{l}}.p_q)]\}.\end{aligned} \tag{A.42}$$

One simply has to inject this into equation 1.5, averaging over colour configuration and taking profit of the ultra-relativistic 4-momenta to express the differential cross section under Z boson propagation:

$$\begin{aligned}\left(\frac{d\sigma_Z}{d\Omega}\right)_{CM} =& \frac{Q_l^2 Q_q^2 e^4}{24(4\pi)^2 E_{CM}^2} |\mathcal{R}|^2 \{[(g_{V_l}^2 + g_{A_l}^2)(g_{V_q}^2 + g_{A_q}^2) + 4 g_{V_l} g_{A_l} g_{V_q} g_{A_q}](1+\cos\theta)^2 \\ &+ [(g_{V_l}^2 + g_{A_l}^2)(g_{V_q}^2 + g_{A_q}^2) - 4 g_{V_l} g_{A_l} g_{V_q} g_{A_q}](1-\cos\theta)^2\}.\end{aligned} \tag{A.43}$$

Strikingly, if none of the couplings are null, this result will vary under a $\theta \to \theta + \pi$ transform: forward-backward asymmetry arises because of the nature of the weak interaction through the presence of the Z gauge boson and interaction through axial and vectorial couplings.

- **$\mathcal{M}_\gamma \mathcal{M}_Z * + c.c.$:**

The interference term mixes the Z boson and photon couplings. Walking through the same steps, one obtains for the lepton and quark parts:

$$\sum_{s_l,s_{\bar{l}}} \bar{u}_l \gamma^\mu v_{\bar{l}} \bar{v}_{\bar{l}} \gamma^\nu (g_{V_l} + g_{A_l}\gamma^5) u_l = \mathrm{Tr}[p_{l\rho} p_{\bar{l}\sigma} \gamma^\rho \gamma^\mu \gamma^\sigma \gamma^\nu (g_{V_l} + g_{A_l}\gamma^5)], \tag{A.44}$$

$$\sum_{s_q,s_{\bar{q}}} \bar{u}_q \gamma_\mu v_{\bar{q}} \bar{v}_{\bar{q}} \gamma_\nu (g_{V_q} + g_{A_q}\gamma^5) u_q = \mathrm{Tr}[p_q^\rho p_{\bar{l}}^\sigma \gamma_\rho \gamma_\mu \gamma_\sigma \gamma_\nu (g_{V_q} + g_{A_q}\gamma^5)]. \tag{A.45}$$

The global term requires a multiplication of the lepton and quark traces. With the same reasoning as for the case of Z boson propagation, one concludes that any product of traces with a γ^5 in only one of the term will be null, due to the contraction of symmetric and antisymmetric tensors. What is left is the case deprived of γ^5, which is proportional to the photon case, and one with both traces taken with a γ^5, proportional to the equivalent Z sub-case, both endowed with a factor $\mathcal{R}^*$. We thus get:

$$32\ (g_{V_l} g_{V_q})[(p_l.p_q)(p_{\bar{l}}.p_{\bar{q}}) + (p_l.p_{\bar{q}})(p_{\bar{l}}.p_q)], \tag{A.46}$$

$$32\ (g_{A_l} g_{A_q})[(p_l.p_q)(p_{\bar{l}}.p_{\bar{q}}) - (p_l.p_{\bar{q}})(p_{\bar{l}}.p_q)]. \tag{A.47}$$

Placing the ultra-relativistic expression of 4-momenta and averaging over colours, one gets for the averaged matrix element:

$$\begin{aligned}\frac{1}{12}\sum_{s_q,s_{\bar{q}},s_l,s_{\bar{l}}} \mathcal{M}_\gamma \mathcal{M}_Z^* = \frac{32 Q_l^2 Q_q^2 e^4}{12(4)^2} \mathcal{R}^* \{ & [g_{V_l} g_{V_q} + g_{A_l} g_{A_q}](1+\cos\theta)^2 \\ & + [g_{V_l} g_{V_q} - g_{A_l} g_{A_q}](1-\cos\theta)^2 \}\end{aligned} \tag{A.48}$$

The complex conjugate of this expression will be the same term with $\mathcal{R}^* \to \mathcal{R}$. Hence summing them and placing the expression in the cross section equation 1.5, one gets:

$$\begin{aligned}\left(\frac{d\sigma_{\text{Int}}}{d\Omega}\right)_{CM} = \frac{32 * 2 Q_l^2 Q_q^2 e^4}{12 * 64 (4\pi)^2 E_{CM}^2} \text{Re}\{\mathcal{R}\}\{ & [g_{V_l} g_{V_q} + g_{A_l} g_{A_q}](1+\cos\theta)^2 \\ & + [g_{V_l} g_{V_q} - g_{A_l} g_{A_q}](1-\cos\theta)^2 \}\end{aligned} \tag{A.49}$$

The cross section for the interference process is thus:

$$\begin{aligned}\left(\frac{d\sigma_{\text{Int}}}{d\Omega}\right)_{CM} = \frac{Q_l^2 Q_q^2 e^4}{24 (4\pi)^2 E_{CM}^2} \text{Re}\{\mathcal{R}\}\{ & 2[g_{V_l} g_{V_q} + g_{A_l} g_{A_q}](1+\cos\theta)^2 \\ & + 2[g_{V_l} g_{V_q} - g_{A_l} g_{A_q}](1-\cos\theta)^2 \}\end{aligned} \tag{A.50}$$

The condition for forward-backward asymmetry in this case is less demanding as solely having the couplings g_{A_l} and g_{A_q} non-null is sufficient. The asymmetry finds its origin in the mixture of axial and vectorial coupling and is independent from the centre of mass energy.

- **The global cross section $|\mathcal{M}_Z|^2 + |\mathcal{M}_\gamma|^2 + (\mathcal{M}_\gamma \mathcal{M}_Z * + c.c.)$:**

Summing equations 1.7, 1.8 and 1.9, one unravels the global Drell-Yan angular differential cross section after some simple algebraic manipulations [15]:

$$\left(\frac{d\sigma_{\text{DY}}}{d\Omega}\right)_{CM} = \frac{Q_l^2 Q_q^2 e^4}{24 (4\pi)^2 E_{CM}^2} [c_1 (1+\cos^2\theta) + c_2 \cos\theta] \tag{A.51}$$

with

$$c_1 = 1 + 2\ \text{Re}\{\mathcal{R}\} g_{V_l} g_{V_q} + |\mathcal{R}|^2 (g_{V_l}^2 + g_{A_l}^2)(g_{V_q}^2 + g_{A_q}^2) \tag{A.52}$$

$$c_2 = 4\ \text{Re}\{\mathcal{R}\} g_{A_l} g_{A_q} + 8\ |\mathcal{R}|^2 g_{V_l} g_{A_l} g_{V_q} g_{A_q} \tag{A.53}$$

The total cross section is simply the integral over the solid angle Ω of equation 1.10:

$$\sigma_{\mathrm{DY}} = \int_{\Omega} \frac{d\sigma_{\mathrm{DY}}}{d\Omega} d\Omega = \frac{4\pi}{9} \left(\frac{e^2}{4\pi}\right)^2 \frac{Q_q^2}{E_{CM}^2}\, c_1 \tag{A.54}$$

A.4 The PI cross section calculation

The Feynman rules are the same as that in the Drell-Yan case with a few additional rules presented in figure A.4:

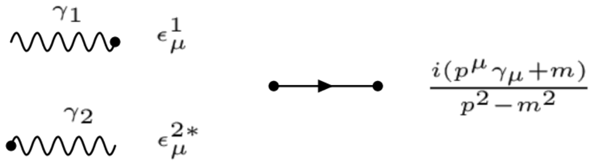


Figure A.4 – Additional Feynman rules for external photon and spinor propagator of mass m.

On-shell conditions for external photons sets their 4-momenta squared to 0: $p_\gamma^2 = 0$. The matrix element of this process is quite straightforward:

$$i\mathcal{M}_{\mathrm{PI}} = Q_l Q_{\bar{l}} (ie)^2 \epsilon^1_\mu \epsilon^2_\nu \bar{u}_l \left[\gamma^\mu \frac{i(\not{p}_l - \not{p}_{\gamma_1} + m)}{(p_l - p_{\gamma_1})^2 - m^2}\gamma^\nu + \gamma^\nu \frac{i(\not{p}_l - \not{p}_{\gamma_2} + m)}{(p_l - p_{\gamma_2})^2 - m^2}\gamma^\mu\right] v_{\bar{l}} \tag{A.55}$$

where the relative sign between the two amplitudes is plus because they are related by a Bose symmetry $p_{\gamma_1} \leftrightarrow p_{\gamma_2}$ and the product $Q_l Q_{\bar{l}}$ is forcibly $+1$ due to the nature of leptons.

The next steps involve squaring the matrix element and average over external lepton and photon polarisations. The leptons follow the same procedure as in the case of the Drell-Yan process. Photons however require a different approach. The prescription to follow is:

$$\sum_{\text{polarisation i}} \epsilon^{i*}_\mu \epsilon^i_\nu \rightarrow -\eta_{\mu\nu} \tag{A.56}$$

This is not an equality but a procedure that holds in any physical matrix element computation (when all relevant diagrams are summed) thanks to the Ward identity $p^\mu M_\mu = 0$. Thus the term will give:

$$\begin{aligned}\sum_{\mathrm{pol.}i,s_l,s_{\bar{l}}} |\mathcal{M}_{\mathrm{PI}}|^2 = e^4 \mathrm{Tr}\Bigg[(\not{p}_l + m)\left(\gamma^\mu \frac{(\not{p}_l - \not{p}_{\gamma_1} + m)}{(p_l - p_{\gamma_1})^2 - m^2}\gamma^\nu + \gamma^\nu \frac{(\not{p}_l - \not{p}_{\gamma_2} + m)}{(p_l - p_{\gamma_2})^2 - m^2}\gamma^\mu\right) \\ .\ (\not{p}_{\bar{l}} - m)\left(\gamma_\nu \frac{(\not{p}_l - \not{p}_{\gamma_1} + m)}{(p_l - p_{\gamma_1})^2 - m^2}\gamma_\mu + \gamma_\mu \frac{(\not{p}_l - \not{p}_{\gamma_2} + m)}{(p_l - p_{\gamma_2})^2 - m^2}\gamma_\nu\right)\Bigg]\end{aligned} \tag{A.57}$$

Now while this truly looks massive it can be quite easily simplified by proceeding as before. First one can simplify the expression using information about the 4-momenta. Indeed:

$$(p_l - p_{\gamma_1})^2 - m^2 = -2p_l.p_{\gamma_1} \tag{A.58}$$

$$(p_l - p_{\gamma_2})^2 - m^2 = -2p_l.p_{\gamma_2} \tag{A.59}$$

since both the photons and the letpons are on-shell. The numerator can also gets simpler using Dirac algebra: $\gamma^\nu \not{p} \gamma_\nu = -2\not{p}$. This leads to rewrite equation A.57 as:

$$\frac{1}{4}\sum_{\text{pol}.i,s_l,s_{\bar{l}}} |\mathcal{M}_{\text{PI}}|^2 = \frac{e^4}{4}\left[\frac{\mathbf{I}}{(2p_l.p_{\gamma_1})^2} + \frac{\mathbf{II}}{(2p_l.p_{\gamma_1})(2p_l.p_{\gamma_2})} + \frac{\mathbf{III}}{(2p_l.p_{\gamma_2})(2p_l.p_{\gamma_1})} + \frac{\mathbf{IV}}{(2p_l.p_{\gamma_2})^2}\right] \tag{A.60}$$

where **I**, **II**, **III** and **IV** are complicated traces. However, **IV** is similar to **I** under $p_{\gamma_1} \to p_{\gamma_2}$ and **II** is equivalent to **III** since one can reverse the order of γ matrices inside a trace. In the first case:

$$\mathbf{I} = \text{Tr}[\gamma_\mu(\not{p}_l + m)\gamma^\mu(\not{p}_l - \not{p}_{\gamma_1} + m)\gamma^\nu(\not{p}_{\bar{l}} - m)\gamma_\nu(\not{p}_l - \not{p}_{\gamma_1} + m)] \tag{A.61}$$

Any term containing an odd number of γ matrices can be left out. Proceeding this way, one can reach the final answer by deriving all terms of equation A.60.

APPENDIX B

The Collins-Soper frame

B.1 Deriving the $\cos\theta_{CS}$ expression

The computation steps presented follow that of [15]. The objective is to reach equation 3.3 from succession of Lorentz transform starting from the laboratory frame, entitled $\mathcal{R}$. In this frame, the proton directions are along the z-axis and they are named A for the proton furnishing the quark parton and B for the other one, offering the antiquark. Three momenta in the laboratory frame are written $\vec{p}_A$ and $\vec{p}_B$ and in the centre of mass of the dilepton pair as $\vec{p'}_A$ and $\vec{p'}_B$. The Collins-Soper frame, $\mathcal{R}_{CS}$, is defined in the centre of mass of the dilepton pair with the axis rotated so that z' bisects $\vec{p'}_A$ and $-\vec{p'}_B$ and the x'-axis, orthogonal to the z'-one, lies in the plane defined by $\vec{p'}_A$ and $-\vec{p'}_B$ with direction set so that both of these vectors projection along it are negative. The angle θ_{CS} is defined by the negatively charged lepton direction and the z'-axis.

The computation assumes head-on collisions of equally energetic protons. Note that there is an actual tiny angle of intersection between the two proton beams in the CMS detector but is in fact very much negligible given the high energy of the particles. In the laboratory frame, the z-axis is along the protons direction, the x-axis matches the direction of the $q\bar{q}$ pair transverse orientation and the y-axis makes the whole set dextrogyral.

To go from the laboratory frame to the Collins-Soper one, a boost (to get to centre of mass frame) along x and z and a rotation around the boosted y-axis (to orientate the frame as prescribed) are necessary. Naming CS frame variables with a " $'$ " and lab-frame ones without any, one can write the connection between both frames in a parametrised (β_x, β_y and α) form:

$$\begin{pmatrix} E' \\ p'_x \\ p'_y \\ p'_z \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\alpha & 0 & \sin\alpha \\ 0 & 0 & 1 & 0 \\ 0 & -\sin\alpha & 0 & \cos\alpha \end{pmatrix} \begin{pmatrix} \gamma & -\beta_x\gamma & 0 & -\beta_z\gamma \\ -\beta_x\gamma & 1+(\gamma-1)\frac{\beta_x^2}{\beta^2} & 0 & (\gamma-1)\frac{\beta_x\beta_z}{\beta^2} \\ 0 & 0 & 1 & 0 \\ -\beta_z\gamma & (\gamma-1)\frac{\beta_x\beta_z}{\beta^2} & 0 & 1+(\gamma-1)\frac{\beta_z^2}{\beta^2} \end{pmatrix} \begin{pmatrix} E \\ p_x \\ p_y \\ p_z \end{pmatrix}, \tag{B.1}$$

where $\beta = \sqrt{\beta_x^2 + \beta_z^2}$ and $\gamma = \frac{1}{\sqrt{1-\beta^2}}$ are the boost and Lorentz factor. For head-on collisions one expects:

$$p_A = (E_A, 0, 0, p_{z,A}), \tag{B.2}$$
$$p_B = (E_B, 0, 0, p_{z,B}), \tag{B.3}$$
$$p'_A = (E'_A, p'_{x,A}, 0, p'_{z,A}), \tag{B.4}$$
$$p'_B = (E'_B, p'_{x,B}, 0, p'_{z,B}). \tag{B.5}$$

α is constrained by the requirement of z' to bisect the momenta:

$$\frac{E'_A}{p'_{x,A}} = \frac{E'_B}{p'_{x,B}}. \tag{B.6}$$

Expressing this requirement with the laboratory frame variables thanks to equation B.1, the following two conditions can be established:

$$\sin\alpha = \frac{\beta_x \beta_z \frac{\gamma}{\gamma+1}}{\sqrt{1-\beta_z^2}}, \tag{B.7}$$
$$\cos\alpha = \frac{1-\beta_z^2 \frac{\gamma}{\gamma+1}}{\sqrt{1-\beta_z^2}}. \tag{B.8}$$

Introducing the dilepton variables $M_{l\bar{l}}$ for its mass and $p_{l\bar{l}} = (E_{l\bar{l}}, p_{x,l\bar{l}}, 0, p_{z,l\bar{l}})$ for four-momentum in laboratory frame, the boost parameters can be re-expressed as:

$$\beta_x = \frac{p_{x,l\bar{l}}}{E_{l\bar{l}}}, \tag{B.9}$$
$$\beta_z = \frac{p_{z,l\bar{l}}}{E_{l\bar{l}}}, \tag{B.10}$$
$$\gamma = \frac{E_{l\bar{l}}}{M_{l\bar{l}}}. \tag{B.11}$$

The variable of interest can be expressed as:

$$\cos\theta_{CS} = \frac{p'_{z,l}}{E'_l}. \tag{B.12}$$

The final expression is then obtained by joining all the different relations and using the fact that in the CS frame, the lepton and antilepton energies are equal so that $E'_l = M_{l\bar{l}}/2$ and $p'_{z,l}$ can be expressed as a function of the laboratory frame variables through equation B.1. Applying these steps, the z momentum in CS frame of the lepton can be expressed as:

$$p'_{z,l} = \frac{p_{z,l}E_{\bar{l}} - p_{z,\bar{l}}E_l}{\sqrt{p_{x,l\bar{l}}^2 + M_{l\bar{l}}^2}}, \tag{B.13}$$

leading equation B.12 to be reformulated as:

$$\cos\theta_{CS} = \frac{2}{M_{l\bar{l}}} \frac{E_{\bar{l}}\, p_{z,l} - E_l\, p_{z,\bar{l}}}{\sqrt{M_{l\bar{l}}^2 + P_{x,l\bar{l}}^2}}, \tag{B.14}$$

which is nothing else but equation 3.3 in a frame such that the transverse momentum is along x. Note that the correction induced by accounting for the very slight angle between the two protons that do not actually collide head-on only insert a modification of the order of the MeV, while the variables one commonly deals with range in the GeV.

APPENDIX C

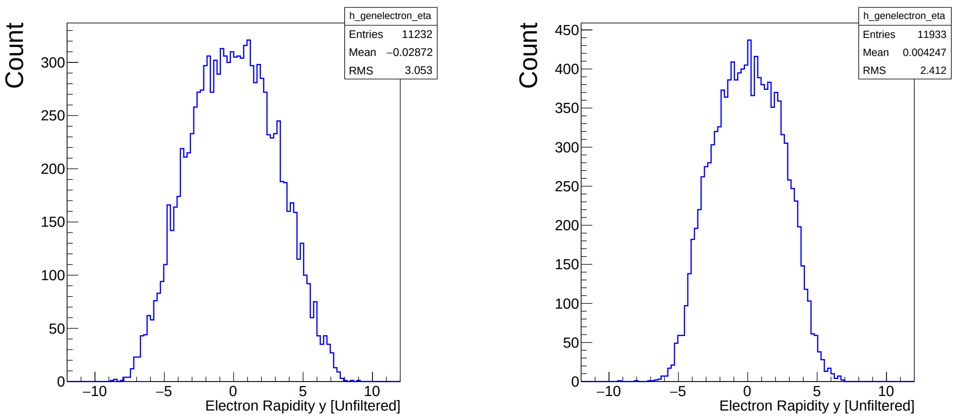
Generated level distributions

C.1 The kinematics of DY and PI processes at the generated level

This section contains plots displaying kinematics information on the dilepton pair produced in DY and PI processes at the generated level. Both plots before and after acceptance cuts are indicated. For some specific variables, lepton-antilepton (electron-positron) separated plots are also available.

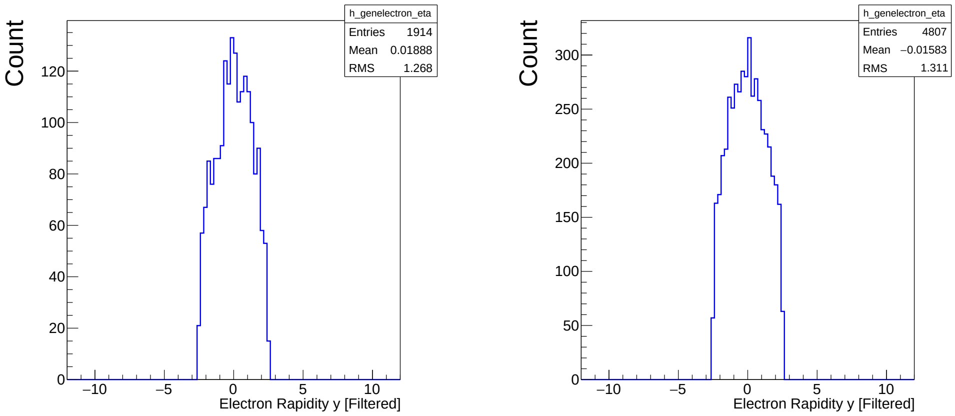
C.1.1 Variable: rapidity of lepton η

This variable offers, at the reconstructed level, a good test on the isotropy of the detector. Indeed, any deviation from symmetry of the distribution in symmetric collisions should be regarded as highly conspicuous. At the generated level, it is rather interesting to compare the behaviour of leptons to antileptons. For example, plots C.5 and C.16 indicate a distinct behaviour for the electron and positron, a sign of the forward-backward asymmetry.



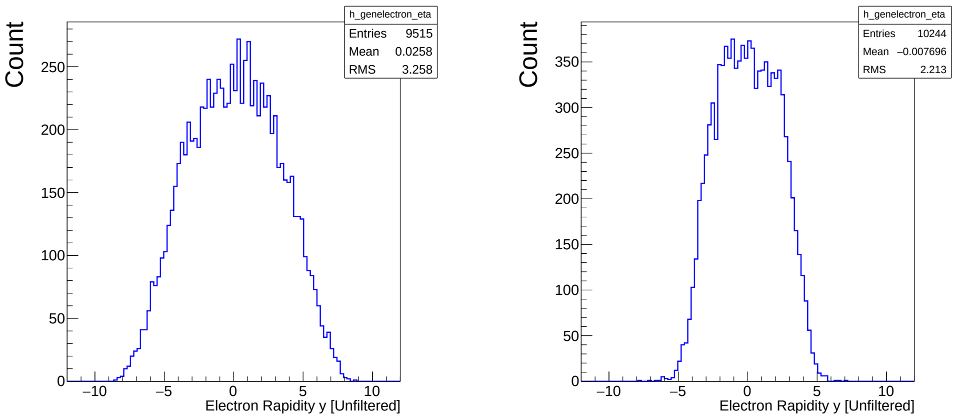
(a) PI *(b) DY*

Figure C.1 – The electron η distributions, $M_{l\bar{l}} > 60$ GeV without acceptance cuts.



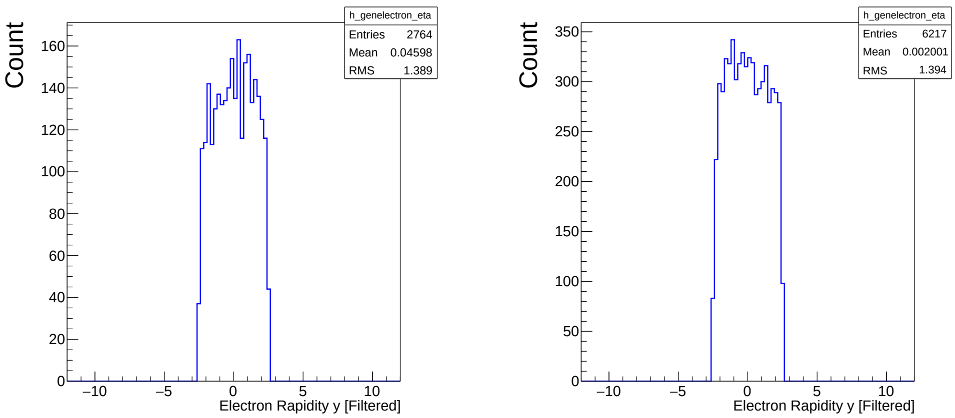
(a) PI *(b) DY*

Figure C.2 – The electron η distributions, $M_{l\bar{l}} > 60$ GeV with acceptance cuts.



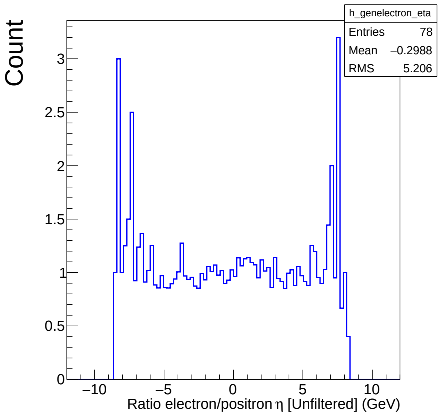
(a) PI *(b) DY*

Figure C.3 – The electron η distributions, $M_{l\bar{l}} \in [200, 400]$ GeV without acceptance cuts.

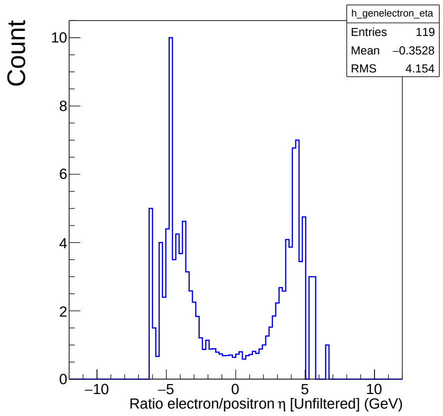


(a) PI *(b) DY*

Figure C.4 – The electron η distributions, $M_{l\bar{l}} \in [200, 400]$ GeV with acceptance cuts.

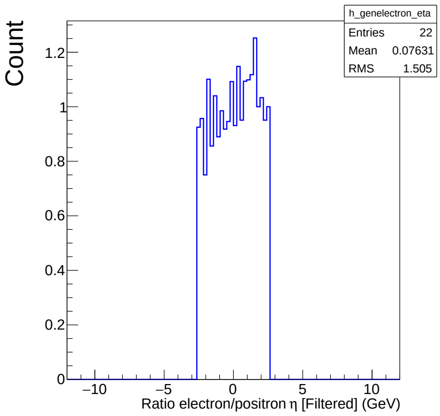


(a) PI

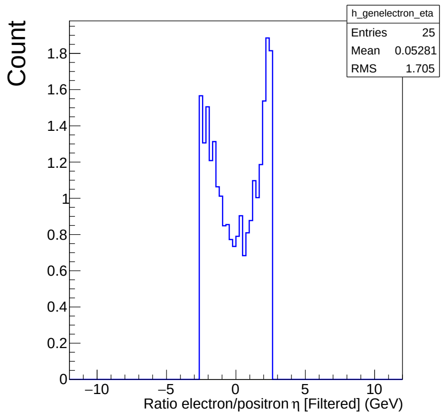


(b) DY

Figure C.5 – The ratio e^-/e^+ as a function of η, $M_{l\bar{l}} \in [200, 400]$ GeV without acceptance cuts. The border of the distributions are impacted by low statistics.



(a) PI

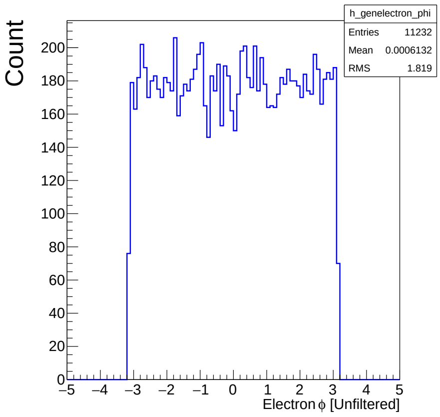


(b) DY

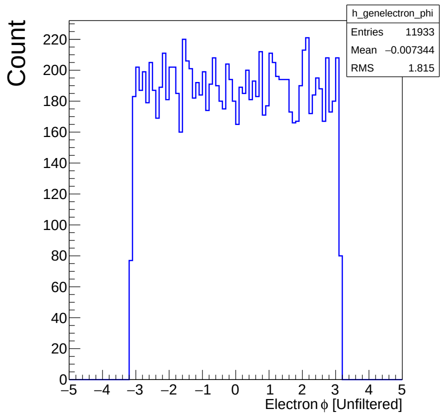
Figure C.6 – The ratio e^-/e^+ as a function of η, $M_{l\bar{l}} \in [200, 400]$ GeV with acceptance cuts.

C.1.2 Variable: lepton ϕ

As expected from symmetry arguments, this variable is flat.

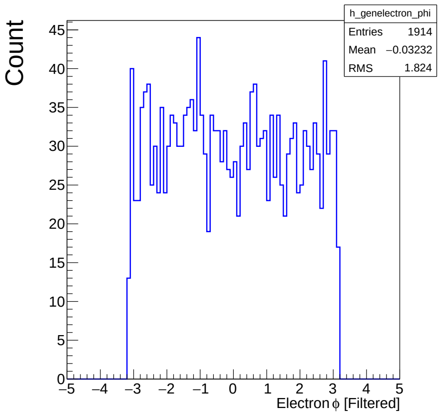


(a) PI

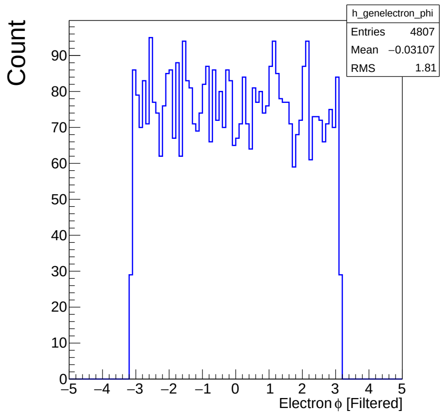


(b) DY

Figure C.7 – The electron ϕ distributions, $M_{l\bar{l}} > 60$ GeV without acceptance cuts.

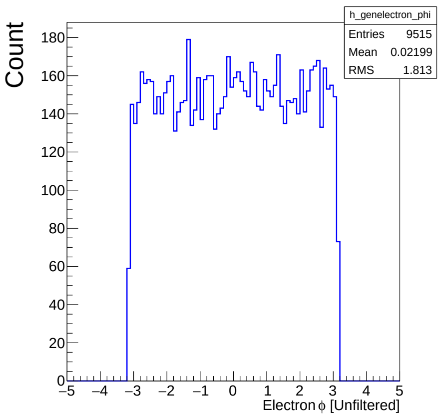


(a) PI

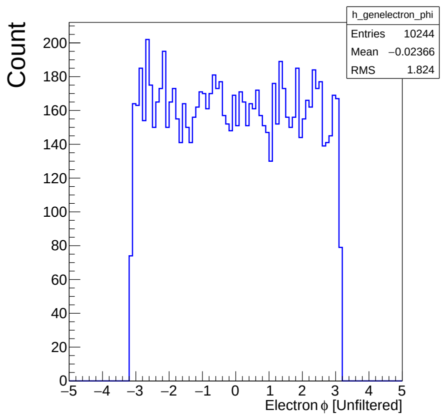


(b) DY

Figure C.8 – The electron ϕ distributions, $M_{l\bar{l}} > 60$ GeV with acceptance cuts.

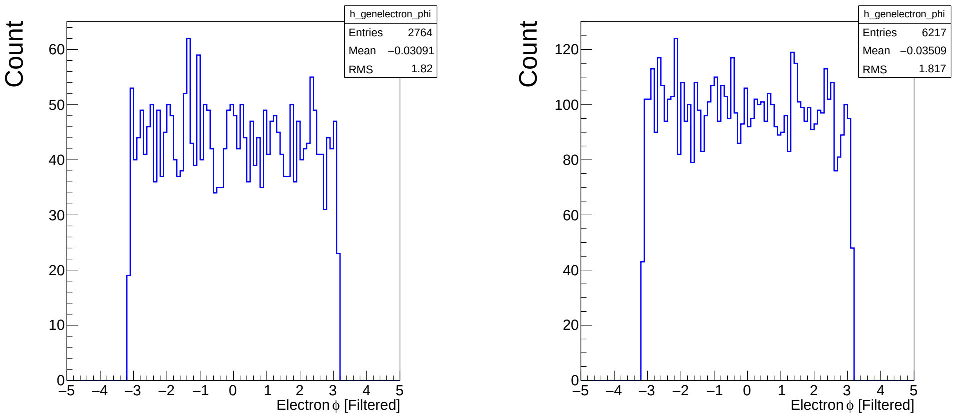


(a) PI



(b) DY

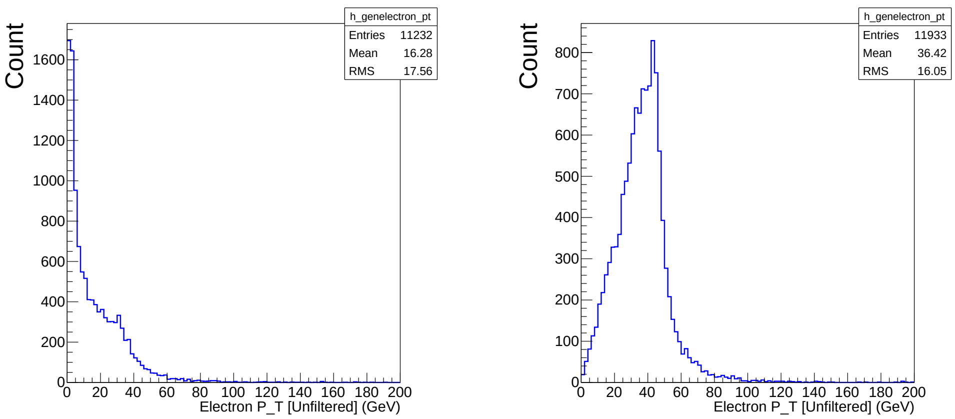
Figure C.9 – The electron ϕ distributions, $M_{l\bar{l}} \in [200, 400]$ GeV without acceptance cuts.



(a) PI *(b) DY*

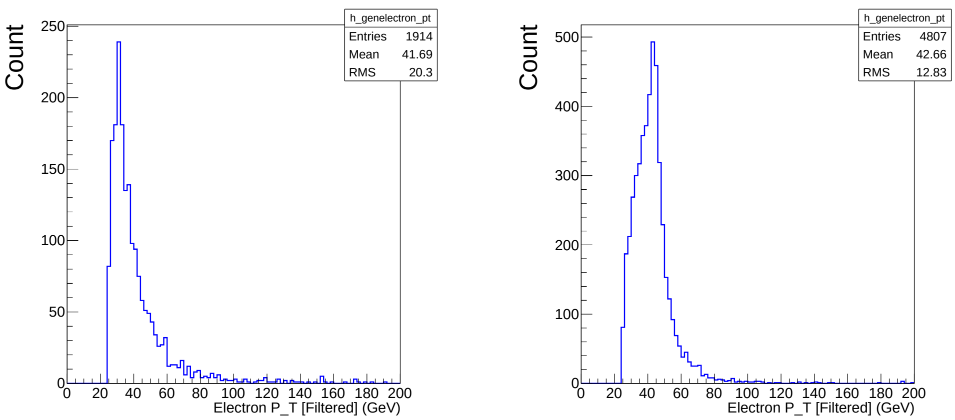
Figure C.10 – The electron ϕ distributions, $M_{l\bar{l}} \in [200, 400]$ GeV with acceptance cuts.

C.1.3 Variable: lepton p_T



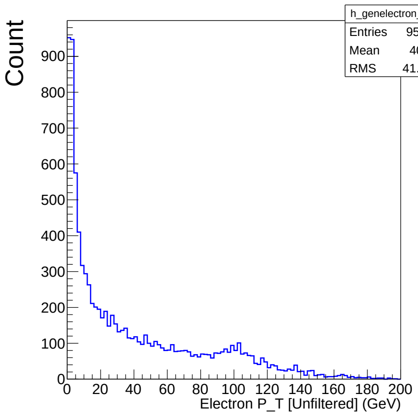
(a) PI *(b) DY*

Figure C.11 – The electron p_T distributions, $M_{l\bar{l}} > 60$ GeV without acceptance cuts.

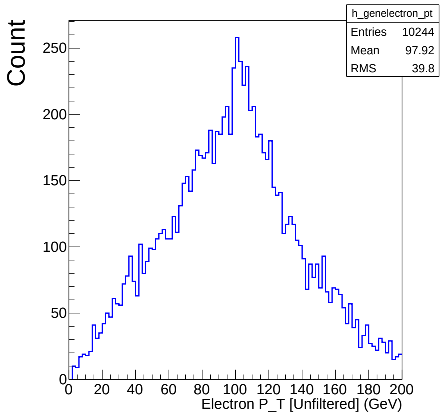


(a) PI *(b) DY*

Figure C.12 – The electron p_T distributions, $M_{l\bar{l}} > 60$ GeV with acceptance cuts.

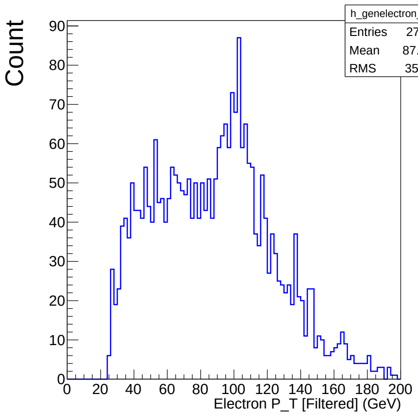


(a) PI

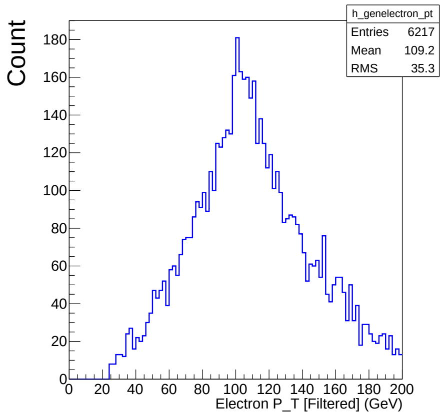


(b) DY

Figure C.13 – The electron p_T distributions, $M_{l\bar{l}} \in [200, 400]$ GeV without acceptance cuts.

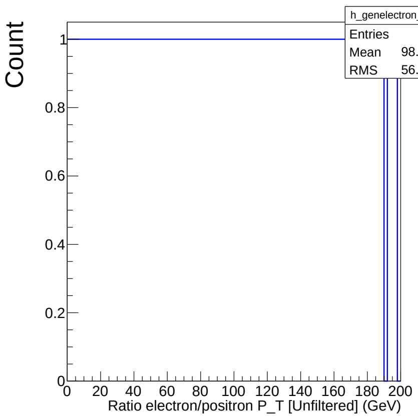


(a) PI

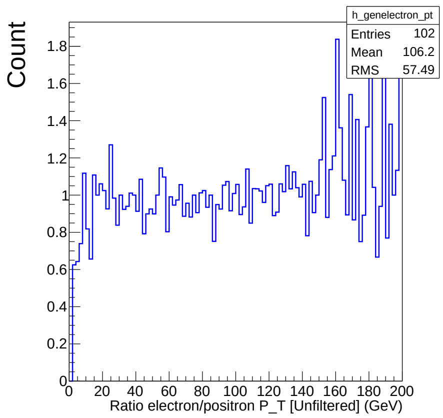


(b) DY

Figure C.14 – The electron p_T distributions, $M_{l\bar{l}} \in [200, 400]$ GeV with acceptance cuts.



(a) PI

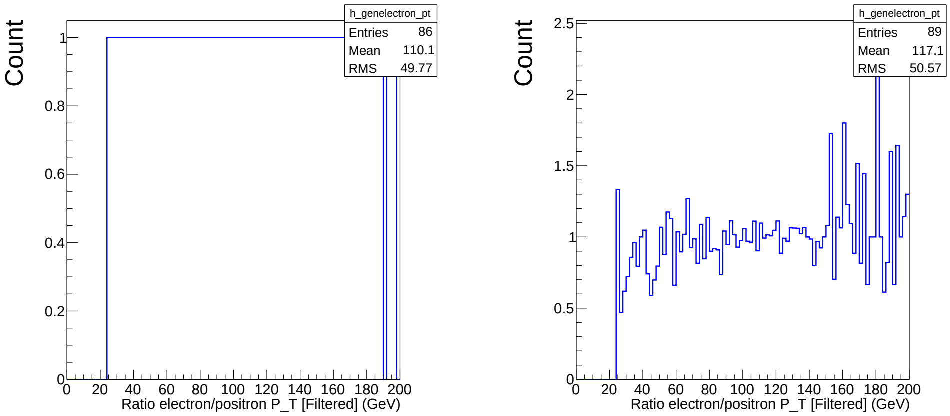


(b) DY

Figure C.15 – The ratio e^-/e^+ as a function of p_T, $M_{l\bar{l}} \in [200, 400]$ GeV without acceptance cuts.

C.2 The DY vs PI cross sections under cuts

Applying the set of cuts presented in section 3.4.2 leads to the set of acceptances in figure 3.11. Rescaling the cross sections of the DY and PI accordingly delivers the set of cross sections presented



(a) PI *(b) DY*

Figure C.16 – The ratio e^-/e^+ as a function of p_T, $M_{l\bar{l}} \in [200, 400]$ GeV with acceptance cuts.

in figure C.17.

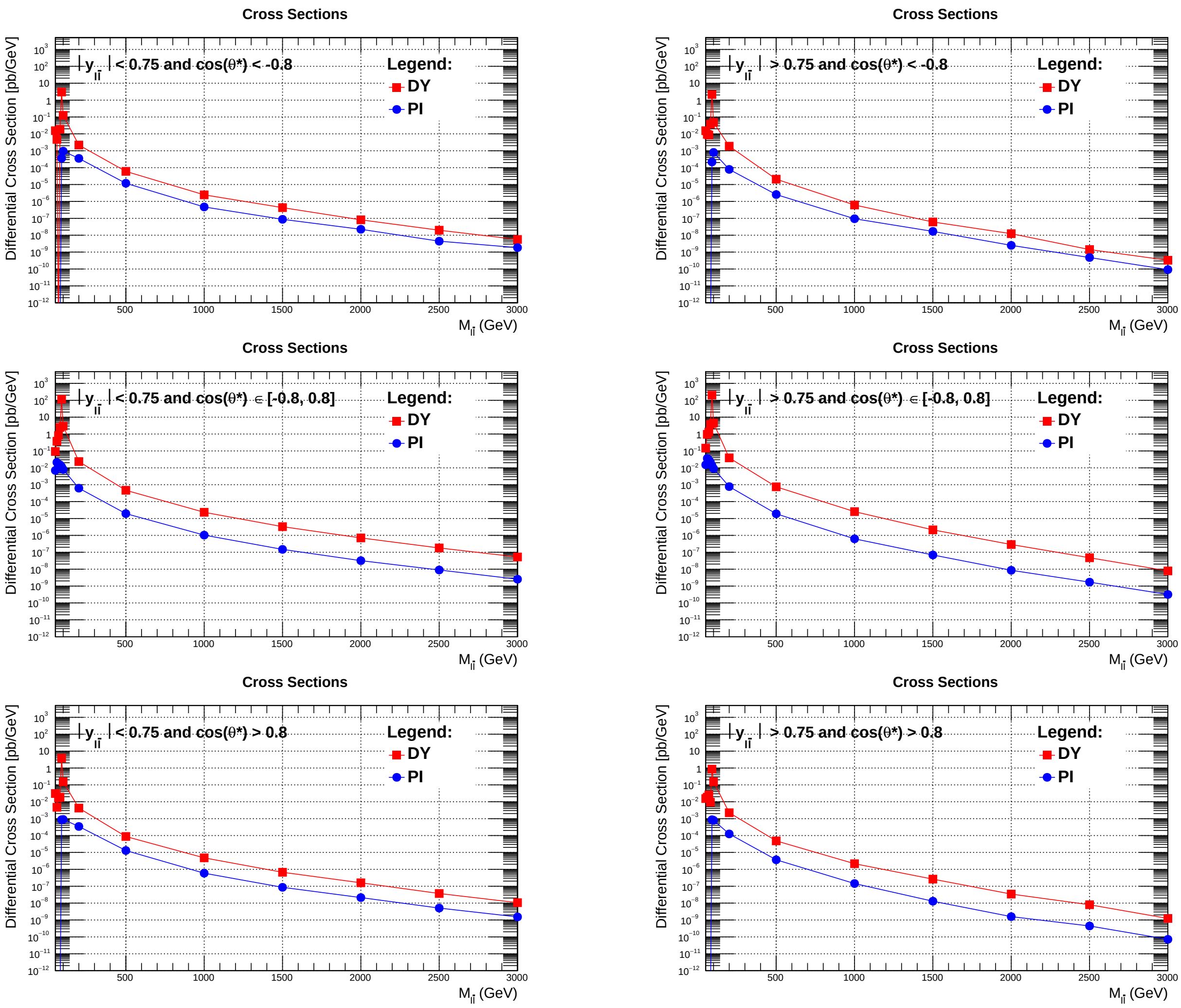


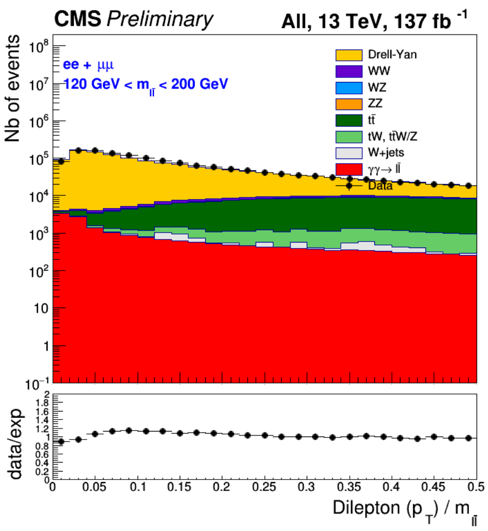
Figure C.17 – Cross sections of the DY (red) and the PI (blue) for the different sets of cuts.

APPENDIX D

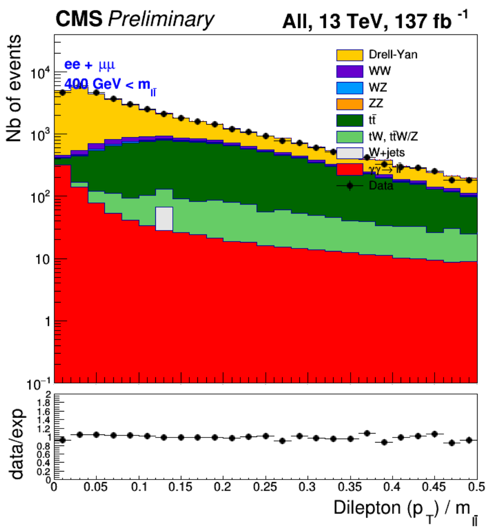
Reconstructed level distributions

D.1 The kinematics of the final state

This section displays three variables that could play a role in disentangling the PI signal from other backgrounds. They are however not necessary in the analysis as the separation power attained by the cuts indicated in the core of the text is sufficient. Further cuts on these variables are hence not demanded but could play a role in a different approach that would implement a higher number of looser cuts. Note that the procedure followed to isolate the signal was designed by hand because of a relative low confidence in the quality of some simulated variables, particularly those attaining to the pile-up. MVA methods could have been employed but at the risk of the procedure relying to heavily on unsatisfyingly modelled variables ... they are thus reserved for precision measurement of well-understood channels and not exploratory searches as the one undergone here.

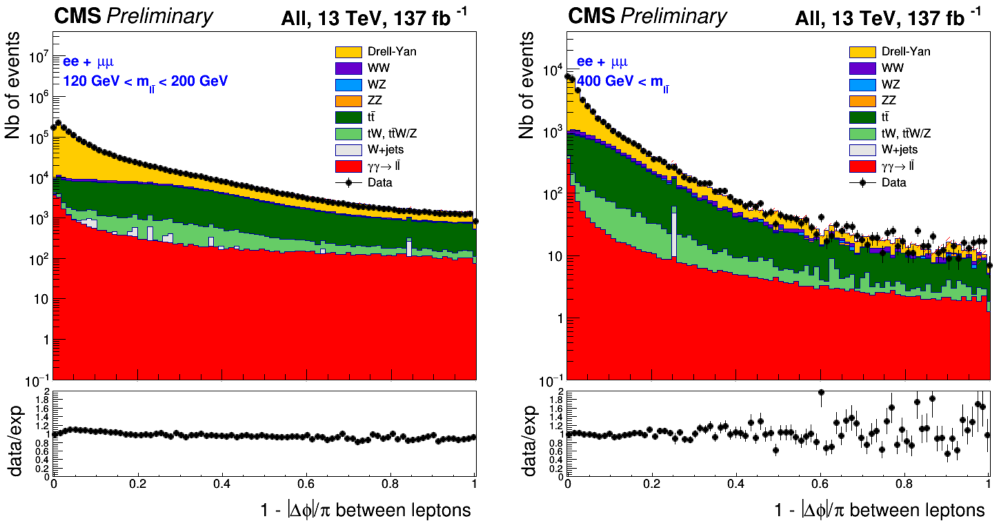


(a) Mass bin [120, 200] GeV.



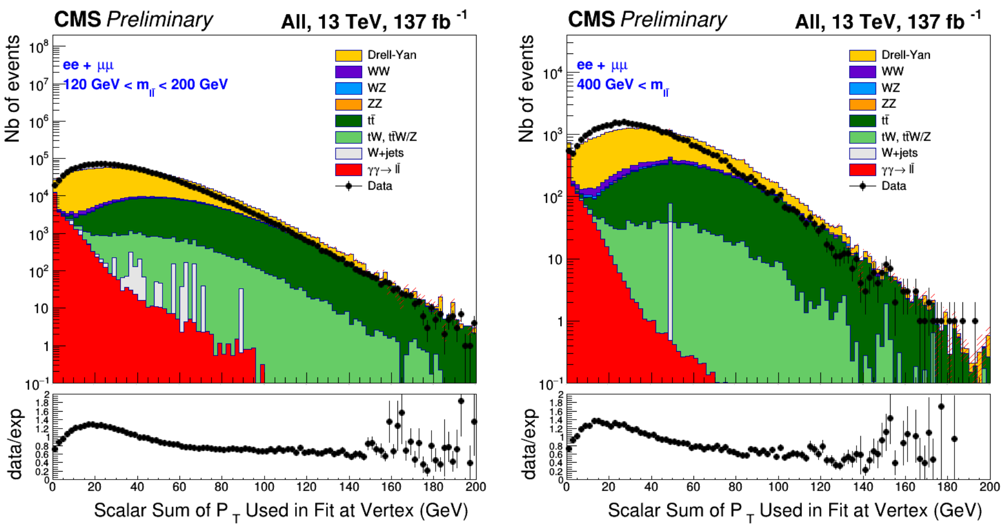
(b) Mass bin [400, 800] GeV.

Figure D.1 – The lepton pair p_T on invariant mass distributions before filters (Run II data).



(a) Mass bin [120, 200] GeV. *(b) Mass bin [400, 800] GeV.*

Figure D.2 – The acoplanarity distributions before filters (Run II data).

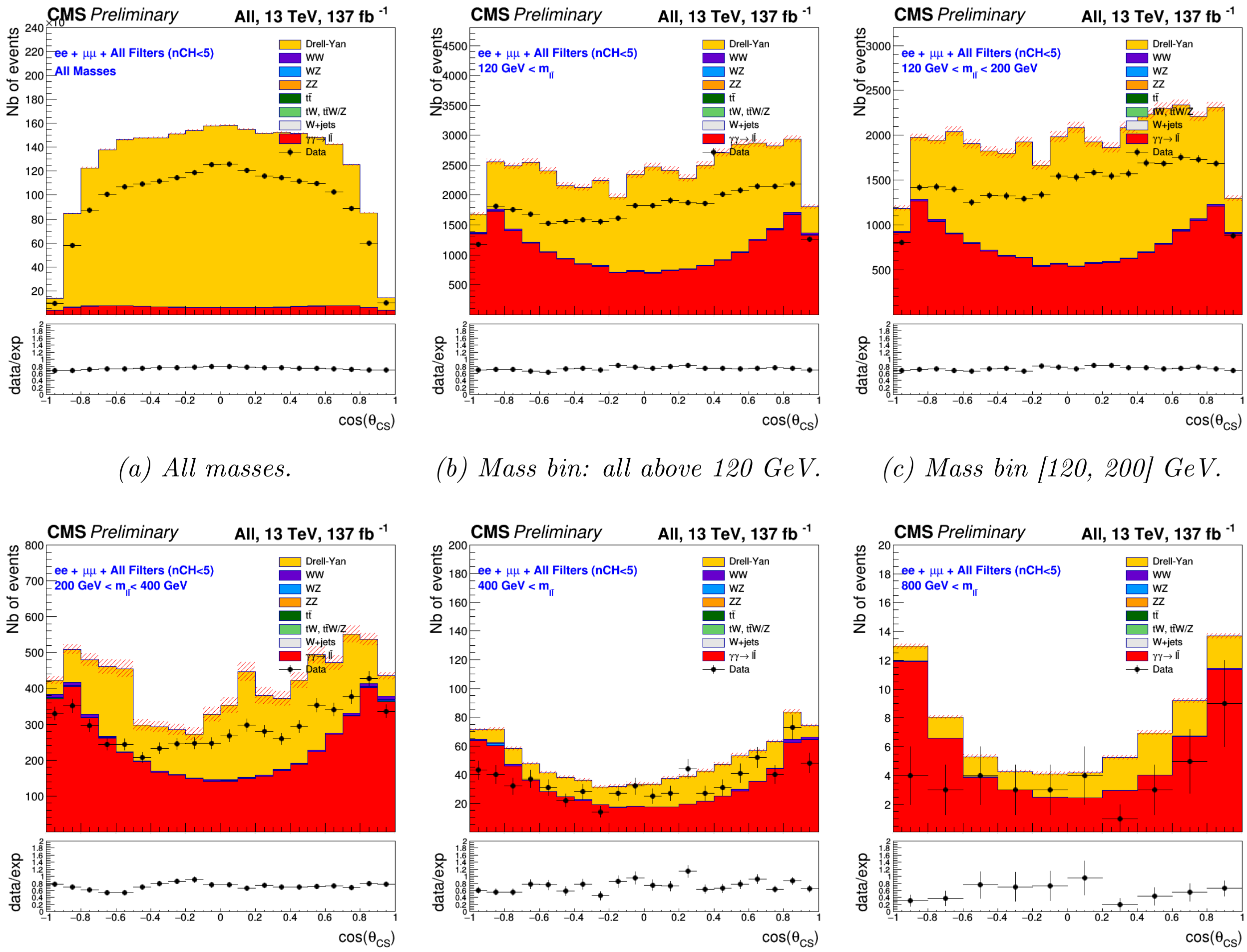


(a) Mass bin [120, 200] GeV. *(b) Mass bin [400, 800] GeV.*

Figure D.3 – The p_T of charged hadron in the vertex fit distributions before filters (Run II data).

D.2 The kinematics of the final state after filters

Figure D.4 displays the filtered $\cos\theta_{CS}$ distributions not weighted for nCH matching in the [70, 110] GeV mass bin. It indeed confirms that the nCH-weights derived in the analysis tend to improve the quality of the simulations in the isolation region.



(a) *All masses.* (b) *Mass bin: all above 120 GeV.* (c) *Mass bin [120, 200] GeV.*

(d) *Mass bin [200, 400] GeV.* (e) *Mass bin: all above 400 GeV.* (f) *Mass bin: all above 800 GeV.*

Figure D.4 – The unweighted (nCH) filtered $\cos\theta_{CS}$ distributions for different mass bins (Run II data).

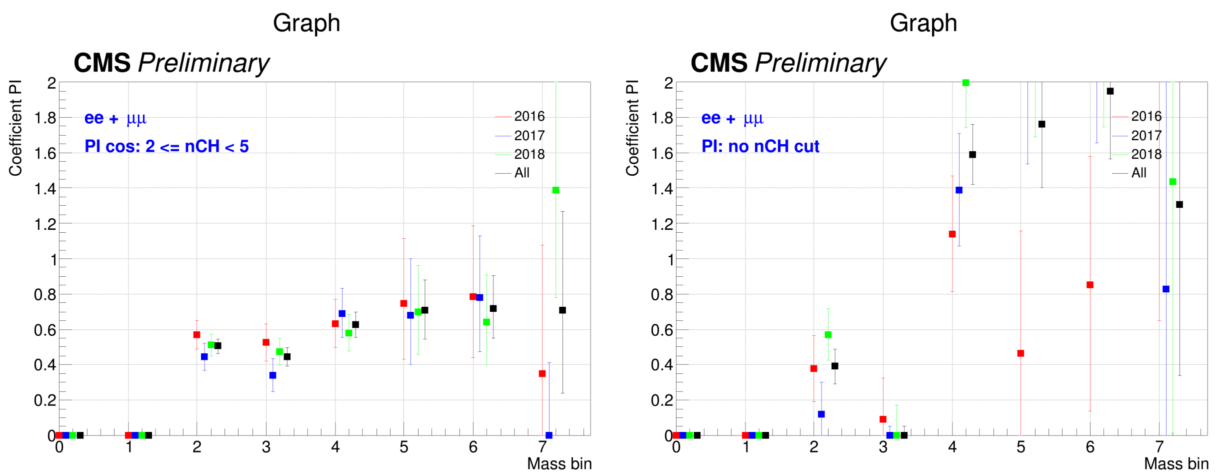
D.3 Additional fit weights results

Figure D.5 displays the overall PI fitted weights for all years of the $2 \leq$ nCH < 5 and no nCH cut. Figure D.6 displays the fitted parameters for both the PI and DY to the $\cos\theta_{CS}$ distributions in nCH < 2 and nCH < 5 configuration. Figure D.7 separates the PI fitted weights over all years between the different lepton configurations. Interestingly, the corrections for the electrons seem to be more important with weights on average much lower than the muons. The precision is obviously better for muons given the larger statistics than electrons (muons/electrons are present in ratio 3/1). Figure D.8 displays the fit parameters for the nCH procedure at different mass bins. Table D.1 summarises the PI fitted weights for the entire Run II data and table D.2 indicates the fractions of PI signal and DY in the data with the various cuts at two different mass bins.

ee			
Mass Interval	$\cos\theta_{CS}$ [nCH < 2]	$\cos\theta_{CS}$ [nCH < 5]	nCH
[60, 120] GeV	*excluded*	*excluded*	0.83± 0.04
[120, 200] GeV	0.55 ± 0.08	0.33 ± 0.07	0.60 ± 0.03
[200, 400] GeV	0.57 ± 0.11	0.61 ± 0.09	0.52 ± 0.05
[400, 800] GeV	0.60 ± 0.27	0.69 ± 0.23	0.70 ± 0.13
Above 800 GeV	0.56 ± 0.47	0.65 ± 0.36	0.39 ± 0.33

$\mu\mu$			
Mass Interval	$\cos\theta_{CS}$ [nCH < 2]	$\cos\theta_{CS}$ [nCH < 5]	nCH
[60, 120] GeV	*excluded*	*excluded*	0.87± 0.02
[120, 200] GeV	0.76 ± 0.04	0.64 ± 0.03	0.67 ± 0.02
[200, 400] GeV	0.75 ± 0.05	0.70 ± 0.05	0.63 ± 0.03
[400, 800] GeV	0.53 ± 0.14	0.59 ± 0.12	0.58 ± 0.07
Above 800 GeV	0.00 ± 0.30	0.32 ± 0.37	0.38 ± 0.22

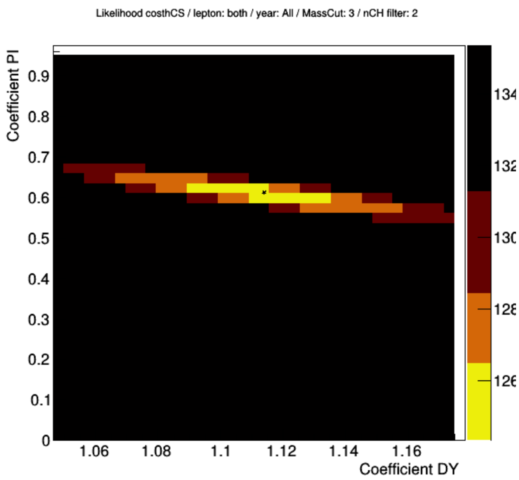
Table D.1 – The different mass intervals and the corrections for ee and $\mu\mu$ suggested by the various fitting variables for all years combined.



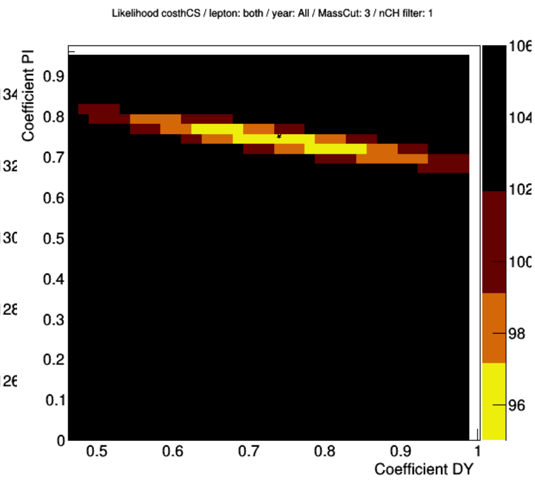
(a) $2 \leq nCH < 5$.

(b) No nCH cut.

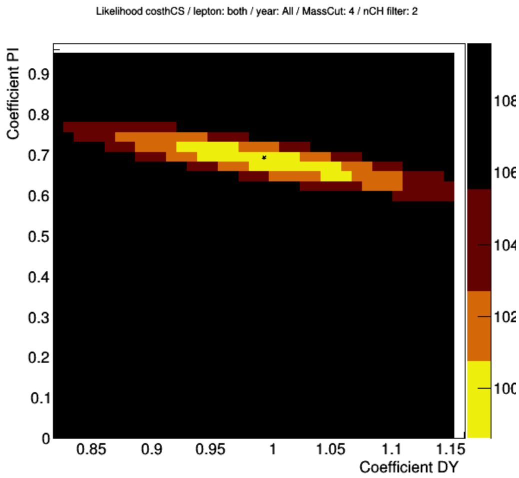
Figure D.5 – The PI fitting weights for the $\cos\theta_{CS}$ distributions for all mass bins (except 0 and 1), Run II data of ee + $\mu\mu$ events.



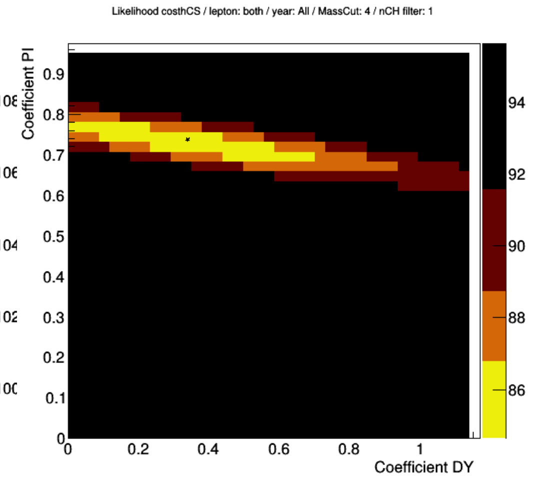
(a) Mass bin [120, 200] GeV.



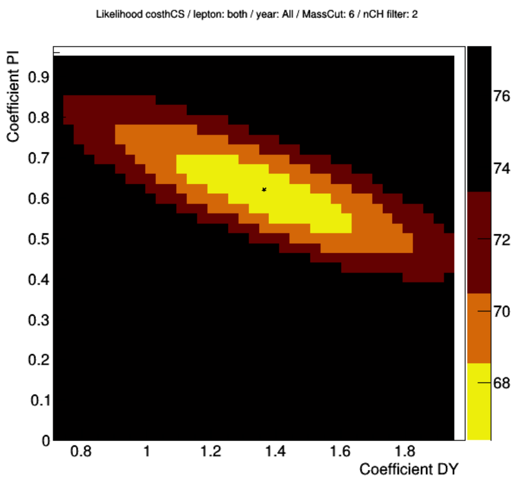
(b) Mass bin [120, 200] GeV.



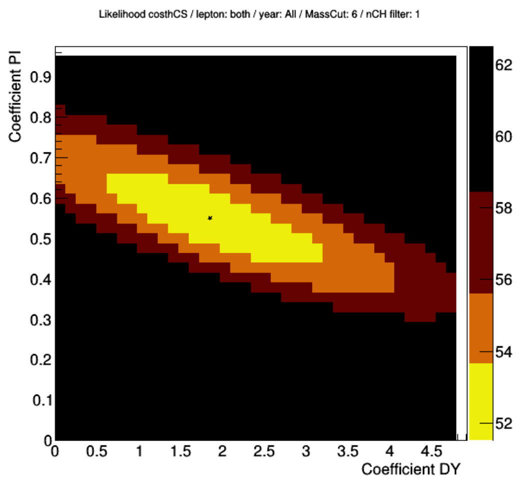
(c) Mass bin [200, 400] GeV.



(d) Mass bin [200, 400] GeV.

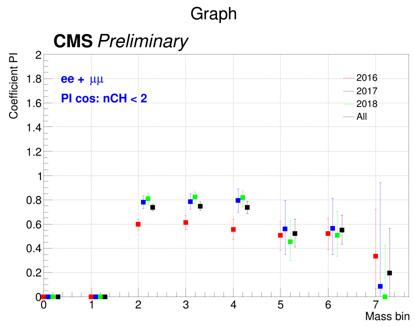


(e) Mass bin [400, 800] GeV.

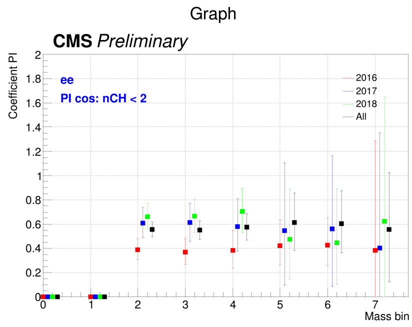


(f) Mass bin [400, 800] GeV.

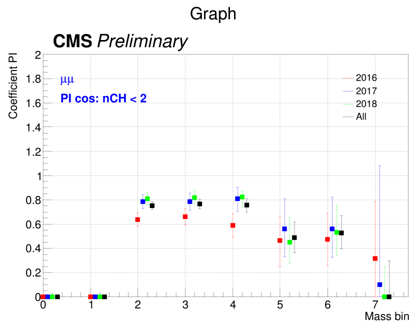
Figure D.6 – The parameters fitted in the different mass regions of the $\cos\theta_{CS}$ distributions (Run II data) for $nCH < 5$ (left) and $nCH < 2$ (right).



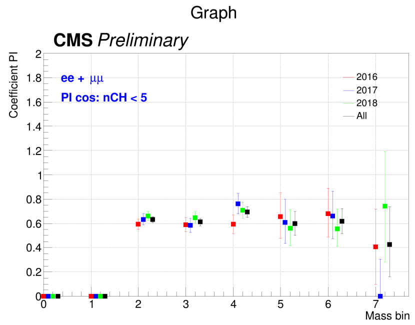
(a) $nCH < 2$ $ee + \mu\mu$.



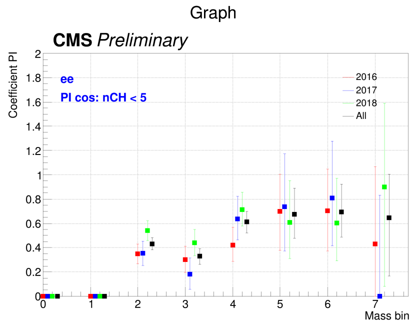
(b) $nCH < 2$ ee.



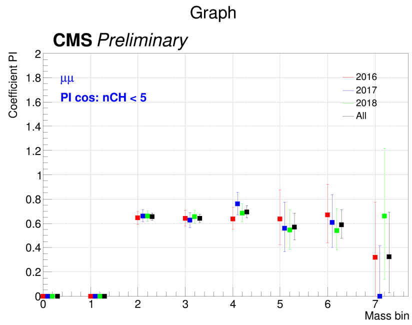
(c) $nCH < 2$ $\mu\mu$.



(d) $nCH < 5$ $ee + \mu\mu$.

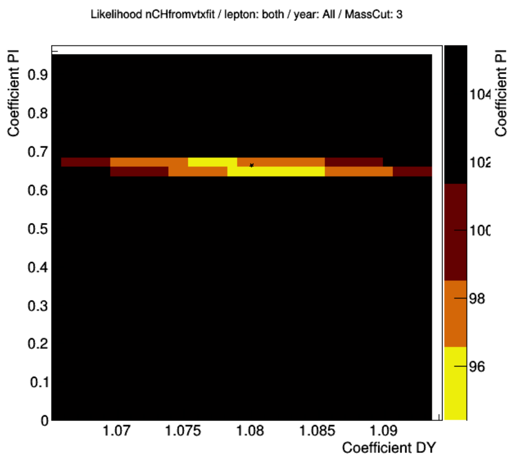


(e) $nCH < 5$ ee.

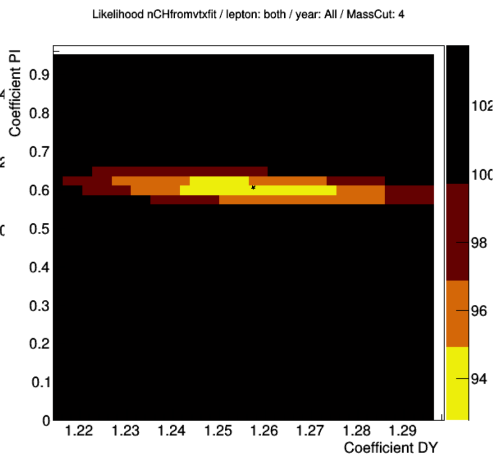


(f) $nCH < 5$ $\mu\mu$.

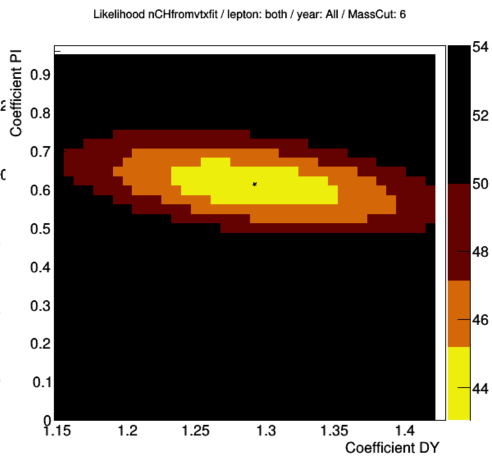
Figure D.7 – The PI fitting weights for the $\cos\theta_{CS}$ distributions for all mass bins (except 0 and 1), all years and different lepton configurations.



(a) Mass bin [120, 200] GeV.



(b) Mass bin [200, 400] GeV.



(c) Mass bin [400, 800] GeV.

Figure D.8 – The parameters fitted in the different mass regions of the nCH distributions (Run II data).

$ee + \mu\mu$: masses above 120 GeV					
Cut Type	Data	DY	PI	DY/Data	PI/Data
Original	$6.19\ 10^6$	$5.34\ 10^6$	$5.65\ 10^4$	86.3 %	0.91 %
Filtered	$5.37\ 10^6$	$5.01\ 10^6$	$4.70\ 10^4$	93.3 %	0.96 %
+ nCH < 5	$9.96\ 10^4$	$7.27\ 10^4$	$1.29\ 10^4$	73.3 %	24.7 %
+ nCH < 2	$2.13\ 10^4$	$8.83\ 10^3$	$1.29\ 10^4$	41.4 %	60.5 %

$ee + \mu\mu$: masses above 400 GeV					
Cut Type	Data	DY	PI	DY/Data	PI/Data
Original	$4.97\ 10^4$	$3.27\ 10^4$	$1.35\ 10^3$	65.8 %	2.72 %
Filtered	$3.38\ 10^4$	$2.90\ 10^4$	$1.20\ 10^3$	88.5 %	3.67 %
+ nCH < 5	$8.56\ 10^2$	$2.88\ 10^2$	$4.80\ 10^2$	33.7 %	56.1 %
+ nCH < 2	$2.99\ 10^2$	$3.45\ 10^1$	$2.64\ 10^2$	11.5 %	88.3 %

Table D.2 – The number of events in the data and in the weighted simulations (the latter being further weighted by set 2 for the nCH cuts) for two mass bins.