

Advanced Phenomenology for Indirect New Physics Searches

Présentée le 16 décembre 2022

Faculté des sciences de base
Laboratoire de physique théorique des particules
Programme doctoral en physique

pour l'obtention du grade de Docteur ès Sciences

par

Siyu CHEN

Acceptée sur proposition du jury

Prof. F. Mila, président du jury
Dr A. Wulzer, directeur de thèse
Dr R. Schöfbeck, rapporteur
Dr R. Torre, rapporteur
Prof. L. Zdeborová, rapporteuse

Acknowledgements

Not many people are as privileged as me to have the opportunity to do a PhD in Theoretical Physics. Though many children will answer “a scientist!”, to the question of what do you want to be in the future, many have not the chance to proceed as far as I have. When my father named me Siyu, which in Chinese means to think about the universe, he probably did not figure that in twenty seven years, I would actually finish a PhD in theoretical physics, a subject as close to what my name suggests as it can be. In any sense, I think that this PhD is a major milestone of my life, and I would like to dedicate the following short paragraphs to thank those who have helped and supported me.

My supervisor, Professor Andrea Wulzer, is a true physicist. He does not only possess a great deal of knowledge in the subject of particle physics, but also knows how to connect the three aspects of theory, phenomenology and experiment. His vision for this subject has greatly expanded mine, and without him I would not even be able to continue after my Masters. We decided to work on the intersection of Effective Field Theory and Machine Learning, and it has been an interesting and fulfilling journey since day one. I think that my work done does not even match one tenths of what he has taught me. In fact, during the entire PhD, Andrea has been teaching me without any reserve. I feel grateful that despite my ignorance and stubbornness, Andrea never gave me up. On a second note, when I was searching for a job in the industry, he has also been very supportive and encouraging, which I should also feel grateful towards as I am setting foot on a different journey.

My second biggest thanks goes to my collaborator Professor Giuliano Panico. During the four year I hardly saw Giuliano in person, but we have spent so much time on Skype discussing the project that I cannot put him any later in this acknowledgement. He has great insights to models beyond the Standard Model, and his attention to details has taught me a lot. I would say that without him the project would have been ten times more difficult. My virtual respect, Giuliano. Then I would also like to thank my colleague, now also a doctor in Theoretical Physics, Dr. Alfredo Glioti, as my closest office mate, my best collaborator, and my dearest friend during the past four years. Learning new topics was much easier and more fun with Alfredo, who is smart, patient, and extremely supportive. I cannot enumerate how many insider jokes we have made during the past four years, but every time we laugh together I feel true companionship. I believe that you will be a great physicist in the future, Alfredo, and I will be most proud to say that we have two papers together.

Acknowledgements

I was lucky to be surrounded by great minds. Professor Riccardo Rattazzi, to start with, exemplifies that a true master is able to explain complicated topics in simple words. Every lunch hour when he is around I learned something new, from concepts in physics to politics and philosophy. The most admirable thing that I find in him is that he really cares about his students. He goes extra miles to help students when necessary. I am really fortunate to be his group. Then there is my colleague, also a doctor now in Theoretical Physics, Lorenzo Ricci, who has helped me a lot in understanding important concepts, especially those in that Muon Collider paper. Lorenzo, thank you for being always encouraging to me. During my PhD I was also lucky to have interacted with Joao, Misha, Luca, Marco, Emmanuel Stamou, Marc, Matthijs, Gabriel, Aditya, Andrea Manenti, Jeanne, Sasha, Brian, Matth, Emmanuel Gendi, Gil, Miguel, Kamran, Eren and Filippo. During the quarantine time when it was difficult, Aditya tried to make me more social, and Gil gave me the first session of ski in life. Thank you for all you have done for me.

I reckon that this piece of work was done during a special time, two years of which was especially complicated due to the pandemic. I have to thank, apart from the names mentioned above, my friends who stayed around and supported me. It would have been impossible without you.

My last and most important words of gratitude goes to my parents and my grandma. I have not seen you for more than three years now because of the pandemic, but there is not a single day that I am not missing home. You never ask for anything but only the best for me. I only hope that this little piece of work will make you proud.

Dublin, November 20, 2022

Siyu

Abstract

Searching indirectly for physics beyond the standard model requires careful investigation of collider data and methodological advances that are the subject of this thesis. In the first part, we develop a multivariate analysis tool to compare data with the new physics prediction in their statistical distributions. The performances are demonstrated through the diboson production process at the Large Hadron Collider, but the method has general applicability. In the second half of the thesis, we study electroweak radiation effects using double log resummation on scattering cross-sections at a future muon collider at 10 TeV or more. We show that these effects play a crucial role in probing new physics, and we make progress towards their complete, accurate evaluation. Both parts of the thesis adopt a general Effective Field Theory description of new physics.

keywords: High energy physics, Effective Field Theory, Beyond the Standard Model, Multivariate Analysis, Neural Networks, Muon Collider

Riassunto

La ricerca indiretta della fisica oltre il modello standard richiede un'attenta indagine dei dati degli acceleratori di particelle e dei progressi metodologici che sono oggetto di questa tesi. Nella prima parte, sviluppiamo uno strumento di analisi multivariata per confrontare i dati con le previsioni di nuova fisica tramite le loro distribuzioni statistiche. Le prestazioni di tale metodo sono verificate prendendo come esempio la produzione di dibosoni al Large Hadron Collider, ma il metodo ha un'applicabilità generale. Nella seconda metà della tesi, studiamo gli effetti delle radiazioni elettrodeboli utilizzando la risommazione dei doppi logaritmi per il calcolo di sezioni d'urto a un futuro acceleratore di muoni a 10 TeV o più. Mostriamo che questi effetti svolgono un ruolo cruciale nello studio della nuova fisica e facciamo progressi verso la loro valutazione completa e accurata. Entrambe le parti della tesi adottano una descrizione generale della nuova fisica basate su teorie di campo efficace.

Parole chiave: Fisica delle Alte Energie, Teorie di Campo Efficaci, Oltre il Modello Standard, Analisi Multivariata, Reti Neurali, Muon Collider

Contents

Acknowledgements	i
Abstract	iii
Riassunto	v
1 Introduction	1
1.1 Effective Field Theory	1
1.2 Probing the Standard Model EFT	5
1.3 Statistical Tools	6
1.3.1 Neyman-Person Lemma	7
1.3.2 Matrix Element Methods and Binned Analysis Method	8
1.4 Neural Networks as Classifiers	9
1.4.1 Basics	9
1.4.2 The Learning Problem	12
1.4.3 Convergence to Distribution Ratio	14
2 Parametrized Classifiers for Optimal EFT Sensitivity	17
2.1 Quadratic Parametrization	17
2.2 The Physical Process	20
2.3 Analytic Approximation and Priors	23
2.3.1 Analytic approximation	23
2.3.2 Monte Carlo Generators	25
2.4 Optimality on Toy data	26
2.4.1 Results	28
2.4.2 MADGRAPH Leading Order	30
2.5 The reach at Next-to-Leading Order	31
2.5.1 Estimating the test statistics distributions	31
2.5.2 Results	34
2.6 Neural Network implementation and validation	35
2.6.1 The Quadratic Classifier	36
2.6.2 The Standard Classifier	38
2.6.3 Validation	39
2.7 Event Reweighting	40
	vii

Contents

2.7.1	The Parametrized Reweighted Classifier	42
3	Electroweak Radiation at Future Muon Colliders	45
3.1	Introduction	45
3.2	Theoretical predictions	48
3.2.1	IR Evolution Equations	48
3.2.2	Di-fermion production	59
3.2.3	Di-boson production	64
3.3	Sensitivity projections	67
3.3.1	W&Y operators	68
3.3.2	Diboson operators	72
3.3.3	BSM sensitivity	79
3.4	Conclusions and outlook	84
A	Formulation of the Quadratic Classifier	87
A.1	The general Quadratic Classifier	87
A.2	Minimization of the parametrized loss	88
B	Resummation of Radiation at the Muon Coller	91
B.1	Radiation Integrals	91
B.2	High-energy EW multiplets	94
B.3	3 rd family operators	98
B.4	Summary plots	99
	Bibliography	114
	Curriculum Vitae	115

1 Introduction

1.1 Effective Field Theory

Effective field theory (EFT) is a powerful strategy for studying Particle Physics. The idea of EFT is simple: when the full theory at the high energy regime is not directly accessible, we construct a low-energy theory that effectively gives the exact predictions for observables at the low-energy scale. EFT captures effects from the UV-complete theory, but it is formulated solely in terms of the degrees of freedom, in the form of quantum fields, that describe particles that are light enough to be produced in low-energy experiments.

A textbook example of EFT is the Fermi theory [1][2]. Before the discovery of the W boson or even today, when dealing with observables whose characteristic scale is much smaller than the W boson mass m_W at 80.385(15) GeV, the Fermi theory provides an adequate description of the theory of Weak interactions. In this theory, interactions emerge from 4-fermion vertices such as

$$-\frac{G_F}{\sqrt{2}} [\bar{\psi}_{\nu_\mu} \gamma^\alpha (1 - \gamma_5) \psi_\mu] [\bar{\psi}_e \gamma_\alpha (1 - \gamma_5) \psi_{\nu_e}] + h.c., \quad (1.1)$$

where ψ_{ν_μ} refers to the quantum field of the muon neutrino, ψ_μ that of the muon, ψ_e the electron and ψ_{ν_e} the electron neutrino. Similar four-fermion interactions emerge for other leptonic particles and for quarks.

G_F is the coupling strength of the interaction and is measured to be $1.663787 \times 10^{-5} \text{GeV}^{-2}$ [3]. On the other hand, the Standard Model (SM) theory, which is currently the established fundamental theory of Weak interactions [4], also accurately predicts the value of G_F . In the SM, W bosons mediate muon decays, as depicted in the right panel of fig.1.1, and G_F can be written in terms of the mass of the W boson m_W and the Weak coupling constant g_W ,

$$G_F = \frac{\sqrt{2}g^2}{8m_W^2}. \quad (1.2)$$

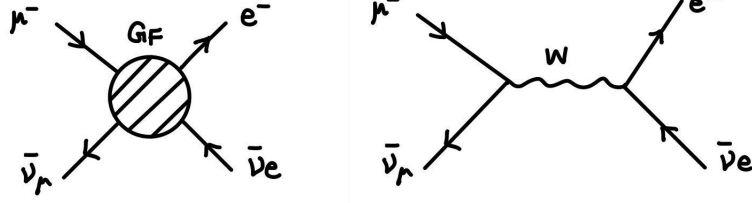


Figure 1.1: Feynman diagram of muon decay in the Fermi theory (left) and in the Standard Model (right)

The SM is the underlying theory for the Fermi theory, as constants in the latter can be computed using parameters in the former, but the other way around does not work. The formula above, eq.(1.2), is an example of “EFT matching”, where we employ the underlying fundamental theory to determine constants in the EFT. From this point of view, EFTs provide a convenient way to carry out computations in low-energy level experiments.

But EFTs are more important than that. When the underlying fundamental theory is unknown, EFT is the only sensible theoretical tool we have, as it was the case for the Fermi theory before the experimental discovery of W bosons and theoretical establishment of Weak interactions. Under this circumstance, eq.(1.1) is all what we know about the Weak interactions. The equation enables us to predict the manifestations of the Weak force by employing the four-fermion vertex as previously explained, but it also provides theoretical assurance of the existence of a more fundamental description of Weak interactions beyond the Fermi theory itself [5]. The argument goes as follows. Taking natural units with $\hbar, c = 1$, we see that each fermionic field carries an energy dimension of $3/2$, and the Lagrangian is of dimension 4, so the coupling strength G_F is naturally suppressed by E^2 , where E is the energy of the system. The amplitude $\mathcal{M}_{\text{Fermi}}$ is a dimensionless quantity, and from Feynman rules we deduce that $\mathcal{M}_{\text{Fermi}}$ is linear in G_F , therefore

$$\mathcal{M}_{\text{Fermi}} \sim G_F E^2, \quad (1.3)$$

suggesting that the amplitude grows quadratically with the energy. Now, we know that Weak interactions are perturbative, a property that bounds the amplitude by $16\pi^2$. This implies that one can expect an energy scale Λ_{Fermi} at which the Fermi theory ceases to be a valid description. Λ_{Fermi} is defined by

$$G_F \Lambda_{\text{Fermi}}^2 < 16\pi^2, \quad (1.4)$$

and is bounded at the TeV scale. This bound is much higher above the mass of W bosons. Still, it is a useful piece of information. The fact that this bound exists suggest that the Fermi theory is not the full picture, and a more fundamental theory awaits.

The SM is also an EFT, characterised by a maximal scale of validity Λ_{SM} . However, the existence of such a bound is not established simply by dimensional analysis as we saw above for the case of the Fermi theory. Manifestly, all the SM couplings have vanishing energy dimension, therefore no scattering amplitude grows with energy polynomially. Nevertheless, SM couplings run under radiative corrections according to the so-called "renormalisation group flow". While QCD is asymptotically free, the Electroweak couplings run into the problem of divergent gauge couplings in the UV, specifically at $\Lambda_{\text{QED}} \sim 10^{31}$ GeV, known as the Landau pole. This energy level is undoubtedly too far away from what we can experimentally probe with colliders at the moment. The situation can change if we include gravity into the picture and quantise the Einstein-Hilbert action. The amplitude of graviton interactions does grow quadratically with energy and by the same analysis a cutoff, lower than Λ_{QED} , of the order of the Planck scale $\sim 10^{19}$ GeV, emerges. While this is still too high to be accessible, other pieces of evidence suggests that new physics should emerge at a closer, foreseeable energy horizon, like the Dark Matter abundance and neutrino masses.

The strongest argument to look for new physics at a much lower energy scale, accessible to present or future collider is, in fact, quantitative and motivated by the Higgs mass. If we carefully examine the Standard Model Lagrangian, we shall find that all operators are of dimension four except for the mass term for the Higgs. Suppose that the Standard Model has a scale Λ_{SM} as high as 10^{31} GeV, then the light mass of the Higgs m_H , measured to be 125 GeV, poses the question of fine tuning. Concretely, the Higgs mass receives loop corrections that involve SM particles,

$$m_H^2 = \int_0^{\Lambda_{\text{SM}}} dE \frac{dm_H^2}{dE} + \int_{\Lambda_{\text{SM}}}^{\infty} dE \frac{dm_H^2}{dE}. \quad (1.5)$$

The following Feynman diagrams estimate the first part of the integral,

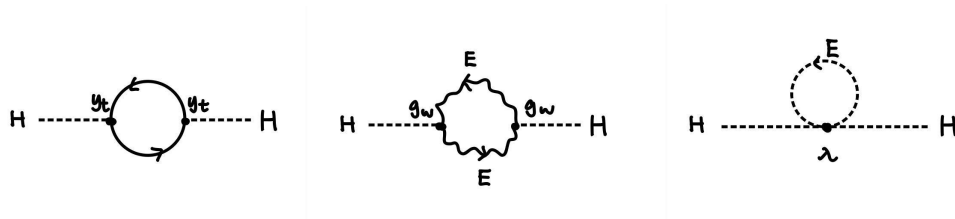


Figure 1.2: Loop corrections to Higgs mass within the Standard Model

and we can compute these contributions. Writing the first integral in eq.(1.5) as $\delta_{\text{SM}} m_H^2$, we see that it is dominated by the top quark loop, therefore

$$\frac{\delta_{\text{SM}}}{m_H^2} \sim \frac{3y_t^2}{8\pi^2} \Lambda_{\text{SM}}^2 = \left(\frac{\Lambda_{\text{SM}}}{645 \text{ GeV}} \right)^2, \quad (1.6)$$

with y_t the Yukawa coupling between the Higgs and the top quark.

This estimate follows dimensional analysis, as Λ_{SM}^2 must enhance the Higgs mass operator. Given that m_H is much smaller than Λ_{SM} , the only sensible way to construct a Higgs mass at the value we observe is to have a fine-tuned cancellation between the SM and the BSM parts. Cancellation of such high precision is *unnatural*; thus, we expect new physics to emerge at a much lower energy level. For example, one could hypothesize that the Higgs is a pseudo Nambu-Goldstone boson generated from the spontaneous symmetry breaking of a bigger symmetry group \mathcal{G} at the new physics resonance scale m_* [5]. In this scenario, m_* can be much higher than the current observed Higgs mass, analogous to the fact that the QCD scale being much higher than that of the π mesons, and a relatively small Higgs mass is conceivable. From this perspective, the naturalness problem motivates us to look for new physics around the corner, i.e. around the TeV energy scale, accessible by colliders.

The presence of new physics can be theoretically described the SM EFT, which is defined as follows. we take a set of BSM operators that have dimensional n , $\mathcal{O}^{(n)}$ with $n > 4$, and modify the SM Lagrangian as follows,

$$\mathcal{L}_{\text{BSM}} = \mathcal{L}_{\text{SM}} + \sum_i \frac{c_i}{\Lambda_{\text{SM}}} \mathcal{O}_i^{(5)} + \sum_j \frac{c_j}{\Lambda_{\text{SM}}^2} \mathcal{O}_j^{(6)} + \dots. \quad (1.7)$$

Specifically, the Wilson coefficients $c_{i,j}$ in front of the operators are now dimension-less. Interestingly, the only dimension-5 operator that is allowed by symmetries of the SM we can construct is the Weinberg operator, which predicts the neutrino mass and is measured to have a very small size. For the interest of this thesis, we are not going to study this operator, but only $\mathcal{O}^{(6)}$, which lead to major corrections next-in-line by naive dimensional analysis.

One of the many advantages of using EFT in studying new physics is that it is model independent. Investigating different UV theories reduces to studying the set of higher dimensional operators that respect symmetries of the Standard Model. For example, one can construct a basis for higher dimensional operators using the Strongly Interacting Light Higgs (SILH) scenario [6], assuming that the Higgs doublet can identify as the fields associated with spontaneous symmetry breaking. This basis accommodates several underlying UV models, including the Holographic composite Higgs model and the Little Higgs model with custodial symmetry. Another basis is the Warsaw basis [7], classified according to the number of fermion fields involved.

In summary, we see that the Standard Model is an EFT theory, and a UV complete theory should emerge at an energy scale much closer than the bound set by Standard Model gauge couplings, hinted by the light mass of the Higgs. To proceed in a model-independent fashion, we will study higher dimensional operators to probe new physics beyond the Standard Model.

1.2 Probing the Standard Model EFT

The Large Hadron Collider (LHC) is a powerful experiment for particle physics research. Two proton beams are guided to collide at the specific detector sites, with a centre-of-mass energy of 14 TeV. However, at the parton level, statistically speaking, at large rate partons in the scattering events only carry a small fraction of the total centre-of-mass energy (~ 100 GeV), and only very rarely their momentum scales up to TeV scale [8]. If new physics particles are above the TeV scale, colliders will not be able to discover them through direct resonance production. EFT then offers an alternative approach. By computing the cross-section in low-energy experiments using higher-dimensional operators, we see that new physics manifests itself in deviations of statistical distributions of observables from that predicted by the SM, including kinematic variables like momentum or scattering angles.

Following the discovery of the Higgs boson at the LHC in 2012, efforts have been made to measure its couplings to other SM particles with a high precision, as many BSM models can be probed through these couplings. Several EFT operators modify the Higgs couplings, therefore, these measurements, together with the future improvement thereof, can be used to probe the SM EFT. On the other hand, such experiments can exploit the large rate of collisions whose energy is relatively low, but high enough to produce Higgs bosons of mass ~ 125 GeV. Involving hadronic interactions, events at the LHC suffer from large systematic uncertainties due to non-perturbative nature of QCD and prominent experimental effects[9]. Large rate helps reducing the statistical error, while progress in the reduction of the systematic uncertainties, theoretical and experimental, is difficult. In view of this, Higgs coupling measurements are, or will be soon, intrinsically limited in their accuracy. However the LHC can also measure processes whose characteristic scale is above the EW scale of 100 GeV. On this type of processes, effects coming from the EFT operators are enhanced. Since EFT operators have dimensions $n > 4$, the Wilson coefficient is always suppressed by some positive powers of Λ_{SM} . Scattering amplitude with one or more insertions of vertices derived from these EFT operators will have to grow with energy. i.e. $\sim E^2/\Lambda_{\text{SM}}^2$, where E is the energy of the system. For example, at 13 TeV, probing through the Drell-Yan process at the LHC already achieves far better sensitivity than at the Large Electron-Positron (LEP) collider, whose primary interactions only involve the electroweak sector [10]. The advantage of

A particular process that benefits from effects growing with energy is the diboson production. However, the advantage of having BSM amplitudes $\sim E^2/\Lambda_{\text{SM}}^2$ is undermined by the "non-interference" theorem, which states that SM and BSM produce transverse bosons of different helicity, therefore killing the interference that contributes as the leading correction to the cross-section. Interestingly enough, it has been discovered that the interference term only disappears if we integrate over the phase space of the azimuthal angles $\phi_{1,2}$ related to the boson decays, and can be restored if $\phi_{1,2}$ are measured [11]. As an example, in $W\gamma$ production process, by examining the distributions of $\phi_{1,2}$, sensitivity to the two dimension-6 trilinear gauge couplings improves. As was pointed out in this thesis, multivariate analysis which includes both $\phi_{1,2}$ and other discriminating kinematic variables, such as polar angle $\theta_{1,2}$ and

scattering angle Θ can further enhance the sensitivity. This conclusion is a major motivation for our work in Chap.2. Before jumping into the details of our method, we will briefly present the existing statistical tools and our method in later parts of this chapter.

Apart from hadron colliders, searches for new physics continue beyond the LHC. EFT remains a powerful method also for future leptonic colliders, for example, the muon collider, which is projected to have muons with a centre-of-mass energy of 10 TeV, targeting at a wide range of dimension-6 BSM operators through multiple channels [12][13][14]. The most salient advantage of a muon collider is that all of its centre of mass energy is readily available for relevant collisions, as compared to the scenario for protons at the LHC. For the muon collider, sophisticated statistical tools are not yet urgent, as it still remains a question to sort out the relevant observables and to compute them. A careful examination of radiation effects, on the other hand, is necessary. When investigating the reach on new physics parameters in this setup, we notice that the precision in measurement is clouded by electroweak radiation corrections at such a high energy level [15]. A detailed discussion on using double log resummation to quantify radiation corrections will therefore follow in Chap.3.

1.3 Statistical Tools

To develop necessary statistical tools to study SM EFT at the LHC, we focus on the statistical problem of determining Wilson coefficients. The problem can be viewed either as a problem of regression on the value of c , or a hypothesis testing for or against a non-vanishing value of c . Either way, we establish confidence on how much our experiments signal the breakdown or the validity of the SM. The problem can be addressed with a classical or Bayesian statistical approach. However, we soon realise that the most important ingredient in carrying out such analysis, the likelihood of data given a certain Wilson coefficient, is not easy to obtain. Accurate theoretical predictions on the cross-sections have a limited degree of reality, because the LHC is a complicated machine. The only additional powerful tool we have at hand is the Monte Carlo generators that employ an enormous number of "latent" variables in modelling the theoretical and experimental processes to generate data that are close to reality, without relying on a closed form expression of the probability distribution functions of the underlying processes. Thus, obtaining an accurate form of the likelihood can be seen as the "inverse" problem to simulation.

Therefore, we discuss our approach in solving this "inverse" problem of reconstructing the likelihood function using simulated data, and dedicate the whole of Chap.2 to study on a specific physical channel. The approach is the frequentist setup of hypothesis testing, one corresponding to the SM and the other to SM EFT, or BSM models, parametrised by a Wilson coefficient. Notice, however, that the reconstructed likelihood can still be used in both ways to assist the study EFT, as was discussed in the last paragraph. The reason to choose to focus on hypothesis testing will become clear soon, but in essence it is the fact that we can use the Neyman-Person lemma (sec.(1.3.1)) to build the notion of optimality of sensitivity. We

will present a simplified example, whose underlying probability function is known, and use Neyman-Pearson test statistics as a concrete figure of merit for assessing the quality of other approximation techniques, which will be discussed in this section and sec.(1.4).

1.3.1 Neyman-Person Lemma

Consider the Standard Model as the null hypothesis H_0 and the addition of a set of EFT operators as the alternative hypothesis H_1 , we can design a statistical test on the observed sample data to decide which hypothesis is preferred. Generally, to reject the null hypothesis, one computes the p-value on the probability distribution function f of some test statistic t in the null hypothesis, defined as

$$p = \int_{t_{obs}}^{\infty} f(t|H_0) dt, \quad (1.8)$$

with t_{obs} referring to t computed on observations. One checks if the p-value is smaller than a specific criterion α . Normally, α is taken to be 0.05, corresponding roughly to 2 standard deviations from the mean in Gaussian distributions. In general, the test statistic can be chosen to be any variable derived from the sample. In a simplified BSM scenario, one can choose to only check the centre of mass energy s , and plot the distribution of s according to the null hypothesis. If the observed s is much higher than what the Standard Model predicts, concretely, with a p-value smaller than 0.05, one can say that the null hypothesis is rejected. Obviously, in a more complicated setup, such a choice is not optimal.

The Neyman-Pearson lemma states that the optimal test statistic is unique and is the log likelihood ratio between the two hypotheses,

$$t_{NP} = \log \frac{L(\mathbf{x}|H_1)}{L(\mathbf{x}|H_0)}, \quad (1.9)$$

where \mathbf{x} refers to observables of the sample, which is either obtained in experiments or more viably, with simulations. Optimality in this statement refers to maximum power for a fixed size or significance level. In a scattering event, the likelihood can be written in terms of the differential cross-section $d\sigma(\mathbf{x}|H)$. On a set of observations $\mathbf{x} = \{\mathbf{x}_1, \mathbf{x}_2 \dots \mathbf{x}_n\}$ of n events,

$$L(\mathbf{x}|H) = \frac{N^n e^{-N}}{n!} \prod_{i=1}^n p(\mathbf{x}_i|H) \quad (1.10)$$

with N the predicted number of events, computed as the product of luminosity and total cross-section $\sigma(H) = \int d\sigma(\mathbf{x}|H)$. The p.d.f. $p(\mathbf{x}|H)$ of event \mathbf{x} in hypothesis H is the normalised differential cross-section, written as

$$p(\mathbf{x}|H) = \frac{1}{\sigma(H)} \frac{d\sigma(\mathbf{x}|H)}{d\mathbf{x}}, \quad (1.11)$$

and consequently for the test statistics t_{NP}

$$t_{NP} = \sum_i^n \log \frac{d\sigma(\mathbf{x}_i|H_1)}{d\sigma(\mathbf{x}_i|H_0)} - N(H_1) + N(H_0). \quad (1.12)$$

In practice, one should compute the Neyman-Pearson test statistic t_{NP} to check if the null hypothesis can be rejected or not. However, as was discussed in sec.(1.2), the closed form of the likelihood $L(\mathbf{x}|H)$ is not available, or often termed “intractable” in the literature [16]. It has to do with the fact that observables \mathbf{x} evolve from parton-level momenta \mathbf{z} through inevitable experimental effects like showering and detector interactions, which is impossible to model with an analytical form.

To overcome this difficulty, several methods have been developed in recent years to approximate the log likelihood ratio $t_{NP}(\mathbf{x})$, and the goal is to come up with a reasonably accurate estimate of this quantity, preferably with the help of simulation data and without resorting to hand-picked functions in modelling the evolution from \mathbf{z} to \mathbf{x} .

1.3.2 Matrix Element Methods and Binned Analysis Method

The Matrix Element Method (MEM) is a well-established method in approximating the likelihood required for optimal sensitivity [17]]. In principle, in a scattering event, the different cross-section can be expressed in the following form,

$$d\sigma(p_i^{\text{vis}}|H) = \sum_{k,l} \int dx_1 dx_2 \frac{f_k(x_1) f_l(x_2)}{2s x_1 x_2} \left(\prod_{j \in \text{inv.}} \int \frac{d^3 p_j}{(2\pi)^3 2E_j} \right) |\mathcal{M}_{kl}(p_i^{\text{vis}}, p_j|H)|^2 \quad (1.13)$$

where p_i^{vis} refers to the measured momenta of the observed particles, p_j refers to that of the invisible particles, and $f(x)$ are parton distribution functions. The matrix element \mathcal{M}_{kl} should take into account the translation from parton-level momenta \mathbf{z} to observables \mathbf{x} , QCD radiation, hadronisation effects, detector interactions and so on. In reality, the matrix element \mathcal{M}_{kl} is approximated by the tree-level matrix element and a set of hand-picked transfer functions that model the rest of the process [18], [19], [20]. Efforts have also been put into extending the tree-level matrix element to the next-to-leading (NLO) order [21], [22] and jet substructure [23].

If we had access to the hidden processes beyond the tree-level matrix element, eq.(1.13) would serve as an optimal device to compute the Neyman-Pearson test statistic. However, the matrix element \mathcal{M}_{kl} can only be approximated and effects from NLO, hadronisation as well as detector interactions are hard to model. One simple technique to apply the Matrix Element

Method is to use the scattering amplitude computed at the tree-level and integrate over all the hidden variables, but that obviously will result in a crude approximation.

Another commonly used method to approximate the likelihood is the Binned Analysis. The idea is to select a discriminating variable and study its histogram. For each hypothesis, suppose that this variable follows a certain distribution, then we can compute the number of expected events that fall inside each pre-selected bins. Denote the bins as $\{b_1, b_2, \dots, b_s\}$ and the number of expected events by $\{N_1, N_2, \dots, N_s\}$, then the likelihood of a sample data-set n_1, n_2, \dots, n_s is

$$L(\mathbf{n}|H) = \prod_{i=1}^s \frac{N_i(H)^{n_i} e^{-N_i}}{n_i!}, \quad (1.14)$$

which is to say if the data is indeed generated in the H hypothesis, the number of events that fall inside each bin, n_i , will be equal to the number of expected events N_i up to some statistical error and hence maximise the likelihood. We can approximate t_{NP} using the log ratio of this likelihood computed in the test and null hypotheses, denoted as t_{bins}

$$t_{\text{bins}} = \log \frac{L(\mathbf{n}|H_1)}{L(\mathbf{n}|H_0)} = \sum_i n_i \log \frac{N_{i,1}}{N_{i,0}} - N_{i,1} + N_{i,0}. \quad (1.15)$$

A good candidate for such a choice of discriminating variables can be the centre of mass energy or the scattering angle. It is also common to apply some quality cuts, for example on the transverse momentum, in order to gain more sensitivity.

The limitation of this method is obvious. Since only one variable is chosen, information in other variables is lost. One can perhaps consider a Binned Analysis in two or more dimensions, or *multivariate* Binned Analysis, however, as formula eq.(1.14) suggests, the validity of the likelihood resides in enough events falling inside each bin so that one can approximate the probability of observing n_i events as a Poisson distribution with mean N_i . Increasing the dimensionality of the analysis implies that the size of the sample data-set has to multiply exponentially in order to populate each bin in a statistically satisfying way. It is practically only reasonable to carry out analysis in 2 and perhaps 3 dimensions, but generically going beyond is impossible. A further comparison of this method against neural-network based multivariate analysis will be presented in the next subsection.

1.4 Neural Networks as Classifiers

1.4.1 Basics

In recent years, Machine Learning has emerged as a popular method to study statistical distributions. Machine Learning has its applications extend in multiple directions, as diverse as to zip code recognition [24], cancer prediction [25] and sentiment analysis in Tweets [26],

but at the core of the method is the approximation of the underlying statistical distribution, or in other words, statistical inference.

Deep neural network is suitable for problems where the underlying statistical distribution is complicated, primarily because it has more parameters than traditional methods [27]. A more mathematical statement of the approximation capabilities of deep neural networks is presented in the Universal Approximation theorem [28]. On the practical side, the success in the realm of image-processing related applications [29], [30] & [31] have also advanced its development. We choose to apply the method of deep neural networks as we understand that the underlying physical process associated with the problem we study is of a complicated nature that requires a high-level of abstraction, and the signal borne inside the data-set is indirect, as was suggested earlier in this chapter.

In the basic supervised learning setup, deep neural networks consist of layers of neurons connected in a customised way. Each layer l takes its multivariate input $x_i^{(l)}$, linearly combines the input with the weight matrix $w_{ij}^{(l)}$ and bias $b_i^{(l)}$, and passes the result through an activation function σ so as to create an output $s_i^{(l)}$. Explicitly, in the case of a fully-connected neural network,

$$s_i^{(l)} = \sigma(x_i^{(l)} \cdot w_i^{(l)} + b_i^{(l)}) \quad (1.16)$$

Popular activation functions include the Sigmoid function, rectifier and Gaussian, and they account for the non-linearity of the final output. Practically, the network can be seen as a function of the input with weights and biases as its parameters, i.e. $f(\mathbf{x}; \mathbf{w}, \mathbf{b})$. This is called the forward pass.

The output of the last layer is compared to some labels y , and the discrepancy is computed as the loss $l(y, f(\mathbf{x}; \mathbf{w}, \mathbf{b}))$. Using the chain rule, one can then compute the derivative of the loss with respect to the individual parameters of the model,

$$\frac{\partial l}{\partial s_i^{(l)}} = \frac{\partial l}{\partial x_i^{(l)}} \sigma'(s_i^{(l)}), \quad \frac{\partial l}{\partial x_j^{(l-1)}} = \sum_i \frac{\partial l}{\partial s_i^{(l)}} w_{ij}^{(l)} \quad (1.17)$$

$$\frac{\partial l}{\partial w_{ij}^{(l)}} = \frac{\partial l}{\partial s_i^{(l)}} x_j^{(l-1)}, \quad \frac{\partial l}{\partial b_i^{(l)}} = \frac{\partial l}{\partial s_i^{(l)}}. \quad (1.18)$$

This is called the backward pass, or in some literature back propagation. This procedure in the algorithm allows us to decrease the loss numerically, in an iterative way. To achieve that, the basic idea is to update the parameters by *stepping down*,

$$w_{ij}^{(l)} \leftarrow w_{ij}^{(l)} - \eta \frac{\partial l}{\partial w_{ij}^{(l)}} \quad b_i^{(l)} \leftarrow b_i^{(l)} - \eta \frac{\partial l}{\partial b_i^{(l)}}. \quad (1.19)$$

The hyper parameter η here is called the learning rate. It is usually chosen to be a constant between 0 and 1, or in most modern practices, a dynamically changing parameter that adjusts the speed of stepping down. This procedure, called gradient step, can be understood in the analogy of mountain climbing. At each point of the journey, one judges the gradient of the mountain locally, and takes a step against the gradient. Each cycle of the three procedures: forward pass, back pass and gradient step, is called an epoch, and during training, a deep neural network will repeat the procedure for epochs after epochs until some convergence criterion is met.

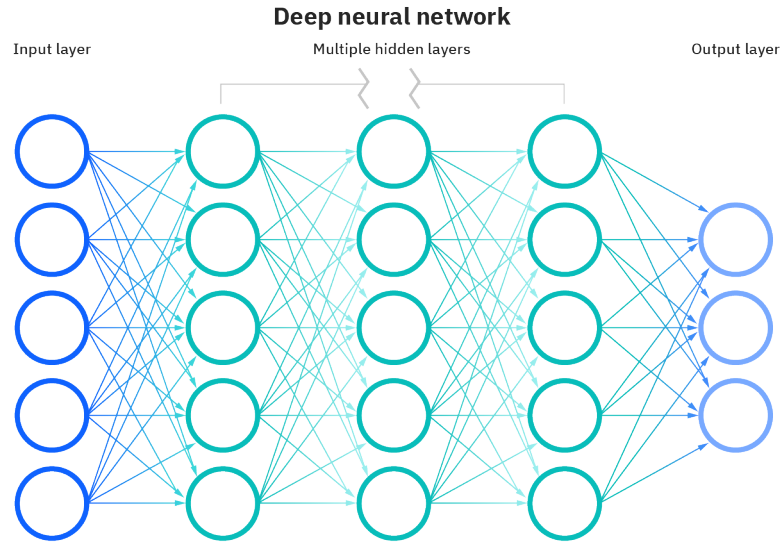


Figure 1.3: Illustration of a fully-connected neural network [32]

Theoretically, gradient stepping only guarantees that after iterations and iterations the algorithm will find a local minimum, not a global one, but there are many techniques studied and developed to help achieve the goal of approaching the global minimum, such as stochastic sampling [33], adaptive optimisation [34] and random drop-out [35].

An important concept to discuss in the engineering of the neural network model is over-fitting. One of the commonly used convergence criterion used for supervised learning is to compute the loss of the model against a test sample, i.e., on a sample data-set that is not seen by the neural network when updating its parameters, and stop training if this loss stops to decrease. Over-fitting refers to the situation where the model has substantially lower loss inside the training sample than the test sample, which implies that the model is more tuned for the noise than the actual signal of the underlying problem. The battle for higher sensitivity of models essentially is a battle of less over-fitting. Details of this struggle will be elaborated in the next

two chapters.

1.4.2 The Learning Problem

To use the neural network in solving the "inverse" problem of simulation, we need to first formulate the learning problem. The physical knowledge of the H_0 and H_1 models gets translated into Monte Carlo generator codes, which allow us to estimate $\sigma_{0,1}(X)$ and to produce samples, $S_{0,1}$, of artificial events following the $\text{pdf}(x|H_{0,1})$ distributions. More precisely, the Monte Carlo generates weighted events $e = (x_e, w_e)$, with x_e one instance of x and w_e the associated weight. If the w_e 's are not all equal, x_e does not follow the pdf of x and the expectation value of the observables $O(x)$ has to be computed as a weighted average. We choose the normalization of the weights such that they sum up to $\sigma_{0,1}(X)$ over the entire sample

$$\sum_{e \in S_{0,1}} w_e = \sigma_{0,1}(X). \quad (1.20)$$

With this convention, the weighted sum of $O(x_e)$ approximates the integral of $O(x) \cdot d\sigma_{0,1}(x)$ on $x \in X$. Namely

$$\sum_{e \in S_{0,1}} w_e O(x_e) \xrightarrow{\text{LS}} \int_{x \in X} d\sigma_{0,1}(x) O(x) = \sigma_{0,1}(X) E[O|H_{0,1}], \quad (1.21)$$

in the Large Sample (LS) limit where $S_{0,1}$ are infinitely large. We will see below how to construct an estimator $\hat{r}(x)$ for $r(x)$ (or, in short, to fit $r(x)$) using finite S_0 and S_1 samples.

For tree-level Monte Carlo generators the previous formulas could be made simpler by employing unweighted samples where all the weights are equal. However radiative corrections need to be included for sufficiently accurate predictions, at least up to next-to-leading (NLO) order in the QCD loop expansion. NLO in QCD simply means allowing 1-loop corrections to the hard scattering process of interest. NLO generators can only produce weighted events, and some of the events have a negative weight due to the subtraction method it employs (detailed discussion on negative weights on NLO can be found in sec.(2.7)). Therefore the NLO Monte Carlo samples cannot be rigorously interpreted as a sampling of the underlying distribution. However provided they consistently obey the LS limiting condition in eq. (1.21), they are equivalent to ordinary samples with positive weights for most applications, including the one described below.

The estimator $\hat{r}(x)$ can be obtained by solving the most basic Machine Learning problem, namely supervised classification with real-output Neural Networks (see Ref. [36] for an in-depth mathematical discussion). One considers a Neural Network acting on the kinematical variables and returning $f(x) \in (0, 1)$. This is trained on the two Monte Carlo samples by

minimizing the loss function

$$L[f(\cdot)] = \sum_{e \in S_0} w_e [f(x_e)]^2 + \sum_{e \in S_1} w_e [1 - f(x_e)]^2, \quad (1.22)$$

with respect to the free parameters (called weights and biases) of the Neural Network. The trained Neural Network, $\hat{f}(x)$, is in one-to-one correspondence with $\hat{r}(x)$, namely

$$\hat{f}(x) = \frac{1}{1 + \hat{r}(x)} \Leftrightarrow \hat{r}(x) = \frac{1}{\hat{f}(x)} - 1. \quad (1.23)$$

The reason why $\hat{r}(x)$, as defined above, approximates $r(x)$ is easily understood as follows. If the Monte Carlo training data are sufficiently abundant, the loss function in eq. (1.22) approaches its Large Sample limit and becomes

$$L[f(\cdot)] \xrightarrow{\text{LS}} \int_{x \in X} d\sigma_0(x) [f(x)]^2 + \int_{x \in X} d\sigma_1(x) [1 - f(x)]^2. \quad (1.24)$$

Furthermore if the Neural Network is sufficiently complex (i.e. contains a large number of adjustable parameters) to be effectively equivalent to an arbitrary function of x , the minimum of the loss can be obtained by variational calculus. By setting to zero the functional derivative of L with respect to f one immediately finds

$$\hat{f}(x) \simeq \frac{d\sigma_1(x)}{d\sigma_1(x) + d\sigma_0(x)} = \frac{1}{1 + r(x)} \Rightarrow \hat{r}(x) \simeq r(x). \quad (1.25)$$

The same result holds for other loss functions such as the standard Cross-Entropy, which has been found in Ref. [37] to have better performances for EFT applications, or the more exotic “Maximum Likelihood” loss [38], which is conceptually appealing because of its connection with the Maximum Likelihood principle. We observed no strikingly different performances with the various options, but we did not investigate this point in full detail. In what follows we stick to the quadratic loss in eq. (1.22).

The simple argument above already illustrates the two main competing aspects that control the performances of the method and its ability to produce a satisfactory approximation of $r(x)$. One is that the Neural Network should be complex in order to attain a configuration that is close enough to the (absolute) minimum, $f(x) = 1/(1 + r(x))$, of the loss functional in eq. (1.24). In ordinary fitting, this is nothing but the request that the fit function should contain enough adjustable parameters to model the target function accurately. On the other hand if the Network is too complex, it can develop sharp features, while we are entitled to take the Large Sample limit in eq. (1.24) only if f is a smooth enough function of x . Namely we need f to vary appreciably only in regions of the X space that contain enough Monte Carlo points. Otherwise the minimization of eq. (1.22) brings f to approach zero at the individual points that belong to S_0 sample, and to approach one at those of the S_1 sample. This phenomenon, called overfitting, makes that for a given finite size of the training sample, optimal performances are obtained by balancing the intrinsic approximation error of the Neural Network against the

complexity penalty due to overfitting. A third aspect, which is extremely important but more difficult to control theoretically, is the concrete ability of the training algorithm to actually reach the global minimum of the loss function in finite time. This requires a judicious choice of the minimization algorithm and of the Neural Network activation functions.

The problem of fitting $r(x)$ is mathematically equivalent to a classification problem. A major practical difference however emerges when considering the level of accuracy that is required on $\hat{r}(x)$ as an approximation of $r(x)$. Not much accuracy is needed for ordinary classification, because $\hat{r}(x)$ (or, equivalently, $\hat{f}(x)$) is used as a discriminant variable to distinguish instances of H_0 from instances of H_1 on an event-by-event basis. Namely, one does not employ $\hat{r}(x)$ directly in the analysis of the data, but a thresholded version of $\hat{r}(x)$ that isolates regions of the X space that are mostly H_0 -like (r is large) or H_1 -like (r is small). Some correlation between $\hat{r}(x)$ and $r(x)$, such that $\hat{r}(x)$ is large/small when $r(x)$ is large/small, is thus sufficient for good classification performances. Furthermore the region where $r(x) \simeq 1$ is irrelevant for classification.

The situation is radically different in our case because the EFT operators are small corrections to the SM. The regions where the EFT/SM distribution ratio is close to one cover most of the phase-space, but these regions can contribute significantly to the sensitivity if they are highly populated in the data sample. Mild departures of $r(x)$ from unity should thus be captured by $\hat{r}(x)$, with good accuracy relative to the magnitude of these departures. Obviously the problem is increasingly severe when the free parameters of the EFT (i.e. the Wilson coefficients “ c ”) approach the SM value $c = 0$ and $r(x)$ approaches one. On the other hand it is precisely when c is small, and the EFT is difficult to see, that a faithful reconstruction of $r(x)$ would be needed in order to improve the sensitivity of the analysis. Limitations of using a single neural network to study the distribution ratio $\hat{r}(x)$ will be addressed in Chap.2. Improvement of the simple classifier is possible. For the rest of this Chapter, we will naively follow the single neural network approach outlined in this section.

1.4.3 Convergence to Distribution Ratio

To demonstrate the benefit of multivariate analysis and also to illustrate how a classifier can reconstruct the likelihood ratio sufficiently accurately, we present the following example on a toy distribution. Consider a scattering process in which the cross-section only depends on two kinematic variables: centre of mass energy E and the scattering angle θ ranging from 0 to π . Without paying reference to a specific underlying BSM model, we design a distribution function $d\sigma_{\text{Toy}}$ as follows,

$$d\sigma_{\text{Toy}}(E, \theta, g) = \frac{1}{20} \left(\frac{E}{1000} \right)^{-4.2} \left(1 + g \frac{E^2}{1000^2} \cos \theta + \frac{E^4}{1000^4} g^2 \right), \quad (1.26)$$

where g refers to the Wilson coefficient of a dimension-six new physics operator. In this demonstration, we take $g_1 = 0.1$ in the BSM hypothesis, and $g_0 = 0$ in SM. These choices

of the Wilson coefficient is only to showcase the differences of the various results we have presented so far in this chapter, and are not restricted to any BSM model. The key feature of this toy distribution is that it is an explicit example where the interference, or the leading order correction, linear in g , would disappear if the distribution is integrated over one of the kinematic variables, in this case, θ .

We compare the following methods with their respective approximation to the Neyman-Pearson test statistics. Specifically,

- t_N : approximation of the log likelihood by the log of the number of expected events. This is a crude approximation, but most direct. The test statistics using solely the information of N can be written as,

$$t_N = n \log \frac{N_{\text{BSM}}}{N_{\text{SM}}} \quad (1.27)$$

with $n(H)$ referring to the actual number of events in the sample, generated from a Poisson distribution whose parameter is given by the number of expected events N_H in hypothesis H . A simple estimate it being, when g is ~ 0.4 , this method will give ~ 0.05 p-value.

- $t_{E, \text{bins}}$: Single-variable binning in energy E . To have a considerable number of events to fall inside each bin, we chose the intervals to be $\{300, 350, 400, 500, 750, 1200, 4000\}$. Denote the expected number of events that fall inside each E bin by $N_{i,E, \text{bins}}$ and the actual number of events by $n_{i,E, \text{bins}}(H)$, where i refers to the index of the bin, we write the test statistics $t_{E, \text{bins}}$ as

$$t_{E, \text{bins}} = \sum_i \left(n_{i,E, \text{bins}} \log \frac{N_{i,E, \text{bins}}(\text{BSM})}{N_{i,E, \text{bins}}(\text{SM})} - N_{i,E, \text{bins}}(\text{BSM}) + N_{i,E, \text{bins}}(\text{SM}) \right). \quad (1.28)$$

- $t_{\theta, \text{bins}}$: Similar single-variable binning in the scattering angle θ . Two sets of bins were chosen, with intervals in $\{0, \pi/4, \pi/2, 3\pi/4, \pi\}$. The test statistic takes a similar form.
- $t_{E, \theta, \text{bins}}$: Two-variable binning in both E and θ . The intervals take the outer product of the two directions $\{300, 350, 400, 500, 750, 1200, 4000\}$ and $\{0, \pi/4, \pi/2, 3\pi/4, \pi\}$.
- t_{NN} : A simple neural network. The network consists of an input layer, two hidden layers of 8 neurons each and ReLU as the activation function, and lastly an output layer that uses the sigmoid function to ensure a correct range of the output. The data set is comprised of 2 million points generated using Monte Carlo. Instead of using the weight $w_{0,1}$ that sums up to respective total cross sections $\sigma_{0,1}$, the sample takes a corresponding number of SM and BSM events such that the ratio of the number of data points generated according to the SM distribution and BSM distribution is equal to the ratio of total cross sections in the two hypotheses. The network is trained for 1000 epochs with a learning rate of 0.001 using RMSProp. To obtain a statistically reliable quantification of the performance, 10 such networks were initialized and trained

independently, the average performance of which will be reported in the following table. An independent test set of 1 million points is generated to evaluate the performance. The output of the network is transformed to the log ratio of likelihood according to eq.(1.23), and the test statistics takes the form

$$t_{\text{NN}} = \sum_{x \in S_{0,1}} \log \left(\frac{1}{y(x)} - 1 \right). \quad (1.29)$$

- t_{ME} : Matrix element method using the actual distribution. In this toy model, the matrix element written in terms of the differential cross-sections is exact and we have the following test statistics,

$$t_{\text{ME}} = \sum_{\text{events} \in H} \log \frac{d\sigma(E, \theta, g_1)}{d\sigma(E, \theta, g_0)}. \quad (1.30)$$

We expect t_{ME} to achieve optimal probing sensitivity according to the Neyman-Pearson lemma. This is indeed the case in our example.

Performances on the toy distribution show that neural networks obtain comparable sensitivity to the ideal Neyman-Pearson test statistics given by the likelihood ratio, equivalently t_{ME} .

method	t_{N}	$t_{\text{E, bins}}$	$t_{\theta, \text{bins}}$	$t_{\text{E}, \theta, \text{bins}}$	t_{NN}	t_{ME}
p-value	0.45568(5)	0.247(6)	0.217(7)	0.077(3)	0.044(2)	0.037(2)

Table 1.1: Various approximation methods on the Neyman-Person test statistic

In summary, the advantage of using a neural network to build log likelihood estimates is obvious. The method is multivariate by nature, and will not depend on the choice of binning. For example, suppose that we choose the intervals of binning on E to be uniformly distributed between the valid range of this variable, namely from 300 to 4000, the sensitivity will be lost (~ 0.5). The essential requirement on successful binning comes from sufficient number of events inside each bin, and even with a conveniently optimised choice of bins, when the number of variables in binning grows, this requirement becomes harder and harder to meet. On the other hand, data sets used for neural network training can be easily scaled. We will discuss in detail in Chapter 2 the more advanced concept of a parametrised network and our work using such a tool in studying diboson process at the LHC.

2 Parametrized Classifiers for Optimal EFT Sensitivity

2.1 Quadratic Parametrization

We take off from the standard classifier introduced in Sec.(1.4.2) and move on to a more realistic physical process. Before deploying the model, we examine further how to take advantage of the leading term of BSM correction in the interference.

Barring special circumstances, the EFT prediction for the differential cross section is a quadratic polynomial in the Wilson coefficients.¹ If a single operator is considered, so that a single free parameter c is present and the SM corresponds to the value $c = 0$, the EFT differential cross section reads

$$d\sigma_0(x; c) = d\sigma_1(x) \{ [1 + c\alpha(x)]^2 + [c\beta(x)]^2 \}, \quad (2.1)$$

where $\alpha(x)$ and $\beta(x)$ are real functions of x . An estimator $\hat{r}(x, c)$ for the distribution ratio in the entire Wilson coefficients parameters space could thus be obtained as

$$\hat{r}(x, c) = [1 + c\hat{\alpha}(x)]^2 + [c\hat{\beta}(x)]^2, \quad (2.2)$$

from estimators $\hat{\alpha}(x)$ and $\hat{\beta}(x)$ of the coefficient functions $\alpha(x)$ and $\beta(x)$. Notice that eq. (2.1) parametrizes, for generic $\alpha(x)$ and $\beta(x)$, the most general function of x and c which is quadratic in c , which is always positive (like a cross section must be) and which reduces to the SM cross section for $c = 0$. The equation admits a straightforward generalization for an arbitrary number of c parameters, which we work out in Appendix A.1.

The estimators $\hat{\alpha}(x)$ and $\hat{\beta}(x)$ are obtained as follows. We first define a function $f(x; c) \in (0, 1)$, in terms of two neural networks n_α and n_β with unbounded output (i.e. $n_{\alpha,\beta} \in (-\infty, +\infty)$) up

¹The only exception is when the relevant EFT effects are modifications of the SM particles total decay widths. Also notice that the cross section is quadratic only at the leading order in the EFT perturbative expansion, which is however normally very well justified since the EFT effects are small. Higher orders could nevertheless be straightforwardly included as higher order polynomial terms.

to weight-clipping regularization), with the following dependence on c :

$$f(x, c) \equiv \frac{1}{1 + [1 + c n_\alpha(x)]^2 + [c n_\beta(x)]^2}. \quad (2.3)$$

Next, we consider a set $\mathcal{C} = \{c_i\}$ of values of c and we generate the corresponding EFT Monte Carlo samples $S_0(c_i)$. At least two distinct values of $c_i \neq 0$ need to be employed, however using more than two values is beneficial for the performances. Monte Carlo samples are also generated for the H_1 (i.e. $c = 0$) hypothesis, one for each of the $S_0(c_i)$ samples. These are denoted as $S_1(c_i)$ in spite of the fact that they are all generated according to the same $c = 0$ hypothesis. The samples are used to train the $n_{\alpha, \beta}$ Networks, with the loss function

$$L[n_\alpha(\cdot), n_\beta(\cdot)] = \sum_{c_i \in \mathcal{C}} \left\{ \sum_{e \in S_0(c_i)} w_e [f(x_e, c_i)]^2 + \sum_{e \in S_1(c_i)} w_e [1 - f(x_e, c_i)]^2 \right\}. \quad (2.4)$$

We stress that in the second term in the curly brackets, the function f is evaluated on the same value of $c = c_i$ that is employed for the generation of the $S_0(c_i)$ Monte Carlo sample which we sum over in the first term.

By taking the Large Sample limit for the loss function as in eq. (1.24), differentiating it with respect to n_α and n_β and using the quadratic condition (2.1), it is easy to show that the trained Networks \hat{n}_α and \hat{n}_β approach α and β , respectively. Namely

$$\hat{\alpha}(x) \equiv \hat{n}_\alpha(x) \simeq \alpha(x), \quad \hat{\beta}(x) \equiv \hat{n}_\beta(x) \simeq \beta(x). \quad (2.5)$$

More precisely, by taking the functional derivative one shows that the configuration $n_\alpha = \alpha$ and $n_\beta = \beta$ is a local minimum of the loss in the Large Sample limit. It is shown in Appendix A.2 that this is actually the unique global minimum of the loss.

It is simple to illustrate the potential advantages of the Quadratic Classifier, based on the analogy with the basic binned histogram approach to EFT searches. In that approach, the X space is divided in bins and the likelihood ratio is approximated as a product of Poisson distributions for the countings observed in each bin. Rather than $\hat{r}(x, c)$, the theoretical input required to evaluate the likelihood are estimates $\hat{\sigma}_0(b; c)$ for the cross sections integrated in each bin “b”. Employing the Standard Classifier approach to determine $\hat{r}(x, c)$ would correspond in this analogy to compute $\hat{\sigma}_0(b; c)$ for each fixed value of c using a dedicated Monte Carlo simulation. By scanning over c on a grid, $\hat{\sigma}_0(b; c)$ would be obtained by interpolation. Every EFT practitioner knows that this is a very demanding and often unfeasible way to proceed. Even leaving aside the computational burden associated with the scan over c , the problem is that the small values of c (say, $c = \bar{c}$) we are interested in probing typically predict cross sections that are very close to the SM value and it is precisely the small relative difference between the EFT and the SM predictions what drives the sensitivity. A very small Monte Carlo error, which in turn requires very accurate and demanding simulations, would be needed in order to be sensitive to these small effects. In the Standard Classifier method, the counterpart

of this issue is the need of generating very large samples for training the Neural Network. Furthermore, this should be done with several values of c for the interpolation. This approach is computationally demanding even when a single Wilson coefficient is considered, and it becomes rapidly unfeasible if c is a higher-dimensional vector of Wilson coefficients to be scanned over.

The strategy that is normally adopted in standard binned analyses is closely analog to a Quadratic Classifier. One enforces the quadratic dependence of $\sigma_0(b; c)$ on c as in eq. (2.1), and estimates the three polynomial coefficients (i.e. the SM cross section and the analog of α and β) in each bin by a χ^2 fit to $\hat{\sigma}_0(b; c)$, as estimated from the Monte Carlo simulations for several values of c . The values of c used for the fit are much larger than the reach of the experiment $c = \bar{c}$, so that their effects are not too small and can be captured by the Monte Carlo simulation. The Quadratic Classifier works in the exact same way. It can learn $\hat{\alpha}(x)$ and $\hat{\beta}(x)$ using training samples generated with large values of c , for which the difference between the $S_0(c)$ and $S_1(c)$ is sizable. The training can thus recognize this difference, producing accurate estimates of $\hat{\alpha}(x)$ and $\hat{\beta}(x)$. This accurate knowledge results in an accurate estimate of $\hat{r}(x, c)$ and of its departures from unity even at the small value $c = \bar{c}$, because our method exploits the exact quadratic relation in eq. (2.1).

It should be noted that the “Quadratic Classifier” introduced in eq. (2.3) is “Parametrized” in the sense that it encapsulates the dependence on the c parameters, but it is the exact opposite of the Parametrized Neural Network (or Parametrized Classifier) of Ref. [39]. In that case, the Wilson Coefficient c is given as an input to the Neural Network, which acts on an enlarged (x, c) features space. The purpose is to let the Neural Network learn also the dependence on c of the distribution ratio in cases where this is unknown. Here instead we want to enforce the quadratic dependence of the distribution ratio on c , in order to simplify the learning task.

An alternative strategy to exploit the analytic dependence on c is the one based on “morphing” [40]. Morphing consists in selecting one point in the parameter space for each of the coefficient functions that parametrize $d\sigma_0(x; c)$ as a function of c , and expressing $d\sigma_0(x; c)$ as a linear combination of the cross-sections computed at these points. For instance, a total of 3 “morphing basis points”, $c_{1,2,3}$, are needed for a single Wilson coefficient and quadratic dependence, and $d\sigma_0(x; c)$ is expressed as a linear combination of $d\sigma_0(x; c_{1,2,3})$. This rewriting can be used to produce two distinct learning algorithms.

The first option is to learn the density ratios $d\sigma_0(x; c_{1,2,3})/d\sigma_1(x)$ individually (one-by-one or simultaneously), by using training data generated at the morphing basis points $c_{1,2,3}$, and to obtain $\hat{r}(x, c)$ using the morphing formula. In the analogy with ordinary binning, this would correspond to extracting the dependence on c of the cross-sections by a quadratic interpolation of $\hat{\sigma}_0(b; c_{1,2,3})$ at the selected points. Of course it is possible to reconstruct the cross sections accurately also by using 3 very accurate simulations, rather than fitting less accurate simulations at several points. However a judicious choice of the values of $c_{1,2,3}$ is essential for a proper reconstruction of the quadratic and of the linear term of the polynomial.

For the former, it is sufficient to take c very large, but for the latter a value of c should be selected that is neither too large, such that the quadratic term dominates by too many orders of magnitude, nor too small such that the constant SM term dominates. Notice that the optimal c depends on the analysis bin because the EFT effects relative to the SM (and the relative importance of the quadratic and linear terms) can be vastly different in different regions of the phase space. With “plain” morphing as described above, we are obliged to employ only few values of c , which might not be enough to cover the entire phase space accurately. With the Quadratic Classifier instead, all values of c that are useful to learn the distribution in some region of the phase space can be included simultaneously in the training set. For example, for the following physical process of fully leptonic ZW decay, we will use

$$\begin{aligned} G_{\varphi q}^{(3)} &: \quad \{\pm 50, \pm 20, \pm 5\} \times 10^{-2} \text{ TeV}^{-2}, \\ G_W &: \quad \{\pm 20, \pm 10, \pm 5\} \times 10^{-2} \text{ TeV}^{-2}. \end{aligned} \quad (2.6)$$

Alternatively, one can use the morphing formula in place of eq. (2.1), producing a different parameterization of the classifier than the one in eq. (2.3), to be trained with values of the parameters that are unrelated with the morphing basis points. The parametrization employed in the Quadratic Classifier is arguably more convenient, as it is simpler, universal and bounded to $f \in (0, 1)$ interval owing to the positivity of eq. (2.1). Importantly, also the condition $\hat{r}(x, 0) \equiv 1$ is built-in in the Quadratic Classifier. However this could be enforced in the morphing parameterization as well by selecting $c = 0$ as one of the basis points. If this is done, we do not expect² a degradation of the performances if employing the morphing-based parametrization rather than ours. Indeed, we believe that the key of the success of the Quadratic Classifier that we observe in this section stems from the appropriate choice of the values of c used for training, and not from the specific parametrization we employ. The non-optimal performances of the morphing strategy observed in Ref. [40] (on a different process than the one we study) are probably to be attributed to a non-optimal choice. Further investigations on this aspect are beyond the scope of the present thesis.

2.2 The Physical Process

We consider ZW production at the LHC with leptonic decays, namely $Z \rightarrow \ell^+ \ell^-$ and $W \rightarrow \ell \nu$, where $\ell = e, \mu$. As this is arguably the simplest process, of established EFT relevance, where a multivariate approach is justified and potentially improves the sensitivity. We focus on the high-energy tail of the process, with a selection cut on the transverse momentum of the Z-boson, $p_{T,Z} > 300$ GeV, because of two independent reasons. First, because at high energy we can approximate the differential cross section analytically and define a realistic enough Toy problem to assess the optimality of the method. Second, because at high-energy the statistics is sufficiently limited (less than 5×10^3 expected events at the HL-LHC, including

²Provided that the possibility of having f outside the $(0, 1)$ interval is not a problem when training, for instance, with the cross-entropy loss function.

both W charges) to expect systematic uncertainties not to play a dominant role. The reach we will estimate in the $p_{T,Z} > 300$ GeV region, on purely statistical bases, should thus be nearly realistic.

The high-energy regime, in spite of the relatively limited statistics, is the most relevant one to probe those EFT operators that induce energy-growing corrections to the SM amplitudes. There are only two CP-preserving and flavor-universal operators in the ZW channel that induce quadratically energy-growing terms, namely ³

$$\mathcal{O}_{\varphi q}^{(3)} = G_{\varphi q}^{(3)} \left(\bar{Q}_L \sigma^a \gamma^\mu Q_L \right) (i H^\dagger \overleftrightarrow{D}_\mu H), \quad \mathcal{O}_W = G_W \varepsilon_{abc} W_\mu^{a\nu} W_\nu^{b\rho} W_\rho^{c\mu}. \quad (2.7)$$

We thus focus on these operators in our analysis. The BSM relevance of these operators is well established in the literature. Searching for $\mathcal{O}_{\varphi q}^{(3)}$ in di-boson final states is among the most promising probes of Higgs compositeness at the HL-LHC [9]. The operator \mathcal{O}_W is generated with loop-suppressed coefficient in perturbative BSM models and in composite Higgs models of SILH [41] type. However it does not experience loop suppression for strongly-coupled BSM and is even potentially enhanced by the strong coupling in particular scenarios [42].

Both $\mathcal{O}_{\varphi q}^{(3)}$ and \mathcal{O}_W contribute to the ZW production amplitudes with quadratically energy-growing terms of order $G \cdot s$, where s is the center-of-mass energy squared of the diboson system. However the way in which this energy growth manifests itself in the cross section is rather different for the two operators (see e.g. Refs. [11, 9]). The $\mathcal{O}_{\varphi q}^{(3)}$ operator mainly contributes to the “00” helicity amplitude, in which the gauge bosons are longitudinally polarized. The SM amplitude in this channel is sizable and has a constant behavior with s at high energy. As a consequence, a sizable quadratically-growing interference term between the SM and the BSM amplitudes is present in the cross section. This happens even at the “inclusive” level, i.e. when only the hard scattering variables describing ZW production (and not the decay ones) are measured.

On the contrary, the \mathcal{O}_W operator induces quadratically-growing contributions only in the transverse polarization channels with equal helicity for the two gauge bosons (namely, ++ and --). In the SM this channel is very suppressed at high energy, since its amplitude decreases as m_W^2/s . Therefore in inclusive observables the interference between \mathcal{O}_W and the SM does not grow with the energy and is very small. In order to access (or “resurrect” [11]) the interference, which is the dominant new physics contribution since the Wilson coefficient of the operator is small, the vector bosons decay variables must be measured. We thus expect that the sensitivity to \mathcal{O}_W will benefit more from an unbinned multivariate analysis technique than the one on $\mathcal{O}_{\varphi q}^{(3)}$.

The relevant kinematical variables that characterize the four-leptons final state are defined as in Ref. [11] and depicted in Figure 2.1, where V_1 is identified with the Z and V_2 with the W boson. The figure displays the kinematics in the rest frame of the ZW system, obtained from

³We use the definition $H^\dagger \overleftrightarrow{D}_\mu H = H^\dagger D_\mu H - (D_\mu H)^\dagger H$.

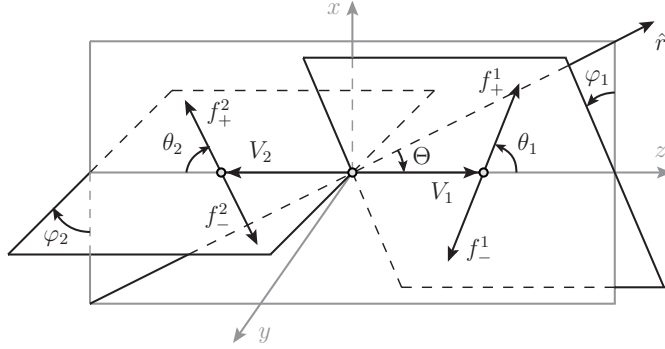


Figure 2.1: The kinematical variables in the “special” coordinate frame [11].

the lab frame by a boost along the direction of motion (denoted as \hat{r} in the figure) of the ZW pair, followed by a suitable rotation that places the Z along the positive z axis and \hat{r} on the x - z plane with positive x component. The “inclusive” variables associated with ZW production are the center-of-mass energy squared s and $\Theta \in [0, \pi]$, which is defined as the angle between \hat{r} and the Z-boson momentum. The decay kinematics is described by the polar and azimuthal decay angles $\theta_{1,2}$ and $\varphi_{1,2}$. The latter angles are in the rest frame of each boson and they are defined as those of the final fermion or anti-fermion with helicity $+1/2$ (e.g. the ℓ^+ in the case of a W^+ and the $\bar{\nu}$ for a W^-), denoted as $f_{+}^{1,2}$ in the figure.⁴ The remaining variables that are needed to characterize the four leptons completely are weakly sensitive to the presence of the EFT operators and can be ignored, with the exception of the total transverse momentum of the ZW system, $p_{T,ZW}$, which is a useful discriminant at NLO [9].

The variables described above are useful for the theoretical calculation of the cross section, but they cannot be used for our analysis because they are not experimentally accessible. The “measured” variables we employ are defined as follows. First, since we do not measure the neutrino (longitudinal) momentum, this needs to be reconstructed by imposing the on-shell condition for the W. The reconstructed neutrino momentum, rather than the true one, is used to define the kinematical variables and in particular s and Θ . Moreover, since we do not measure the helicity of the fermions but only their charge, the decay angles of the Z boson, denoted as θ_Z and φ_Z , are defined in terms of the charge-plus lepton rather than of the helicity plus lepton. Depending on the (unobserved) leptons helicities these angles are either equal to θ_1 and φ_1 , or to $\pi - \theta_1$ and $\varphi_1 + \pi$, respectively. The W decay angles, defined in terms of the lepton or the reconstructed neutrino depending on the charge of the W as previously explained, are denoted as θ_W and φ_W . In summary, the variables we employ in the analysis are

$$\{s, \Theta, \theta_W, \varphi_W, \theta_Z, \varphi_Z, p_{T,ZW}\}, \quad (2.8)$$

where of course $p_{T,ZW}$ is non-vanishing only at NLO.

The on-shell condition for the W boson has no real solution if the W-boson transverse mass is

⁴The correct definition of φ_2 appears in version four of Ref. [11].

larger than the W pole mass m_W . The neutrino is reconstructed in this case by assuming that the neutrino rapidity is equal to the one of the lepton. If instead the transverse mass is smaller than m_W , the condition has two distinct real solutions, each of which produces a different reconstructed kinematics. For our analysis we picked one of the two solutions at random on an event-by-event basis, while for the analysis of the actual data it would be arguably more convenient to duplicate the kinematical variables vector by including both solutions. Nothing changes in the discussion that follows if this second option is adopted.

2.3 Analytic Approximation and Priors

2.3.1 Analytic approximation

At the tree-level order, and based on the narrow-width approximation for the decays, it is easy to approximate the cross section analytically in the high-energy regime. The crucial simplification is that the reconstructed 3-momentum of the W boson (with any of the two solutions for the neutrino) becomes exact when the W is boosted, so that the reconstructed Θ and s variables approach the “true” ones of Figure 2.1. Notice that Θ is the angle between the Z and the direction of motion of the ZW system in the lab frame, which corresponds at tree-level to the direction of motion of the most energetic incoming parton. In the kinematical region we are interested in, the (valence) quark is more energetic than the anti-quark in more than 80% of the events. Therefore we can identify Θ as the angle between the Z and the u quark or the d quark in the $u\bar{d} \rightarrow ZW^+$ and $d\bar{u} \rightarrow ZW^-$ processes, respectively.

With these identifications, the non-vanishing on-shell helicity amplitudes $\mathcal{M}_{h_Z h_W}$ for the hard scattering process $u\bar{d} \rightarrow ZW^+$, at leading order in the high-energy expansion, read

$$\begin{aligned} \mathcal{M}_{00} &= -\frac{g^2 \sin \Theta}{2\sqrt{2}} - \sqrt{2} G_{\phi q}^{(3)} s \sin \Theta, & \mathcal{M}_{++} = \mathcal{M}_{--} &= \frac{3g c_W G_W s \sin \Theta}{\sqrt{2}}, \\ \mathcal{M}_{-+} &= -\frac{g^2 (s_W^2 - 3c_W^2 \cos \Theta)}{3\sqrt{2} c_W} \cot \frac{\Theta}{2}, & \mathcal{M}_{+-} &= \frac{g^2 (s_W^2 - 3c_W^2 \cos \Theta)}{3\sqrt{2} c_W} \tan \frac{\Theta}{2}, \end{aligned} \quad (2.9)$$

where g is the $SU(2)_L$ coupling, c_W and s_W are the cosine and the sine of the Weak angle. An overall factor equal to the cosine of the Cabibbo angle has not been reported for shortness. The amplitudes for the $d\bar{u} \rightarrow ZW^-$ process can be obtained from the ones above with the formal substitutions $\Theta \rightarrow -\Theta$ and $s_W^2 \rightarrow -s_W^2$. The amplitudes are non-vanishing only for left-chirality initial quarks. Notice that the above formulas depend on the conventions in the definition of the wave-function of the external particles, and that these conventions must match the ones employed in the decay amplitude for the consistency of the final results. The wave-function reported in Ref. [43] are employed.

Let us now turn to the vector bosons decays. The decay amplitudes assume a very simple form

in terms of the $\theta = \theta_{1,2}$ and $\varphi = \varphi_{1,2}$ variables, namely

$$\mathcal{A}_h = -\sqrt{2}g_V m_V e^{ih\varphi} d_h(\theta), \quad (2.10)$$

where h is the helicity of the decaying vector boson ($V = V_{1,2} = Z, W$) and $d_h(\theta)$ are the Wigner d -functions. The overall coupling factor g_V depends on the nature of the boson and, in the case of the Z , on the electric charge of the helicity-plus fermion it decays to. Specifically, $g_W = g/\sqrt{2}$ for the W , $g_Z = g_L = -g(1 - 2s_W^2)/2c_W$ if the Z decays to an helicity-plus ℓ^+ and $g_Z = g_R = g s_W^2/c_W$ if the Z decays to an helicity-plus ℓ^- . The two options for the helicity (which are physically distinct) correspond to two terms in the cross section. In the first one the Z decay amplitude is evaluated with the g_L coupling, with $\theta = \theta_1 = \theta_Z$ and $\varphi = \varphi_1 = \varphi_Z$. In the second one we employ g_R , $\theta = \theta_1 = \pi - \theta_Z$ and $\varphi = \varphi_1 = \varphi_Z + \pi$. There is no helicity ambiguity in the W -boson decay angles. However the reconstruction of the azimuthal decay angle is exact in the high-energy limit only up to a twofold ambiguity [11]. Namely the reconstructed φ_W approaches φ_1 on one of the two solutions for the neutrino (and we do not know which one), and $\pi - \varphi_1$ on the other. Since we are selecting one solution at random, we should average the cross section over the two possibilities $\varphi = \varphi_2 = \varphi_W$ and $\varphi = \varphi_2 = \pi - \varphi_W$ for the W azimuthal angle. The polar angle is instead $\theta = \theta_2 = \theta_W$ in both cases.

Production and decay are conveniently combined using the density matrix notation. We define the hard density matrix

$$d\rho_{h_Z h_W h'_Z h'_W}^{\text{hard}} = \frac{1}{24s} \mathcal{M}_{h_Z h_W} (\mathcal{M}_{h'_Z h'_W})^* d\Phi_{ZW}, \quad (2.11)$$

where $d\Phi_{ZW}$ is the two-body phase space and the factor $1/24s$ takes care of the flux and of the averages over the colors and the helicities of the initial quarks. The decay processes are instead encoded into decay density matrices. The one for the Z -boson, including the sum over the ℓ^\pm helicities as previously explained, reads

$$d\rho_{h_Z h'_Z}^Z = \frac{1}{2m_Z \Gamma_Z} \left[\mathcal{A}_{h_Z} \mathcal{A}_{h'_Z}^* \Big|_{g_L, \theta_Z, \varphi_Z} + \mathcal{A}_{h_Z} \mathcal{A}_{h'_Z}^* \Big|_{g_R, \pi - \theta_Z, \varphi_Z + \pi} \right] d\Phi_{\ell^+ \ell^-}, \quad (2.12)$$

where Γ_Z is the Z decay width. For the W , since we average on the neutrino reconstruction ambiguity, we have

$$d\rho_{h_W h'_W}^W = \frac{1}{2m_W \Gamma_W} \frac{1}{2} \left[\mathcal{A}_{h_W} \mathcal{A}_{h'_W}^* \Big|_{\frac{g}{\sqrt{2}}, \theta_W, \varphi_W} + \mathcal{A}_{h_W} \mathcal{A}_{h'_W}^* \Big|_{\frac{g}{\sqrt{2}}, \theta_W, \pi - \varphi_W} \right] d\Phi_{\ell\nu}. \quad (2.13)$$

The complete partonic differential cross section is finally simply given by

$$d\hat{\sigma} = 4 \sum d\rho_{h_Z h_W h'_Z h'_W}^{\text{hard}} d\rho_{h_Z h'_Z}^Z d\rho_{h_W h'_W}^W, \quad (2.14)$$

where the sum is performed on the four helicity indices and the factor of 4 takes into account the decay channels into electrons and muons.

2.3.2 Monte Carlo Generators

For our analysis we use three Monte Carlo generators, of increasing accuracy.

The first one is the Toy generator that implements the analytic approximation of the cross section in eq. (2.14), with the hard amplitudes expanded up to order $G \cdot s$ in the EFT contribution and up to order s^0 in the SM term, as in eq. (2.9). This implies, in particular, that in the Toy Monte Carlo all the mixed transverse/longitudinal helicity channels vanish exactly, that only the $\pm\mp$ and 00 channels are retained in the SM and that new physics is just in the 00 and $\pm\pm$ channels for $\mathcal{O}_{\varphi q}^{(3)}$ and \mathcal{O}_W , respectively. The Toy Monte Carlo employs a simple fit to the ($u\bar{d}$ or $d\bar{u}$) parton luminosities obtained from the nCTEQ15 [44] PDF set (implemented through the ManeParse [45] Mathematica package). The variable s is sampled according to the parton luminosity, while all the other variables are sampled uniformly. The cut $p_{T,Z} = \sqrt{s}/2 \sin\Theta > 300$ GeV is implemented at generation level. Since the analytical distribution is extremely fast to evaluate, this basic approach is sufficient to obtain accurate Monte Carlo integrals and large unweighted event samples in a very short time.

The second generator is MADGRAPH [46] at LO, with the EFT operators implemented in the UFO model of Ref. [47]. We simulate the $2 \rightarrow 4$ process $pp \rightarrow \mu^+ \mu^- e \nu_e$, with the Z and the W decaying to opposite flavor leptons for a simpler reconstruction, and multiply the resulting cross section by 4. The cut on $p_{T,Z}$, defined as the sum of the μ^+ and μ^- momenta, is imposed at generation level, as well as the cuts

$$m_{T,ev} \leq 90 \text{ GeV}, \quad 70 \text{ GeV} \leq m_{\mu\mu} \leq 110 \text{ GeV}, \quad (2.15)$$

on the transverse mass of the virtual W and the invariant mass of the virtual Z . These are needed to suppress non-resonant contributions to the production of the 4 leptons. Standard acceptance cuts on the charged leptons are also applied. The unweighted events obtained with MADGRAPH are further processed to compute the kinematical variables in eq. (2.8) after neutrino reconstruction, as detailed at the beginning of this section.

The MADGRAPH LO generator is slightly more accurate than the Toy one. It contains all the ZW helicity amplitudes and no high-energy approximation. Furthermore, it describes non-resonant topologies and off-shell vector bosons production, which affects the reconstruction of the neutrino and in turn the reconstruction of the Z and W decay variables [11]. Nevertheless on single-variable distributions the Toy Monte Carlo and the LO one agree reasonably well, at around 10%.

The third and most refined generator is MADGRAPH at NLO in QCD, interfaced with PYTHIA 8.244 [48, 49] for QCD parton showering. The complete $2 \rightarrow 4$ process is generated like at LO, but no cuts could be applied at generation level apart from default acceptance cuts on the leptons and the lower cut on $m_{\mu\mu}$ in eq. (2.15). At NLO, the cut on $p_{T,Z}$ needs to be replaced with the cut $p_{T,V} > 300$ GeV, with $p_{T,V} = \min[p_{T,Z}, p_{T,W}]$. This cut suppresses soft or collinear vector boson emission processes, which are insensitive to the EFT. In order to populate the $p_{T,V} > 300$ GeV

tail of the distribution with sufficient statistics, events were generated with a bias. The bias function was equal to one for $p_{T,V}$ above 290 GeV, and equal to $(p_{T,V}/290 \text{ GeV})^4$ below. The momenta of the charged leptons and the transverse momentum of the neutrino in the generated events were read with MADANALYSIS [50] and the kinematical variables in eq. (2.8) reconstructed like at LO. The cut $p_{T,V} > 300 \text{ GeV}$ and the remaining cuts in eq. (2.15) were imposed on the reconstructed events. The total cut efficiency on the Monte Carlo data, thanks to the bias, was large enough (around 17%) to allow for an accurate prediction of the cross section and for the generation of large enough event samples.

Even if ours is an electroweak process, it is known that NLO QCD corrections can in principle affect significantly the sensitivity to the EFT operators. Relevant effects are related with the tree-level zero [51] in the transverse amplitude, which is lifted at NLO, and with the appearance of same-helicity transverse high-energy amplitudes due to real NLO radiation [52]. All these effects are properly modeled by the MADGRAPH NLO generator.

2.4 Optimality on Toy data

Our goal is to reconstruct the EFT-over-SM cross section ratio $r(x, c)$ as accurately as possible 1.4.2. Since r is known analytically for the Toy problem, a simple qualitative assessment of the performances could be obtained by a point-by-point comparison (see Figures 2.7 and 2.8) of $r(x, c)$ with its approximation $\hat{r}(x, c)$ provided by the trained Neural Network. However a point-by-point comparison is not quantitatively relevant, since the level of accuracy that is needed for $\hat{r}(x, c)$ can be vastly different in different regions of the phase-space, depending on the volume of expected data and on the discriminating power of each region (i.e. on how much r is different from one).

The final aim of the entire construction is to obtain an accurate modeling of the extended log-likelihood ratio in eq. (1.12), to be eventually employed in the actual statistical analysis. A quantitative measure of the r reconstruction performances is thus best defined in terms of the performances of the final analysis that employs \hat{r} , instead of r , in the likelihood ratio. Among all possible statistical analyses that could be carried out, frequentist tests to the EFT hypothesis $H_0(c)$ (regarded as a simple hypothesis for each given value of c), against the SM one, H_1 , are considered for the illustration of the performances.

Four alternative test statistic variables are employed. One is the standard Poisson binned likelihood ratio (see below). The others are unbinned and take the form

$$t_c(\mathcal{D}) = N(X|H_0) - N(X|H_1) - \sum_{i=1}^{\mathcal{N}} \tau_c(x_i), \quad (2.16)$$

where $\tau_c(x)$ is either equal to the exact $\log[r(x, c)]$ or to $\log[\hat{r}(x, c)]$, as reconstructed either with the Standard Classifier or with the Quadratic Classifier described in Section 1.4.2 and 2.1, respectively. In each case the probability distributions of t in the two hypotheses are computed

with toy experiments (or with the simpler strategy of Section 2.5.1), and used to estimate the expected (median) exclusion reach on c at 95% Confidence Level if the SM hypothesis is true. In formulas, the 95% reaches ($c_{2\sigma}$) we quote in what follows are solutions to the implicit equation

$$p(t_{\text{med}}(c_{2\sigma}); c_{2\sigma}) = 0.05, \text{ with } t_{\text{med}}(c) = \text{Median}[t_c(\mathcal{D})|H_1], \quad (2.17)$$

where the p -value is defined as

$$p(t_c; c) = \int_{t_c}^{\infty} dt'_c \text{pdf}(t'_c|H_0(c)). \quad (2.18)$$

The two Wilson coefficients $c = G_{\varphi q}^{(3)}$ and $c = G_W$ are considered separately. Therefore the results that follow are single-operator expected exclusion reaches.

Summarizing, the four methodologies we employ are

i) *Matrix Element (ME)*

In this case we set $\tau_c(x) = \log[r(x, c)]$ in eq. (2.16), with r computed analytically using eq. (2.14). Therefore t coincides with the log-likelihood ratio λ in eq. (1.12), which in turn is the optimal discriminant between H_0 and H_1 due to the Neyman–Pearson lemma [53]. Namely, a straightforward application of the lemma guarantees that by employing $t = \lambda$ as test statistic we will obtain the optimal (smallest) $c_{2\sigma}$ reach, better than the one we could have obtained using any other variable. The Matrix Element Method is thus optimal in this case, and the optimality of the other methods can be assessed by comparing their $c_{2\sigma}$ reach with the one of the Matrix Element.

ii) *Standard Classifier (SC)*

The second method consists in setting $\tau_c(x) = \log[\hat{r}(x, c)]$ in eq. (2.16), with \hat{r} reconstructed by the Standard Classifier as in Section 1.4.2. Notice that a separate training is needed to reconstruct $\hat{r}(x, c)$ for each value of the Wilson Coefficient. Therefore computing $c_{2\sigma}$, as defined in eq. (2.17), requires scanning over c , performing first the Neural Network training and next the calculation of the distributions of t by toy experiments. For the Quadratic Classifier (and for the Matrix Element Method), the first step is not needed. The details on the Neural Network architecture and training, and of its optimization, will be discussed in Section 2.6.

iii) *Quadratic Classifier (QC)*

The third approach is to employ $\hat{r}(x, c)$ as reconstructed by the Quadratic Classifier of Section 2.1. Implementation details are again postponed to Section 2.6, however it is worth anticipating that the key for a successful reconstruction is to train using values for the Wilson coefficients that are significantly larger than the actual reach. Specifically, we used eq.(2.6) in the beginning of this chapter.

These values have been selected as those that induce order one departures from the SM cross section in the low, medium and high regions of the $p_{T,Z}$ distribution. If willing to compute cross-section in each $p_{T,Z}$ region by quadratic interpolation, using the values selected with this criterion can be shown to maximize the accuracy on the reconstruction of the linear term, while still allowing for a good determination of the quadratic term. We expect this choice to be optimal for the Quadratic Classifier training as well. Also notice that the total number of training Monte Carlo events is the same one (6M, see Section 2.6) employed for each of the separate trainings performed on the Standard Classifier.

iv) *Binned Analysis (BA)*

Finally, in order to quantify the potential gain of the unbinned strategy, we also perform a binned analysis. The test statistic in this case is provided by the sum over the bins of the log-ratio of the SM over EFT Poisson likelihoods, with the expected countings as a function of the Wilson coefficients computed from Monte Carlo simulations. The test statistic distributions, and in turn the reach by eq. (2.17), are computed with toy experiments like for the other methods and no asymptotic formulas are employed.

For both $G_{\varphi q}^{(3)}$ and G_W we considered 3 bins in $p_{T,Z}$, with the following boundaries

$$p_{T,Z}[\text{GeV}] : \quad \{300, 500, 750, 1200\} \text{ GeV}. \quad (2.19)$$

The $p_{T,Z}$ variable is an extremely important discriminant because it is sensitive to the energy growth induced by the EFT. The three bins are selected based on the studies in Refs. [11, 9], and a narrower binning has been checked not to improve the sensitivity significantly. A cut $\cos\Theta < 0.5$ is imposed in the analysis targeting $G_{\varphi q}^{(3)}$, because this helps [9] in isolating the longitudinal helicity channel thanks to the amplitude zero in the transverse SM amplitudes. For G_W , no $\cos\Theta$ cut is performed, and each $p_{T,Z}$ is split in two bins, for $\cos 2\varphi_W$ larger and smaller than zero. This is sufficient to partially capture the leading EFT/SM interference term as discussed in Ref. [11].

Most likely the binned analysis could be improved by considering more (and/or better) variables and a narrower binning. However it should be noticed that the simple strategies described above already result from an optimization, targeted to the specific operators at hand, and that the reach we obtain is consistent with the sensitivity projections available in the literature.

2.4.1 Results

The results of the four methods are shown in Figure 2.2 (see also Table 2.1), together with the ones obtained with the MADGRAPH LO description of the ZW process, to be discussed in Section 2.4.2. The 2σ sensitivities reported in the figure are obtained by interpolating the median p -value as a function of the Wilson Coefficient c , in the vicinity of the reach, and computing $c_{2\sigma}$ by solving eq. (2.17). Further details on this procedure, and the associated

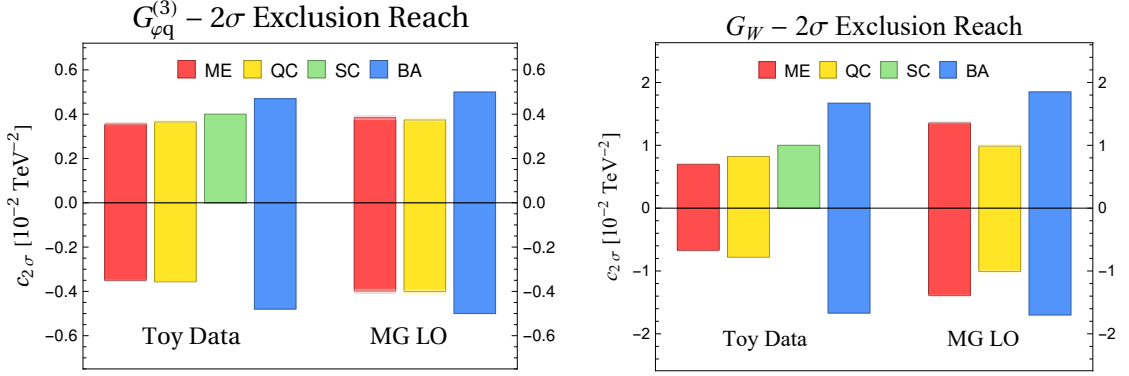


Figure 2.2: Expected exclusion reach on $G_{\varphi q}^{(3)}$ (left) and on G_W (right). The results are also reported in Table 2.1. Light-color stacked bars represent the errors.

error, are given in Section 2.5.1.

The figure reveals a number of interesting aspects. First, by comparing the Matrix Element reach with the one of the Binned Analysis we can quantify the potential gain in sensitivity offered by a multivariate strategy. The improvement is moderate (around 30%) for $G_{\varphi q}^{(3)}$, but it is more than a factor of 2 (of 2.4) in the case of the G_W operator coefficient. The different behavior of the two operators was expected on physical grounds, as discussed in details below. The figure also shows that the Quadratic Classifier is nearly optimal. More precisely, the reach is identical to the one of the Matrix Element for $G_{\varphi q}^{(3)}$, and $< 20\%$ worse for G_W . We will see in Section 2.6 that the residual gap for G_W can be eliminated with more training points than the ones used to produce Figure 2.2. Suboptimal performances are shown in the figure in order to outline more clearly, in Section 2.6, that our method is systematically improvable as long as larger and larger Monte Carlo samples are available.

Finally, we see in the figure that the Standard Classifier is slightly less sensitive than the Quadratic one, but still its performances are not far from optimality. This is reassuring in light of possible applications of Statistical Learning methodologies to different problems, where the dependence of the distribution ratio on the new physics parameters is not known and the Quadratic Classifier approach cannot be adopted. On the other hand, the Standard Classifier method is rather demanding. First, because it requires separate trainings on a grid of values of c , out of which the reach should be extracted by interpolation. In turn, this requires a much larger number of training points than the Quadratic Classifier, since at each point of the grid we use as many training points as those the Quadratic Classifier needs in total for its training. Second, because we observed hyperparameters optimization depends on the specific value of c that is selected for training. Because of these technical difficulties, we only report sensitivity estimates for the positive Wilson coefficients reach. Furthermore these estimates (see Table 2.1) are based on the p -value obtained at a given point of the c grid without interpolation. For the same reason, we did not try to apply the Standard Classifier methodology to the LO and to the NLO data and we focus on the Quadratic Classifier in what

follows.

Let us discuss now the physical origin of the different behaviors observed for the $\mathcal{O}_{\varphi q}^{(3)}$ and for the \mathcal{O}_W operator. The point is that the new physics effects due to $\mathcal{O}_{\varphi q}^{(3)}$ have very distinctive features which can be easily isolated even with a simple binned analysis with few variables. Indeed $\mathcal{O}_{\varphi q}^{(3)}$ (see eq. (2.9)) only contributes to the 00 polarization amplitude, which is non-vanishing in the SM as well and proportional to $\sin\Theta$. The squared 00 amplitude thus contributes to the cross section with a sizable interference term, which is peaked in the central scattering region $\cos\Theta \sim 0$. The other helicity channels play the role of background, and are peaked instead in the forward region. They are actually almost zero (at LO) at $\cos\Theta \simeq 0$. Therefore a binned analysis targeting central scattering (this is why we imposed the cut $\cos\Theta < 0.5$) is sufficient to isolate the effects of $\mathcal{O}_{\varphi q}^{(3)}$ at the interference level and thus to probe $G_{\varphi q}^{(3)}$ accurately. By including the decay variables as in the multivariate analysis we gain sensitivity to new terms in the cross section, namely to the interference between the 00 and the transverse amplitudes, however these new terms are comparable with those that are probed already in the Binned Analysis and thus they improve the reach only slightly.

The situation is very different for the \mathcal{O}_W operator. It contributes to the $++$ and $--$ helicity channels, that are highly suppressed in the SM and set exactly to zero in the Toy version of the problem we are studying here. The $p_{T,Z}$ (and Θ) distribution depends only at the quartic level on G_W , i.e. through the square of the BSM amplitude, because the interference between different helicity channels cancels out if we integrate the cross section in eq. (2.14) over the ZW azimuthal decay angles. Our Binned Analysis is sensitive to the interference term through the binning in φ_W , however this is not enough to approach the optimal reach because all the other decay variables (and Θ as well) do possess some discriminating power, from which we can benefit only through a multivariate analysis. More specifically, one can readily see by direct calculation that the dependence on all our kinematical variables of the G_W interference contribution to the differential cross section is different from the SM term. By integrating on any of this variables we partially lose sensitivity to this difference, and this is why the multivariate analysis performs much better than the binned one.

2.4.2 MADGRAPH Leading Order

The analyses performed for the Toy dataset can be easily replicated for the MADGRAPH LO Monte Carlo description of the process, obtaining the results shown in Figure 2.2.

The most noticeable difference with what was found with the Toy Monte Carlo is the strong degradation of the Matrix Element reach, and the fact that it gets weaker than the one of the Quadratic Classifier. As usual, the effect is more pronounced for the \mathcal{O}_W operator. This is not mathematically inconsistent because the analytic ratio $r(x, c)$ we employ for the Matrix Element test statistic is not equal anymore to the ratio of the true distributions according to which the data are generated. Therefore it is not supposed to give optimal performances. On the other hand the observed degradation is quantitatively surprising for G_W , especially in light

of the fact that the MADGRAPH LO Monte Carlo distributions seem quite similar to the ones of the Toy data at a superficial look. The degradation is not due to the high-energy approximation in the ZW production amplitude, indeed the results we are reporting are obtained with the exact tree-level helicity amplitudes, which are employed in eq. (2.14) in place of the ones in eq. (2.9). It is due to the other approximations we performed in the calculation of the cross section, namely to the assumption that the initial quark is always more energetic than the anti-quark, which allows us to interpret Θ as the angle between the quark and the Z, and to the one of a perfect reconstruction (up to the ambiguity) of the neutrino momentum. We verified that this is the case by repeating the Matrix Element analysis using the true neutrino momentum and the actual direction of motion of the quark in the Monte Carlo events. In this case the reach on G_W gets closer to the one obtained with the Toy data.

The degradation of the Matrix Element reach should be contrasted with the relative stability of the Quadratic Classifier method. Notice that the method is applied on the MADGRAPH LO data in the exact same way as on the Toy data, namely the architecture is the same, as well as the number of training point and the values of the Wilson coefficients in eq. (2.6) used for training. The computational complexity of the distribution ratio reconstruction is thus identical in the two cases, in spite of the fact that the MADGRAPH LO Monte Carlo offers a slight more complete (or “complex”) description of the data. The total computational cost is somewhat higher in the MADGRAPH LO case, but just because the process of Monte Carlo events generation is in itself more costly. Similar considerations hold at NLO, where the cost of event generation substantially increases.

2.5 The reach at Next-to-Leading Order

Including NLO QCD corrections is in general essential for an accurate modeling of the LHC data. Therefore it is imperative to check if and to what extent the findings of the previous section are confirmed with the MADGRAPH NLO Monte Carlo description of the process, introduced in Section 2.3.2. As far as the reconstruction of $\hat{r}(x, c)$ is concerned, using MADGRAPH NLO does not pose any conceptual or technical difficulty, provided of course the (positive and negative) Monte Carlo weights are properly included in the loss function as explained in Section 1.4.2. Computing the distribution of the test statistic variable that we obtain after the reconstruction (or of the one we employ with the Matrix Element method, for which the exact same issue is encountered) is instead slightly more complicated than with the Toy and MADGRAPH LO data. This point is discussed in the following section, while the illustration of the results is postponed to Section 2.5.2.

2.5.1 Estimating the test statistics distributions

As soon as $\tau_c(x)$ is known, either as an analytic function in the case of the Matrix Element or as a (trained) Neural Network in the case of the Quadratic Classifier, the test statistic $t_c(\mathcal{D})$, as defined in eq. (2.16), is fully specified. Namely we can concretely evaluate it on any dataset

$\mathcal{D} = \{x_i\}$, consisting of \mathcal{N} repeated instances of the variable x , for each given value of c . However in order to perform the hypothesis test, and eventually to estimate the reach $c_{2\sigma}$, we also need to estimate the probability distribution of $t_c(\mathcal{D})$ under the H_0 and under the H_1 hypotheses. This is the problematic step at NLO, after which the evaluation of $c_{2\sigma}$ proceeds in the exact same way as for the Toy and for the LO data. Specifically, once we are given with

$$\text{pdf}(t_c|H_0(c)) \text{ and } \text{pdf}(t_c|H_1), \quad (2.20)$$

we obtain the p -value as a function of t_c and c as in eq. (2.18) from the former, while from the latter we compute the median value of t_c and in turn

$$p_{\text{med}}(c) \equiv p(t_{\text{med}}(c); c), \quad (2.21)$$

as a function of c . After scanning over c and interpolating $p_{\text{med}}(c)$ in the vicinity of the reach (actually we interpolate the logarithm of $p_{\text{med}}(c)$, using three points in c and quadratic interpolation), we can solve the equation $p_{\text{med}}(c_{2\sigma}) = 0.05$ and obtain the reach as defined in eq. (2.17). Given the error on $p_{\text{med}}(c)$ at the three points used for the interpolation, the error on the estimate of $c_{2\sigma}$ is obtained by error propagation.

It is conceptually trivial (but numerically demanding) to estimate the distributions if artificial instances of the dataset \mathcal{D} (aka “toy” datasets) are available. In this case one can simply evaluate $t_c(\mathcal{D})$ on many toy datasets following the $H_0(c)$ and the H_1 hypotheses and estimate the distributions. More precisely, one just needs the empirical cumulative in $H_0(c)$ and the median of t_c in H_1 . Toy datasets are readily obtained from unweighted Monte Carlo samples by throwing \mathcal{N} random instances of x from the sample, with \mathcal{N} itself thrown Poissonianly around the total expected number of events. This is impossible at NLO because the events are necessarily weighted, therefore they are not a sampling of the underlying distribution of the variable x . As emphasized in Section 1.4.2, NLO Monte Carlo data can only be used to compute expectation values of observables $O(x)$ as in eq. (1.21). For instance we can compute the cross section in any region of the X space, and the mean or the higher order moments of the variable of interest, $\tau_c(x)$.

This suggests two options to estimate the distributions of the test statistic at NLO. The first one is to compute the distribution of $\tau_c(x)$ by means of a (weighted) histogram with many and very narrow bins. By knowing the cross section of each bin in τ_c , we know how many events are expected to fall in that bin and generate toy datasets for τ_c accordingly. This procedure is quite demanding, and it relies on a careful choice of the τ_c binning, which can only be performed on a case-by-case basis. It is still useful to validate the strategy we actually adopt, described below.

The second option is to approximate the distribution of t_c in a “nearly Gaussian” form, based on the Central Limit theorem. Namely we notice that t_c is in a trivial linear relation (see

2.5 The reach at Next-to-Leading Order

		Toy Data	LO	NLO
$G_{\varphi q}^{(3)}$	ME	$[-0.350(6), 0.356(8)]$	$[-0.399(13), 0.384(12)]$	$[-0.55(4), 0.464(14)]$
	SC	$\gtrsim 0.4$ ($p = 0.077(5)$)	—	—
	QC	$[-0.357(6), 0.365(8)]$	$[-0.401(12), 0.374(10)]$	$[-0.426(22), 0.401(21)]$
	BA	$[-0.48, 0.47]$	$[-0.50, 0.50]$	$[-0.58, 0.55]$
G_W	ME	$[-0.673(14), 0.697(11)]$	$[-1.390(21), 1.357(22)]$	$[-1.51(7), 1.93(14)]$
	SC	$\lesssim 1$ ($p = 0.038(3)$)	—	—
	QC	$[-0.781(13), 0.822(13)]$	$[-1.007(27), 0.987(26)]$	$[-0.99(4), 1.08(12)]$
	BA	$[-1.67, 1.67]$	$[-1.70, 1.85]$	$[-1.63, 1.98]$

Table 2.1: Bounds on the $G_{\varphi q}^{(3)}$ and G_W coefficients obtained for the Toy, LO and NLO datasets. The rows correspond to the Matrix Element (ME), Standard Classifier (SC), Quadratic Classifier (QC) and Binned Analysis (BA) approach. Notice that the errors on the Binned Analysis bounds are negligible. The results are given in 10^{-2} TeV^{-2} units.

eq. (2.16)) with the variable

$$\mathcal{T}_c(\mathcal{D}) \equiv \frac{1}{N} \sum_{i=1}^{\mathcal{N}} \tau_c(x_i), \quad (2.22)$$

where \mathcal{N} is Poisson-distributed with expected N , with $N = N(X|H)$ and $H = H_0$ or $H = H_1$. The x_i 's are independent and sampled according to $\text{pdf}(x|H)$. The cumulant-generating function of \mathcal{T}_c (which is a so-called ‘‘compound’’ Poisson variable [54]) is readily computed

$$K_{\mathcal{T}_c}(\xi) \equiv \log \left\{ \mathbb{E} \left[e^{\xi \mathcal{T}_c} \middle| H \right] \right\} = N \mathbb{E} \left[e^{\frac{\xi}{N} \tau_c} \middle| H \right] - N, \quad (2.23)$$

by first taking the expectation on the x_i 's conditional to \mathcal{N} and next averaging over the Poisson distribution of \mathcal{N} . Therefore the cumulants of \mathcal{T}_c ,

$$\kappa_{\mathcal{T}_c}^n \equiv \left. \frac{d^n K_{\mathcal{T}_c}(\xi)}{d\xi^n} \right|_{\xi=0} = N^{1-n} \mathbb{E} \left[\tau_c^n \middle| H \right], \quad (2.24)$$

are increasingly suppressed with N for larger and larger $n > 1$. Since N is of the order of several thousands in our case, neglecting all cumulants apart from the first and the second one, i.e. adopting a Gaussian distribution for \mathcal{T}_c , might be a good approximation.

Actually it turns out that in order to model properly the 5% tail of the distribution, which we need to probe for the exclusion limit, non-Gaussianity effects can be relevant. These are included by using a skew-normal distribution for \mathcal{T}_c , which contains one more adjustable parameter than the Gaussian to model the skewness. The mean, standard deviation and

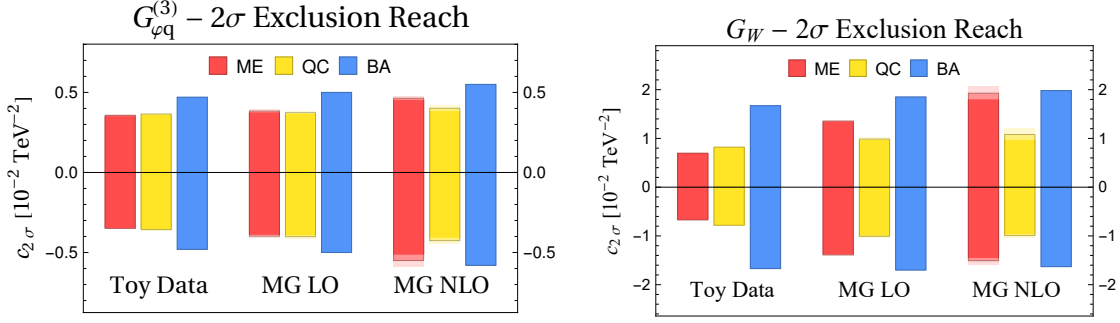


Figure 2.3: Expected exclusion reach on $G_{\varphi q}^{(3)}$ (left) and on G_W (right) with the various methodologies described in the text. The results are also reported in Table 2.1.

skewness of \mathcal{T}_c are immediately obtained from eq. (2.24)

$$\mu(\mathcal{T}_c) = \langle \tau_c \rangle, \quad \sigma(\mathcal{T}_c) = \frac{1}{\sqrt{N}} \sqrt{\langle \tau_c^2 \rangle}, \quad \mu_3(\mathcal{T}_c) = \frac{1}{\sqrt{N}} \frac{\langle \tau_c^3 \rangle}{\langle \tau_c^2 \rangle^{3/2}}, \quad (2.25)$$

where $\langle \cdot \rangle$ is used to denote expectation for brevity. By computing the expectation values of τ_c , τ_c^2 and τ_c^3 using the Monte Carlo data, we thus find the parameters of the skew-normal distribution for \mathcal{T}_c and in turn the distribution of t_c . We finally obtain the median p-value from the definition in eq. (2.21). The errors on the expectation values of τ_c are estimated from the fluctuations in the means on subsets of the entire Monte Carlo sample. These errors are propagated to the p-value and eventually to the $c_{2\sigma}$ estimated reach as previously explained. Accurate results (see Table 2.1) are obtained with relatively small Monte Carlo samples. Namely, 500k event were used at NLO, 1M at LO and 3M for the Toy data.

We cross-checked the above procedure in multiple ways. First, it reproduces within errors the LO and Toy p-values obtained with the toy experiments. Second, we validated it against the approach based on τ_c binning on NLO data, as previously mentioned. We also verified that including the skewness changes the results only slightly, with respect to those obtained in the Gaussian limit. Further improving the modeling of the non-Gaussianity with more complex distributions than the skew-normal, with more adjustable parameters in order to fit higher order moments of \mathcal{T}_c , is therefore not expected to affect the results.

2.5.2 Results

Our results with the MADGRAPH NLO Monte Carlo are reported in Figure 2.3 and in Table 2.1. They essentially confirm the trend we already observed in the transition from the Toy to the MADGRAPH LO data. The Matrix Element keeps losing sensitivity because the analytic distribution ratio is now even more faraway from the actual distribution ratio since it does not include NLO QCD effects. The reach of the Binned Analysis deteriorates less, so that it becomes comparable to the one of the Matrix Element. The Quadratic Classifier reach is

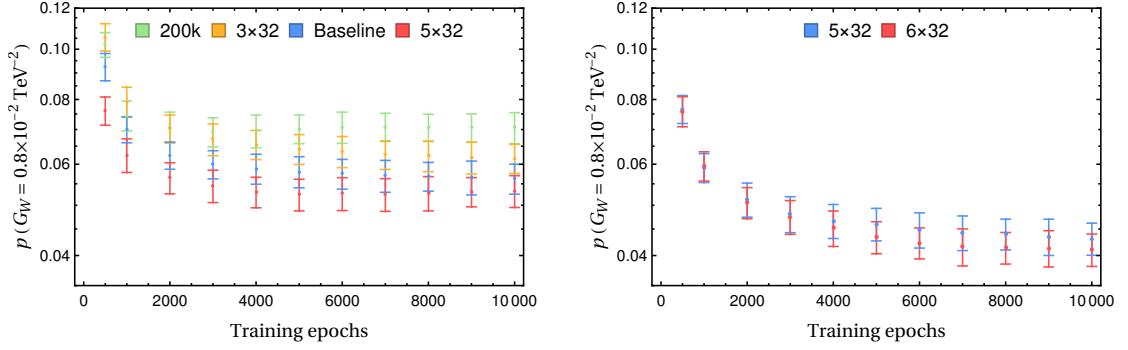


Figure 2.4: Evolution of the p -value for different architectures and training sample sizes. On the left plot we compare the baseline setup with the baseline architecture Network trained with 200k points per value of c (for a total of 2.4M points), and with the baseline number of training points (500k, times 12) on architectures with one less (“ 3×32 ”) and one more (“ 5×32 ”) hidden layer. On the right plot, a similar analysis is performed, but with 3M points per value of c .

remarkably stable. Actually it slightly improves with respect to the LO one for G_W . This is probably due to the appearance of same-helicity SM transverse amplitudes (see Section 2.3.2) and of the corresponding interference term for the \mathcal{O}_W operators.

Notice few minor differences in the implementation of the Quadratic Classifier and of the Binned Analysis at NLO. The Quadratic Classifier now also employs the variable $p_{T,ZW}$, as discussed in Section 2.2. The Binned Analysis for $G_{\varphi q}^{(3)}$ employs $p_{T,ZW}$ as well, through a cut $p_{T,ZW}/p_{T,V} < 0.5$. This improves the reach [9] because it helps recovering (partially) the background suppression due to the zero of the transverse amplitudes in the central region.

2.6 Neural Network implementation and validation

The strategies described in Section 1.4.2 were implemented in Pytorch [55] and run on NVIDIA GeForce GTX 1070 graphics card. Fully connected feedforward deep Neural Networks were employed, acting on the features vector

$$x = \{s, \Theta, \theta_W, \theta_Z, p_{T,ZW}, p_{T,Z}, \sin \varphi_W, \cos \varphi_W, \sin \varphi_Z, \cos \varphi_Z\}, \quad (2.26)$$

for a total of 10 features. Each feature is standardized with a linear transformation to have zero mean and unit variance on the training sample. For the Quadratic Classifier training, the Wilson coefficient employed in the parametrization (2.3) were scaled to have unit variance on the training sample. Employing the redundant variables (i.e., $p_{T,Z}$, and the cosines and sines of $\varphi_{W,Z}$) is helpful for the performances, especially the angular ones, which enforce the periodicity of the azimuthal angular variables. The “baseline” results presented in Figures 2.2, 2.3 and in Table 2.1 were all obtained with the features vector above and employing a total of

6 million training Monte Carlo points for each of the two Wilson coefficients. Training was always performed with a single batch (which was found to perform better in all cases), even if in practice the gradients calculation was split in mini-batches of 100k points in order to avoid saturating the memory of the GPU. Apart from these common aspects, the optimization of the Neural Network design and of the training strategy is rather different for the Quadratic and for the Standard Classifier methods. They are thus discussed separately in what follows.

2.6.1 The Quadratic Classifier

For the Quadratic Classifier, best performances were obtained with ReLU activation functions and with the Adam Pytorch optimizer. The initial learning rate (set to 10^{-3}) does not strongly affect the performances. Other attempts, with Sigmoid activation and/or with SGD optimizer, produced longer execution time and worse performances. The baseline architecture for the two Neural Networks n_α and n_β in eq. (2.3) consists of 4 hidden layers with 32 neurons, namely the architecture $\{10, 32, 32, 32, 32, 1\}$, including the input and the output layers. Weight Clipping was implemented as a bound on the L_1 norm of the weights in each layer, but found not to play a significant role. The total training time, for 10^4 training epochs, is around 5 hours for the baseline architecture and with the baseline number (6 million) of training points.

The Neural Network architecture was selected based on plots like those in Figure 2.4. The left panel shows the evolution with the number of training epochs of the median p -value (see eq. (2.21)) on Toy data for $c = G_W = 0.8 \times 10^{-2} \text{ TeV}^{-2}$, with the baseline and with larger and smaller Networks. We see that adding or removing one hidden layer to the baseline architecture does not change the performances significantly. The plot also shows that 10^4 epochs are sufficient for the convergence and that no overfitting occurs. The degradation of the performances with less training point is also illustrated in the plot. Of course, the p -value is evaluated using independent Monte Carlo samples, not employed for training. The errors on the p -value are estimated from the error on the skew-normal distribution parameters as explained in Section 2.5.1. In the baseline configuration we used 500k EFT Monte Carlo training points for each of the 6 values of G_W in eq. (2.6), plus 500k for each associated SM sample. Each sample consists instead of 3M points in the extended configuration employed on the right panel of Figure 2.4, for a total of 36M. The same value of $G_W = 0.8 \times 10^{-2} \text{ TeV}^{-2}$ is employed. The baseline architecture becomes insufficient, and best results are obtained with the 6 hidden layers of 32 neurons each.

The figure also demonstrates that the method is systematically improvable towards optimality. The value of G_W considered in the figure was not within the 95% CL reach with the baseline setup, while it becomes visible with the extended configuration. All the reaches reported in Table 2.1 would expectedly improve with the extended configuration. The G_W reach on Toy data becomes $[-0.732(9), 0.764(14)] 10^{-2} \text{ TeV}^{-2}$, which is now only less than 10% worse than the optimal Matrix Element reach. Training takes around 30 hours with the extended configuration, while generating and processing the required training points with MADGRAPH NLO

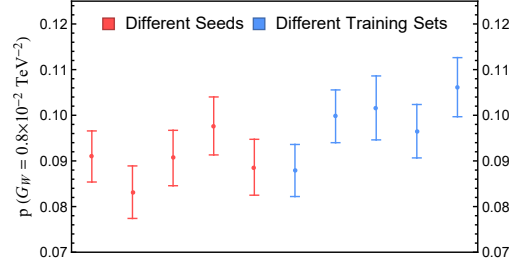


Figure 2.5: Results of 5 different trainings of the same architecture (Baseline architecture trained with 2.4M points) using: the same training data but different initialization seeds (red points) and the same initialization but different training data samples (blue points).

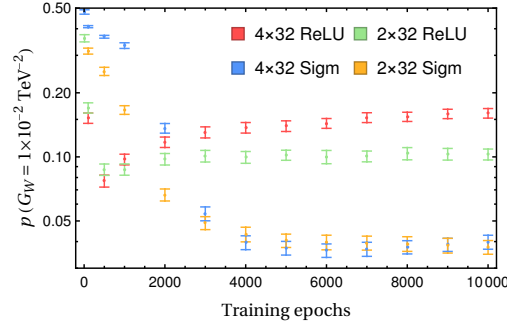


Figure 2.6: The p -value evolution during training for the Standard Classifier using different architectures and activation functions. The value $G_W = 1 \times 10^{-2} \text{ TeV}^{-2}$ is employed.

(which is the most demanding generator) would take around 10 days on a 32-cores workstation. We could thus try to improve also the NLO reach even with limited computing resources.

For the reproducibility of our results we also study how the performances depend on the Neural Network initialization and on the statistical fluctuations of the Monte Carlo training sample. This analysis is performed in a reduced setup, with a total of 2.4 million training point, and for $G_W = 0.8 \times 10^{-2} \text{ TeV}^{-2}$. We see in Figure 2.5 that the p -value fluctuates by varying the random seed used for training at a level comparable with the error on its determination. Similar results are observed by employing different independent Monte Carlo training samples. Notice that these fluctuations should not be interpreted as additional contributions to the error on the p -value. Each individual Neural Network obtained from each individual training defines a valid test statistic variable, on which we are allowed to base our statistical analysis. Since the fluctuations are comparable to the p -value estimate errors, our sensitivity projections were obtained by randomly selecting one of the seed/training set configuration.

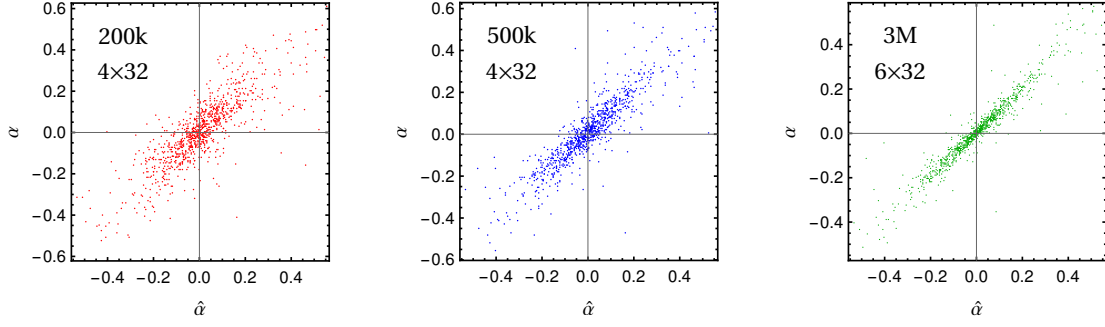


Figure 2.7: Comparison between the reconstructed ($\hat{\alpha}$) and true (α) linear term of the distribution ratio for the G_W operator.

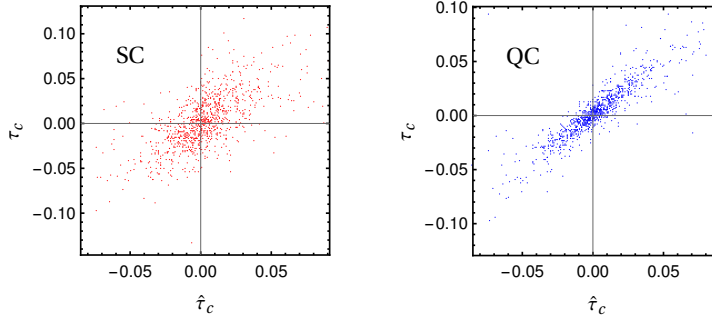


Figure 2.8: Comparison between the reconstructed ($\hat{\tau}_c$) and true (τ_c) distribution log-ratio for $G_W = 1 \times 10^{-2} \text{ TeV}^{-2}$. The Standard Classifier and the Quadratic one are considered in the left and right panel of the figure, respectively.

2.6.2 The Standard Classifier

Hyperparameters optimization is rather different for the Standard Classifier. We see in Figure 2.6 that Networks with ReLU activation like those we employed for the Quadratic Classifier displays overfitting, and Sigmoid activations need to be employed. The results in Figures 2.2 and in Table 2.1 were obtained with 2 hidden layers with 32 neurons each and Sigmoid activation. The figure shows that increasing the complexity does not improve the performances.

This different behavior of the Standard Classifier compared with the Quadratic one is probably due to the fact that training is performed on small Wilson coefficient EFT data, whose underlying distribution is very similar to the one of the SM data sample. Therefore there is not much genuine difference between the two training sets, and the Network is sensitive to statistical fluctuations in the training samples. The Quadratic Classifier instead is trained with large values of the Wilson coefficients. The optimizer thus drives the Neural Networks towards the deep minimum that corresponds to a proper modeling of the distribution ratio, which is more stable against statistical fluctuations of the training samples.

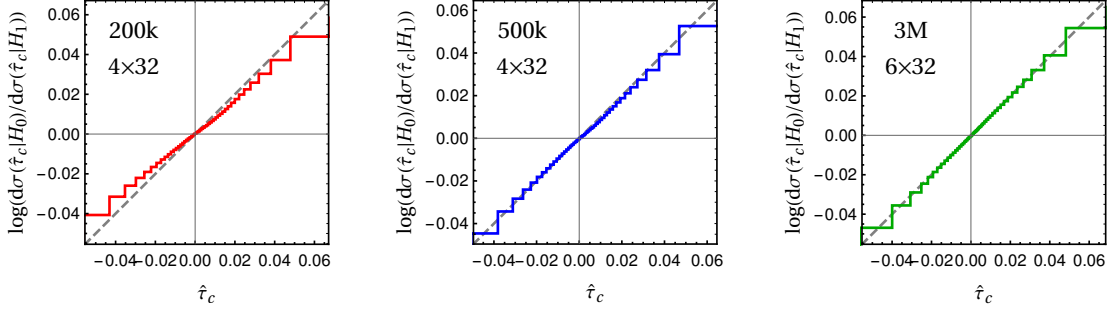


Figure 2.9: Distribution log-ratio for $\hat{\tau}_c$, for $c = G_W = 0.8 \times 10^{-2} \text{ TeV}^{-2}$. The accurate determination displayed in the plots is obtained by the reweighting of a single Toy SM Monte Carlo sample. The same approach, based on reweighting, could have been adopted to assess the quality of the distribution ratio reconstruction on MADGRAPH Monte Carlo data, using MADWEIGHT.

2.6.3 Validation

An important question is how to validate as “satisfactory” the outcome of the hyperparameters optimization described above. This is straightforward for the Toy version of the problem, because we have to our disposal a rigorous notion of statistical optimality, through the Neyman–Pearson lemma, and we do have direct access to the true distribution ratio through which the data are generated. Therefore we know that we can stop optimization as soon as the reach of the Neural Network becomes sufficiently close to the one of the Matrix Element method. We can also rely on a more naive validation test, based on comparing point-by-point the distribution ratio learned by the Neural Network with the true one, which is known analytically. For instance in Figure 2.7 we compare the true linear term $\alpha(x)$ in eq. (2.1) (for the \mathcal{O}_W operator) with its estimator $\hat{\alpha}(x) \equiv \hat{n}_\alpha(x)$ provided by the trained Neural Network. The baseline architecture is employed, with increasing number of training points. While it is impossible to extract quantitative information, a qualitative comparison between the three scatter plots confirms that more training points improve the quality of the reconstruction. We also show, in Figure 2.8, the correlation between the true and the reconstructed ratios (for $G_W = 1 \times 10^{-2} \text{ TeV}^{-2}$, which corresponds to the Standard Classifier 95% reach) obtained with the Quadratic and with the Standard Classifier. The reconstruction obtained with the Quadratic Classifier is more accurate as expected.

Validation is of course less easy if, as it is always the case on real problems, the true distribution ratio is not known. One option is to proceed like we did in the present chapter. Namely to identify a Toy version of the problem that is sufficiently close to the real one and for which the distribution ratio is known. Since it is unlikely that the true distribution is much harder to learn than the Toy distribution, and since we can establish optimality on the Toy data using a certain architecture and training dataset size, we can argue heuristically that the same configuration will be optimal also with a more refined Monte Carlo description.

Finally, one can monitor heuristically how accurately the distribution ratio is reconstructed,

as follows. The true distribution log-ratio $\tau_c(x) = \log r(x, c)$, seen as a statistical variable for each fixed value of c , obeys, by definition, the equation

$$\frac{d\sigma_0}{d\tau_c} = e^{\tau_c} \frac{d\sigma_1}{d\tau_c}. \quad (2.27)$$

Therefore if we computed the distribution of τ_c (if it was known) in the EFT hypothesis $H_0(c)$ and in the SM hypothesis H_1 , and take the log-ratio, the result would be a straight line as a function of τ_c . By computing the same distributions for the reconstructed distribution log-ratio $\hat{\tau}_c = \log \hat{r}(x, c)$, we can thus get an indication of how closely $\hat{r}(x, c)$ approximates $r(x, c)$. While no quantitative information can be extracted from these plots, they clearly illustrate the improvement achieved by enlarging the size of the training sample and the Neural Network architecture, as Figure 2.9 shows.

2.7 Event Reweighting

Training neural network based classifiers requires a large amount of simulation data, as we have seen in sec.(2.6.1). In fact, to study statistical distributions for BSM hypotheses, in general, is data consuming. In the previous section, to retrieve information from events of various Wilson coefficients, we generated data sets under different values of c independently. This is highly inefficient. To relieve the burden of rerunning simulations, we consider the *reweighting* technique [56][57][58][59], which exploits the usage of latent variables in Monte Carlo generators, to generate data sets for multiple Wilson coefficients at once.

Monte Carlo generators designed for the LHC consist of two parts. At the first level, events are registered with parton-level latent variables \mathbf{z}_{pl} , which includes the four-momenta as well as helicity, and charge information, as discrete labels, of the initial and final states. Scattering amplitudes can be computed theoretically at this level. However, the parton level latent variables are generally impossible to measure in experiments. To start with, the detectors do not measure helicities. As for the initial states, because it is the quarks that participate in the scattering process, only a fraction x of the energy carried by the protons will be available during the collision. This process is carefully modelled by the proton distribution functions $f(x)$. The value of x , or more precisely, for the two initial states, x_1 and x_2 , can be determined through momentum conservation, if the momenta of the final states are known. However, the final states usually include particles that cannot be directly observed or invisible, concretely, hadrons or neutrinos. In short, the parton level latent variables \mathbf{z}_{pl} are not accessible in the form of observables.

In addition, at the second level of simulation, the generator throws random variables that model effects from hadronic showering or detector interactions to make the events closer to reality. This statistical smearing process maps \mathbf{z}_{pl} to experimentally visible measurements \mathbf{x}_{vis} .

$$\mathbf{z}_{\text{pl}} \xrightarrow{\text{random variables}} \mathbf{x}_{\text{vis}} \quad (2.28)$$

This concept was also mentioned in sec.(1.3.2), where the Matrix Element method is developed to employ a set of hand-picked transfer functions to approximate this process in order to express the likelihood in terms of observables \mathbf{x}_{vis} . One has to keep in mind that in our assumption, this mapping is done independently from the first level of simulation, or in other words, BSM effects do not affect hadron showering or detector interaction, at least to a measurable extend.

Simulated events, depending on the specific Monte Carlo algorithm, can be weighted or unweighted, as long as the weights are positive definite. This is the case for the Toy generator and leading order (LO) generator. For events simulated at the next-to-leading order (NLO) level, however, the weights are not positive-definite. 1-loop correction or emission is allowed, and in order to model photon emission and deal with infrared divergence of the energy of photons, *subtraction method* is used. This involves subtracting two independent integrals, which can result in negative weights [58][60]. This blocks unweighting of the events. The negative do not represent a "negative" probability, but only serve in the sense that the expectation value of the observables using a weighted data set is accurate.

Now, consider a BSM hypothesis parametrised by c . The correction to the SM amplitude $\mathcal{M}(\mathbf{z}_{\text{pl},i}|0)$ of event i with weight $w_{0,i}$ can be computed theoretically. We assign

$$w_{c,i} = \frac{|\mathcal{M}(\mathbf{z}_{\text{pl},i}|c)|^2}{|\mathcal{M}(\mathbf{z}_{\text{pl},i}|0)|^2} w_{0,i} \quad (2.29)$$

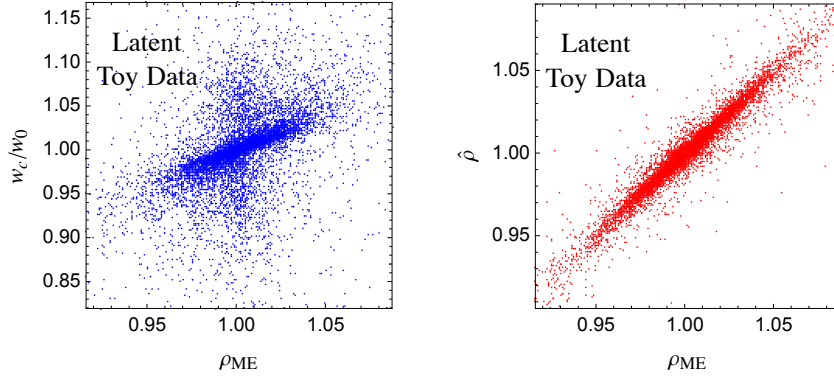
as its weight under the BSM hypothesis. Because the BSM effects and the hadronisation or detector response functions factorise, this weight remains the same after going through the second level of simulation

$$\{\mathbf{z}_{\text{pl}}, w_c\} \xrightarrow{\text{random variables}} \{\mathbf{x}_{\text{vis}}, w_c\}. \quad (2.30)$$

Through this technique, we have exploit the algorithm of Monte Carlo generators such that we do not have to run the second level of simulation multiple times in order to carry out statistical analysis, such that having a data set with multiple reweights is equivalent to having data sets corresponding to all the BSM parameters under which the reweights are computed. Event-by-event reweighting is a common technique employed in studying collision events at the LHC [59] and is particularly useful for our study of BSM parameters or in the future, nuisance parameters.

Neural networks can take advantage of this technique naturally. To include reweights in the algorithm, the formula for computing the loss eq.(1.22), needs to be extended with another summation over the set of reweight parameters $\mathcal{C} = \{c_1, c_2, \dots, c_n\}$,

$$L[f(\cdot)] = \sum_{c \in \mathcal{C}} \left[\sum_{e \in S_0} w_{0,e} [f(x_e)]^2 + \sum_{e \in S_1} w_{c,e} [1 - f(x_e)]^2 \right]. \quad (2.31)$$


 Figure 2.10: Correlation between ρ_{rw} and ρ_{ME} , and $\hat{\rho}$ and ρ_{rmME} .

The network can also be parametrized as was explained in eq.(2.4)

$$L[\mathbf{n}_\alpha(\cdot), \mathbf{n}_\beta(\cdot)] = \sum_{c \in \mathcal{C}} \left[\sum_{e \in S_0(\mu_i)} w_{0,e} [f(x_e, c_i)]^2 + \sum_{e \in S_1(\mu_i)} w_{c,e} [1 - f(x_e, c_i)]^2 \right]. \quad (2.32)$$

In this manner, we multiply the efficiency of the algorithm, as the neural network only has to be trained once to approximate $f(x_e, c_i)$, apart from saving a considerable amount of time by not having to rerun the simulation for multiple data sets.

2.7.1 The Parametrized Reweighted Classifier

The parametrized classifier introduced in sec.(2.1) is now upgraded to work with reweighted events. The data set is generated using the following set of Wilson coefficients,

$$G_W : \quad \{\pm 5, \pm 2.5, \pm 1.25\} \times 10^{-2} \text{ TeV}^{-2}. \quad (2.33)$$

Different from the three generators described in sec.(2.3.2), we employ only the latent Toy generator. In this simulation, two variables are latent, namely the helicities of the leptons of the Z decay, and the azimuthal angle ϕ_ν of the neutrino; the angle is randomly selected to be the true ϕ_ν or $\pi/2 - \phi_\nu$. This modification is realistic because the detectors do not measure helicity, and neutrinos are invisible to the detectors; the angle ϕ_ν cannot be reconstructed kinematically. The nature of the learning problem now goes beyond regression even for the Toy data, and this is what we would like to investigate using our classifiers.

To clarify what we mean by having latent variables in the latent Toy generator, we show the comparison between the ratio of differential cross-sections given by the reweight, w_c/w_0 , and that computed by ME, ρ_{ME} , in the left panel of fig.(2.10). The reweights have information to the underlying latent variables, while ME does not.

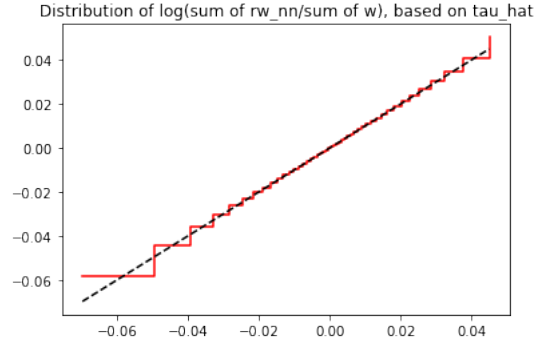
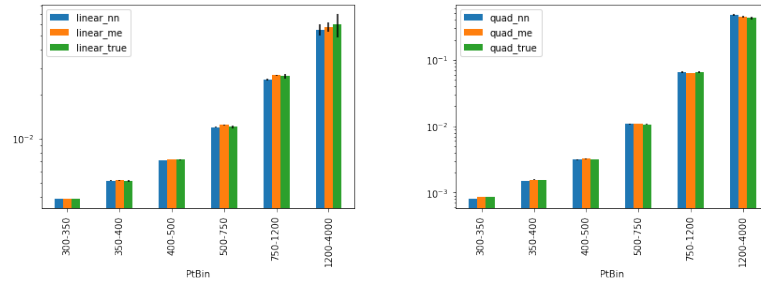

 Figure 2.11: Distribution log-ratio for $\hat{\tau}_{rw}$.


Figure 2.12: Reconstructed linear and quadratic term as compared to that by ME and reweights

The network is now designed to be smaller, with only two hidden layers of 32 neurons each, and uses Sigmoid functions instead of RELU as activation functions as they help avoid the “dead neuron” effect. We used an initial learning rate of 3×10^{-3} for the first 40k epochs and then 1×10^{-3} for another 20k epochs. Optimisation was done using ADAM. Notably, the size of the data set is reduced to 3 million points, and each point comes with reweights of the mentioned Wilson coefficients (eq.2.33). Training takes roughly 2 hours for one such data set on a A30 GPU. Reweighting clearly improves the training efficiency.

The reach of the reweighted neural network is $0.7 \times 10^{-2} \text{ TeV}^{-2}$, which coincides with that of the ME. To assess the reconstruction accuracy of our classifier, we display the comparison between the reconstructed ratio $\hat{\rho}$ and ρ_{ME} in the right panel fig.(2.10).

As a validation check, we produce the “staircase” plot in the same style as fig.(2.9) for $\hat{\tau}_{rw}$, which is the reweighted version of $\hat{\tau}_c$, in fig.(2.11).

To further examine the reconstruction accuracy of different components of ρ , we check in fig.(2.12) the reconstructed linear and quadratic terms as defined in the quadratic form presented in eq.(2.1). We select a Wilson coefficient of $G_W = 1 \times 10^{-2} \text{ TeV}^{-2}$ close to the reach and take a binning in both the transverse momentum $p_{T,Z}$ and a cut on the azimuthal angles $\cos \phi_{W,Z} < -0.5$ or $\cos \phi_{W,Z} > 0.5$. Physically, these binning and cut should outline new physics signals.

Chapter 2. Parametrized Classifiers for Optimal EFT Sensitivity

The plots show that the reweighted neural network reconstructs the linear and the quadratic term with satisfying accuracy, and qualify as a promising technique to be further developed and applied.

3 Electroweak Radiation at Future Muon Colliders

3.1 Introduction

The perspective of a future muon collider with high energy and high luminosity [61], whose feasibility is being investigated by the International Muon Collider Collaboration [62], has triggered a growing interest in the physics potential of lepton colliders with a center of mass energy of 10 TeV or more [63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102]. Such a Very High Energy Lepton Collider (VHEL) could greatly advance the post-LHC knowledge of fundamental physics [61] by directly searching for new heavy particles (see e.g. [63, 64, 65, 66]), and by precise measurements of ElectroWeak (EW) scale processes exploiting the high luminosity of virtual vector bosons pairs (see e.g. [66, 67, 68]). By these two search modes, the VHEL reach on new physics is generally comparable to that of the most ambitious future collider projects, in the corresponding domains of investigation. In particular it is comparable to the combined reach of the -ee and -hh stages of the FCC program. The sensitivity is slightly weaker or slightly stronger depending on the specific new physics target and, obviously, on the assumed VHEL energy and luminosity.

At a VHEL, however, there also exists a third strategy of investigation [61, 67], based on hard processes with energy scale comparable to the collider energy $E_{\text{cm}} \sim 10$ TeV. As the indirect effects of new heavy particles are enhanced at high energy, these processes are extremely sensitive probes of new physics. With the target integrated luminosity of 10 ab^{-1} , $2 \rightarrow 2$ scattering processes at $E_{\text{cm}} = 10$ TeV can be generically measured with percent or few-percent precision. Such measurements are therefore sensitive to putative new physics at a scale $\Lambda \sim 100$ TeV when its effects, relative to the SM cross-section, scale like $(E_{\text{cm}}/\Lambda)^2$. In an Effective Field Theory (EFT) description of new physics, this corresponds to an enhanced sensitivity to those dimension-6 operators that contribute to the $2 \rightarrow 2$ processes with a quadratically growing-with-energy term. The VHEL sensitivity to new physics through this kind of “high-energy” probes vastly and generically exceeds the potential of any other future project that is currently under consideration [67]. In particular it exceeds the sensitivity of

precision measurements of EW-scale (~ 100 GeV) processes at future Higgs factories, where new physics at $\Lambda \sim 100$ TeV produces effects at the unobservable level of one part per million. It also exceeds the potential sensitivity of a 100 TeV proton collider, because the effective luminosity for partonic collisions above 10 TeV is significantly lower than that of the VHEL. The possibility of probing new physics at the 100 TeV scale, and in particular of probing new physics that is either flavor-universal or endowed with a flavor protection mechanism, is thus a unique feature of the VHEL that deserves an extensive investigation.¹

The above mentioned high-energy strategy exploits simple $2 \rightarrow 2$ processes with extremely low or negligible background, whose target accuracy is statistically limited to 1%. At a superficial look, it might thus seem that its implementation will not pose any specific challenge, neither as concerns the feasibility of the measurements, nor as concerns the theoretical predictions that are needed for their interpretation. However the situation is slightly more complex, both experimentally and theoretically, owing to the phenomenon of EW radiation [103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123], which becomes prominent at 10 TeV or above. This happens because of the existence of a large separation between the hard scale E of the process and the vector boson mass scale m_W , which acts as an IR cutoff. As the hard scale is increased, large logarithms $\log E^2/m_W^2$ (squared) enhance both virtual corrections and real emission, up to the point where fixed-order perturbation theory becomes insufficient and resummation is needed. The experimental implications of the copious emission of real EW massive vector bosons should be assessed. In particular it should be studied if and how it impacts the accuracy of the relevant cross-section measurements. In this chapter we investigate the implications of EW radiation on the theoretical predictions and, assuming purely statistical experimental errors, on the VHEL sensitivity to new physics.

EW radiation obviously displays some similarities with QCD radiation, but also remarkable differences. First, EW scattering processes violate the KLN theorem assumptions [124, 125] because the initial state particles are not EW singlets. Therefore no cancellation takes place between virtual and real contributions, not even in “fully-inclusive” cross-sections [103, 104]. Moreover the observables that are fully inclusive in the sense of Ref. [103] are insufficient to characterize new physics because they require summing over the “color” of the hard final-state particles. In the EW context, color sum means, for instance, including transversely-polarized W and Z bosons and photons (or, longitudinal W , Z and Higgs) in the same observable, while we need to keep them separate for a comprehensive exploration of new physics. Unlike QCD (and QED), EW radiation effects cannot be eliminated or systematically mitigated with a judicious choice of the observables. They unavoidably play an important role in the predictions.

The second peculiarity of EW radiation is that the IR cutoff scale is physical, and the theory is weakly-coupled at the IR scale. It should thus be possible to characterize the radiation

¹Hard processes are also useful to investigate flavor non-universal effects, as we will see in Section 3.3.3 for Top Compositeness. See also Ref. [97] for a study of new physics potentially responsible for the $g - 2$ muon anomaly.

fully by first-principle suitably resummed perturbative calculations. Unlike QCD and QED, one can consider observables with an arbitrary degree of radiation exclusiveness, among which “exclusive” scattering cross-sections with a fixed number (2, in $2 \rightarrow 2$ processes) of hard partons in the final state and no extra massive vector bosons.² Fully-inclusive cross-sections can be also considered, however they are not sufficiently informative on new physics as previously mentioned. In this chapter we employ “semi-inclusive” final states, consisting of 2 hard partons with fixed EW color and flavor carrying a large fraction of the total available energy E_{cm} and accompanied by an arbitrary number of massive vectors, photons and other light particles³. Our resummed predictions for semi-inclusive observables at the double logarithm (DL) accuracy are obtained by extending the IR Evolution Equation (IREE) treatment of EW radiation developed in Ref. [105]. Similarly, we employ the IREE to compute the more standard exclusive cross-sections. Single-logarithmic terms turn out to be relevant, and they are included at fixed one-loop order in the exclusive cross-sections using the results of Ref. [127].

Finally, there is an interplay between EW radiation and short-distance physics that has no analog in QED and QCD [67]. Based on a simplistic fixed order intuition, this interplay can be exemplified by noticing that the emission of a soft W from one initial lepton changes the total charge of the initial state of the hard $2 \rightarrow 2$ scattering process. By allowing for the charged W emission one thus obtains a scattering cross-section that is sensitive to charged current new physics interactions, while the exclusive process where no radiation is allowed is only sensitive to neutral currents.⁴ The combined measurement of the two types of cross-section thus enables a more effective and complete exploration of new physics. In reality the situation is slightly more complex, because the neutral and the charged current hard amplitudes both contribute to the resummed expression of the neutral exclusive and of the charged and neutral semi-inclusive cross-sections. However, since they appear in different combinations in the different cross-sections, the conclusion is unchanged.

At a more quantitative level, we will see that the VHEL energy sits in an interestingly “intermediate” regime for EW radiation. The energy is on one hand large enough for the radiation effects to be important and require resummation. Accurate resummation techniques, more accurate than the one considered in the present chapter, will thus be needed to fully exploit the potential of the machine. On the other hand, the energy is not yet in the asymptotic regime where the Sudakov exponentials entail a strong suppression of the non-emission probability. Therefore in this regime the exclusive cross-sections are still large, and comparable with the

²In order to cope with QED and QCD radiation, the observable must still be inclusive over the emission of photons and other light particles. The cross-section we define as “exclusive” coincides with the “semi-inclusive” cross-section of Ref. [105]. Correspondingly, the “semi-inclusive” cross-section we will readily introduce was not considered in Ref. [105]. See Section 3.2 for details.

³A similar observable is discussed in [126] to show the impact of weak gauge boson emission at LHC. The final state they consider is somehow intermediate between the “fully-inclusive” of [103] and the “semi-inclusive” we study in this chapter

⁴More precisely, the charged and neutral current process depend on different linear combinations of the Wilson coefficients of the operators parametrizing new physics.

semi-inclusive (and fully-inclusive) ones. Observables with a different degree of radiation inclusiveness can thus be measured with comparable statistical accuracy and combined, potentially boosting, as previously explained, the indirect sensitivity to heavy new physics.

3.2 Theoretical predictions

Several approaches have been considered in the literature for the resummation of the effects of EW radiation. Double logarithm (DL) contributions, of the form $(\alpha \log^2 E^2 / m_W^2)^n$ with arbitrary n , have been resummed in fully-inclusive and exclusive cross-sections, using respectively Asymptotic Dynamics [103, 104] and IREE [105, 106]. In Soft-Collinear Effective Theory (SCET) the expansion is already organized in exponential form. In that case the resummation of leading logarithms (LL)⁵ as well as its extension to next-to-leading (NLL) logarithm [112] has been studied. The study of EW parton distribution or fragmentation functions [115, 116, 117, 118, 119] is obviously connected, but not directly relevant for very high energy processes, where the main role is played by the emission of EW radiation that is both collinear and soft. Notice however that in some framework [115, 116, 117, 118, 119] soft effects are partially or completely included in parton distributions and fragmentation functions.

In this chapter we employ DL predictions based on the robust diagrammatic methodology of the IREE [105], which we review and further develop in Section 3.2.1. We also supplement our computations by the available virtual single logarithms (SL) at 1-loop [127, 128]. Based on these results, we present in Sections 3.2.2 and 3.2.3 our theoretical predictions for the di-fermion and di-boson production processes at the VHEL. While it will emerge that single logarithms are potentially relevant, a systematic consideration of these effects goes beyond the scope of the thesis. Our finding that electroweak radiation can be used to our own advantage in the exploration of new physics, strongly motivates the systematic inclusion of subleading effects. A first simple step would be to include in our predictions the single logarithms from real emissions at tree level. A complete treatment including resummation would likely best be achieved by using SCET.

3.2.1 IR Evolution Equations

The basic idea of the IREE is to introduce an unphysical IR regulator λ with dimension (energy)² in the calculation of the observables and to write down a differential equation for the evolution with λ of the result. Denoting by “ E^2 ” the hardness of the observable under consideration, the choice $\lambda \sim E^2$ eliminates all the IR enhancements and makes fixed-order perturbation theory well-behaved. For $\lambda \sim E^2$, the Born level computation therefore offers a reliable approximation, which can be used as the initial condition for the evolution equation to lower λ . The physical predictions are obtained from the solution of the IREE in the limit $\lambda \rightarrow 0$.

⁵These include but do not coincide with the pure DL, as explained, for instance, in Ref. [112].

In order to define the IR regulator, consider the 4-momenta k_i of the external hard particles that participate in the scattering process. They will correspond in our setup to the 4 legs of a central energetic $2 \rightarrow 2$ processes. With the exception of the masses $k_i^2 \ll E^2$, all the Lorentz invariants constructed with the k_i 's are therefore large and of order E^2 . Given a pair ij of external hard particles and given a radiation particle with 4-momentum q we define its hardness relative to the ij pair as

$$\mathfrak{h}^{ij}(q) \equiv 2 \left| \frac{(k_i \cdot q)(k_j \cdot q)}{(k_i \cdot k_j)} \right|. \quad (3.1)$$

The IR regulator is provided by the lower bound λ

$$\mathfrak{h}(q) \equiv \min_{i \neq j} \mathfrak{h}^{ij}(q) > \lambda. \quad (3.2)$$

on the “absolute” hardness \mathfrak{h} of the radiation. Notice that the bound is imposed on the 4-momentum of each individual radiation particle, either virtual or real. Specifically, eq. (3.2) applies to the off-shell loop momenta describing virtual radiation, as well as to the on-shell momenta of real radiation particles in the final state of the process. The specific definition of the radiation hardness in eq. (3.1) is convenient because it reflects the structure of the denominators that appear in the calculation of the IR-enhanced diagrams, as we will readily see. At this stage, it is enough to remark that the lower cut on $\mathfrak{h}(q)$ in eq. (3.2) is a valid IR regulator as it eliminates both the regions where q is soft and those where it is collinear to one of the hard partons.

The main peculiarity of the IREE formalism applied to EW radiation stems from the presence of the physical scale $m_W \sim 100$ GeV associated to the masses of the EW bosons. We will see that m_W acts as a threshold that separates two different regimes, $\lambda \gg m_W^2$ and $\lambda \ll m_W^2$. In the former regime, the cut on the radiation hardness in eq. (3.2) is so strong that the mass of the radiation particles can be safely neglected and the IREE computed in the unbroken $SU(2)_L \times U(1)_Y$ EW gauge theory. Starting from the initial condition at $\lambda \sim E^2$, the evolution is thus performed with the $SU(2)_L \times U(1)_Y$ evolution kernel down to $\lambda \sim m_W^2$. At $\lambda \ll m_W^2$, the massive vector bosons do not contribute to the evolution and the kernel is purely determined by the unbroken $U(1)_Q$ group of electromagnetism.

Amplitude evolution

We start, following Ref. [105], from the IREE for the scattering amplitude with purely hard external quanta and with regulator λ on the internal lines. While the discussion holds for an arbitrary number of external legs, we focus for definiteness on $2 \rightarrow 2$ amplitudes, which we indicate by

$$\mathcal{M}_\lambda^\alpha = \mathcal{M}_\lambda [p_1(k_1, \alpha_1) p_2(k_2, \alpha_2) \rightarrow p_3(k_3, \alpha_3) p_4(k_4, \alpha_4)], \quad (3.3)$$

where α_i denotes the gauge group index of the external state p_i , which is taken to transform in an irreducible representation of the group. The amplitude is labeled by the 4 indices

$\alpha = \alpha_1 \alpha_2 \alpha_3 \alpha_4$, and it is IR-regulated according to eq. (3.2). Since no real radiation is involved, the cut acts only on the momenta of virtual vector bosons in loop diagrams. We aim at writing down the IREE for $\mathcal{M}_\lambda^\alpha$ and to solve it given the initial condition

$$\mathcal{M}_{E^2}^\alpha = \mathcal{B}^\alpha \equiv \text{BornAmplitude}. \quad (3.4)$$

As we explained, for $\lambda \gg m_W^2$ the effects of EW symmetry breaking (EWSB) can be ignored, and $\mathcal{M}_\lambda^\alpha$ equals the (IR-regulated) amplitude of the unbroken EW gauge theory. More precisely, EWSB gives negligible relative corrections of order $m_W/\sqrt{\lambda}$ (or powers thereof) to all those amplitudes that are not forbidden by the $SU(2)_L \times U(1)_Y$ exact symmetry of the unbroken theory. The other amplitudes are sensitive to EWSB at the leading order and therefore they cannot be studied in the unbroken theory.⁶ However their contribution to the physical scattering process is negligible and they can be safely excluded from the discussion⁷. Similarly, for the allowed processes, up to negligible power corrections of order m_W/E , the amplitude $\mathcal{M}_\lambda^\alpha$ is an $SU(2)_L \times U(1)_Y$ invariant tensor satisfying the charge conservation equation

$$(G_{1^c}^A)^\alpha_\beta \mathcal{M}_\lambda^\beta + (G_{2^c}^A)^\alpha_\beta \mathcal{M}_\lambda^\beta + (G_3^A)^\alpha_\beta \mathcal{M}_\lambda^\beta + (G_4^A)^\alpha_\beta \mathcal{M}_\lambda^\beta \stackrel{\lambda \gg m_W^2}{=} 0, \quad \forall A, \alpha. \quad (3.5)$$

In the equation, G_i^A denotes the generators associated with the representation of each hard particle “ i ” under the EW group, acting only on the corresponding index “ α_i ” of the amplitude tensor. For instance

$$(G_3^A)^\alpha_\beta = \delta_{\beta_1}^{\alpha_1} \delta_{\beta_2}^{\alpha_2} (G_3^A)^{\alpha_3}_{\beta_3} \delta_{\beta_4}^{\alpha_4}. \quad (3.6)$$

Notice that, in our notation, (α_3, α_4) run in the representations of the outgoing states, while (α_1, α_2) run in the conjugate representation of the incoming particles. Consequently in eq. (3.5), $G_{1^c} = -G_1^*$ and $G_{2^c} = -G_2^*$.

The IREE is obtained by computing the variation of the amplitude under an infinitesimal variation $\lambda \rightarrow \lambda + \delta\lambda$ of the IR cutoff in eq. (3.2). This computation dramatically simplifies in the leading DL approximation as one can infer by inspecting diagrams involving a number n of soft/collinear virtual vector bosons. Indeed the maximal logarithm power arises from the region where momenta are hierarchically separated $E^2 \gg \mathfrak{h}(q_1) \gg \mathfrak{h}(q_2) \gg \dots \gg \mathfrak{h}(q_n)$ with the softer legs dressing the subdiagrams involving the harder legs, as shown in the left panel of Figure 3.1. In this configuration only the outermost virtual vector can reach a virtuality $\mathfrak{h}(q_n) \sim \lambda$, the inner ones being much harder in the dominant region of integration. The effect on $\mathcal{M}_\lambda^\alpha$ of the variation of λ is then computed by considering the variation of the endpoint of the integral over the momentum of such outermost vector. More precisely we have that $-\delta\mathcal{M}_\lambda^\alpha \equiv \mathcal{M}_\lambda^\alpha - \mathcal{M}_{\lambda+\delta\lambda}^\alpha$ equals the integral over the outermost loop momentum in

⁶For instance the amplitude with 3 transversely- and one longitudinally-polarized W bosons is suppressed by m_W/E already at the Born level, owing to the fact that it is impossible to form an $SU(2)_L$ singlet with one doublet (i.e., the representation of longitudinal W 's owing to the Equivalence Theorem) and three triplets.

⁷Regarding power suppressed amplitudes, in [129] it has been found that they can receive positive Sudakov enhancements. However due to the power-like suppression we don't expect this to affect our discussion.

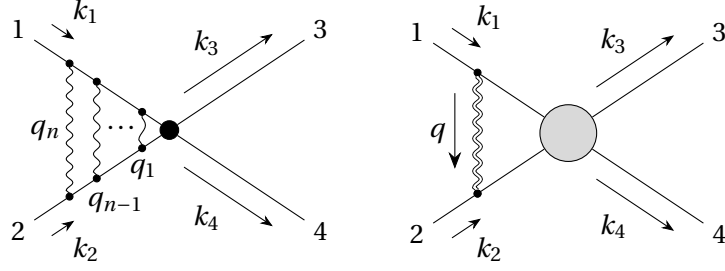


Figure 3.1: Left panel: the leading Sudakov diagrams topology. Right panel: a diagrammatic representation of the contributions to the amplitude variation that are logarithmically enhanced. The momentum q is integrated over the infinitesimal strip $\delta\sigma$ (3.7).

the infinitesimal strip

$$\delta\sigma = \{q : \mathfrak{h}(q) \in [\lambda, \lambda + \delta\lambda]\}. \quad (3.7)$$

The contribution to the variation from the vector that connects a given pair of hard external legs can be depicted like on the right panel of Figure 3.1. The vector boson is represented with a double line to indicate that its momentum q must be integrated only over the strip $\delta\sigma$.

As we already said the leading contribution comes from the integration region where q is soft (and also collinear), in which the vector boson emission is described by the eikonal formula

$$\mathcal{M} \xrightarrow{k_i} \alpha_i \simeq \frac{k_i^\mu}{k_i \cdot q} (G_i^A)^\alpha{}_\beta \mathcal{M}^\beta, \quad \alpha_i \xrightarrow{k_i} \mathcal{M} \simeq \frac{k_i^\mu}{k_i \cdot q} (G_{ic}^A)^\alpha{}_\beta \mathcal{M}^\beta, \quad (3.8)$$

with G_i the group generator acting on particle “ i ” as in eq. (3.6). In line with our conventions, as explained above, the generators of the charge-conjugate representation G_{ic} appear in the eikonal formula for vector boson emission from an incoming particle. For brevity, we have included the gauge coupling constants in the definition of the generators G_i . In terms of the canonical $\text{SU}(2)_L \times \text{U}(1)_Y$ generators we then have

$$G_i^{1,2,3} = g T_i^{1,2,3}, \quad G_i^Y = g' Y_i. \quad (3.9)$$

The integration over the soft q momentum factorizes with respect to the integral over the harder lines, represented as a blob in Figure 3.1. Indeed in the eikonal ($q \rightarrow 0$) limit the virtual vector boson momentum can be neglected and the sub-amplitude blob evaluated on the momenta k_i of the external legs before the virtual vector boson emission/absorption. Therefore the blob gives us back the original amplitude, with one less loop but this is immaterial as $\mathcal{M}_\lambda^\alpha$ is the all-loops amplitude. We can thus express the amplitude variation in terms of the amplitude itself, eventually obtaining an evolution equation. In covariant gauges, the leading DL contributions only arise from virtual vectors connecting two distinct external lines.

Therefore, we have

$$\delta \mathcal{M}_\lambda^\alpha = \frac{-i}{(2\pi)^4} \sum_{j < i} \int_{\delta\sigma} d^4 q \frac{1}{q^2 + i\epsilon} \frac{k_i \cdot k_j}{(q \cdot k_i)(q \cdot k_j)} [\sum_A G_i^A \cdot G_j^A]^\alpha_\beta \mathcal{M}_\lambda^\beta, \quad (3.10)$$

where the sum extends over the unordered ij pairs of distinct external legs and it is understood that the conjugate generators must be employed for the incoming legs $i, j = 1, 2$, due to eq. (3.8).

The evaluation of the integral in eq. (3.10) is quite straightforward, and it is reported in Appendix B.1 for completeness. This gives

$$\delta \mathcal{M}_\lambda^\alpha = -\frac{1}{8\pi^2} \frac{\delta\lambda}{\lambda} \log \frac{E^2}{\lambda} \frac{1}{2} \sum_{A,i} (G_i^A)^\alpha_\beta \sum_{j \neq i} (G_j^A)^\beta_\gamma \mathcal{M}_\lambda^\gamma, \quad (3.11)$$

up to non logarithmically enhanced terms. Notice that in the equation we traded the sum over unordered ij pairs for an ordered sum times $1/2$. The sum over $j \neq i$ can be performed using charge conservation according to eq. (3.5), giving

$$\begin{aligned} \delta \mathcal{M}_\lambda^\alpha &= \frac{1}{16\pi^2} \frac{\delta\lambda}{\lambda} \log \frac{E^2}{\lambda} \sum_i [\sum_A G_i^A G_i^A]^\alpha_\beta \mathcal{M}_\lambda^\beta \\ &= \frac{1}{16\pi^2} \frac{\delta\lambda}{\lambda} \log \frac{E^2}{\lambda} \sum_i [g^2 c_i + g'^2 y_i^2] \mathcal{M}_\lambda^\alpha, \end{aligned} \quad (3.12)$$

where for any given external particle with weak isospin spin t_i and hypercharge y_i , the coefficients $c_i = t_i(t_i + 1)$ and y_i^2 are nothing but the Casimirs of respectively $SU(2)_L$ and $U(1)_Y$. We thus recovered the familiar result that, in DL accuracy, IR effects are universal for each individual external particle and purely determined by the Casimir of the corresponding gauge group representation.

We finally obtain an IREE

$$\frac{d\mathcal{M}_\lambda^\alpha}{d\log^2(E^2/\lambda)} = -\frac{1}{2} \mathcal{K} \mathcal{M}_\lambda^\alpha, \text{ where } \mathcal{K} \stackrel{\lambda \gg m_w^2}{=} \frac{1}{16\pi^2} \sum_i [g^2 c_i + g'^2 y_i^2], \quad (3.13)$$

with, since the Casimir operators are proportional to the identity, an evolution kernel \mathcal{K} that is a mere multiplicative constant. Solving eq. (3.13) starting from the initial condition (3.4) gives the amplitude evaluated with an IR cutoff scale $\lambda = m_w^2$

$$\mathcal{M}_{m_w^2}^\alpha = \exp \left[-\sum_i \frac{g^2 c_i + g'^2 y_i^2}{32\pi^2} \log^2(E^2/m_w^2) \right] \mathcal{B}^\alpha. \quad (3.14)$$

In order to continue the amplitude evolution to lower λ , we should now consider the regime $\lambda \ll m_w^2$, write the corresponding IREE and solve them using eq. (3.14) as initial condition. This is straightforward, because we have seen that all that matters for the derivation of the

IREE are the loop integrals in a strip where the virtual radiation hardness is infinitesimally close to the cutoff λ as in eq. (3.7). In this region, a logarithmic enhancement of the amplitude variation only originates from photon exchange diagrams.⁸ The IREE evolution kernel is thus immediately obtained by specifying the previous formulae to the $U(1)_Q$ gauge group of QED

$$\mathcal{K}^{\lambda \ll m_W^2} = \frac{1}{16\pi^2} \sum_i [e^2 q_i^2]. \quad (3.15)$$

Notice that in order to derive the IREE in this regime, only conservation of electric charge must be employed. The conservation of the full $SU(2)_L \times U(1)_Y$ charges of eq. (3.5) is not valid for $\lambda \ll m_W^2$, where the effects of electroweak symmetry breaking are important.

Solving eq. (3.15) produces the regular QED Sudakov factors, which go to zero in the physical limit $\lambda \rightarrow 0$ where the IR regulator is removed. Therefore the amplitude \mathcal{M}_0^α vanishes, and so does the cross-section of the corresponding fully-exclusive scattering process, in which no extra radiation is present in the final state. More inclusive observables need to be considered for a non-vanishing result. One possibility is to allow for the presence of real photon radiation up to an upper threshold of order m_W^2 on the hardness \mathfrak{h} . This defines a cross-section that we denote as *exclusive* because it indeed excludes the radiation of massive EW bosons. In fact, it is easy to check that $\mathfrak{h}(q) > m^2$ for the emission of a real radiation quantum with $q^2 = m^2$. An upper cut $\mathfrak{h}(q) < m_W^2$ then excludes the presence of massive EW bosons in the final state, but allows for (sufficiently) soft photons. Ref. [105] considered this same observable (but calling it “semi-inclusive”) showing that it stops evolving with λ below m_W^2 , due to the cancellation of real and virtual IR effects in QED. Cross-sections that are exclusive according to our definition can thus be computed at the DL accuracy by just squaring the $\lambda = m_W^2$ amplitude (3.14). At the end of the next section we will re-derive the result of Ref. [105] for exclusive cross-sections by a slightly different methodology, which is also suited for the calculation of the other type of cross-sections we are interested in.

Density matrix evolution

It is possible to extend the IREE methodology to more complex quantities than the hard Feynman amplitude. Specifically, we consider the hard “density matrix”⁹

$$\mathcal{D}_\lambda^{\alpha\bar{\alpha}} \equiv \mathcal{M}_\lambda^\alpha (\mathcal{M}_\lambda^{\bar{\alpha}})^* + \sum_{N=1}^{\infty} \int d\text{Ph}_{N,\lambda}^{\mathcal{H}} \sum_{\rho_1 \dots \rho_N} \mathcal{M}_\lambda^{\alpha;\rho} (\mathcal{M}_\lambda^{\bar{\alpha};\rho})^*, \quad (3.16)$$

which incorporates the emission of an arbitrary number N of radiation particles, with gauge group indices denoted as $\rho = \rho_1 \dots \rho_N$. In the equation, $\mathcal{M}_\lambda^\alpha$ is the hard amplitude with no extra emissions as in the previous section, while $\mathcal{M}_\lambda^{\alpha;\rho}$ is the amplitude for the production of the 2 hard particles plus the radiation. The virtual radiation particles exchanged in the

⁸The calculation of the loop integral in Appendix B.1 shows explicitly that no enhancement emerges from the exchange of vectors with mass m_V much larger than λ .

⁹The same object was dubbed “overlap matrix” in Ref. [104].

Feynman diagrams for the amplitude are subject to the IR hardness cutoff λ as in eq. (3.2). The phase-space volume element $d\text{Ph}_{N,\lambda}^{\mathcal{H}} = \prod_{k=1}^N d\text{Ph}_{k,\lambda}^{\mathcal{H}}$ for the emission of real radiation is also constrained by eq. (3.2). The \mathcal{H} superscript refers to the possible presence of an upper cutoff on the radiation hardness $\mathfrak{h}(q) < \mathcal{H}$. In what follows we will first consider processes we define as *semi-inclusive*, for which $\mathcal{H} \sim E^2$. For these processes the upper radiation cut is effectively absent, and plays no role in the discussion. The exclusive processes defined in the previous section instead simply correspond to $\mathcal{H} = m_W^2$.

It should be noted that eq. (3.16) formally violates the conservation of the total energy and momentum, because in the radiation terms we are employing the same hard 4-momenta that obey energy and momentum conservation in the absence of radiation. It is understood that this makes sense only in the presence of an upper cutoff on the total energy and momentum of the radiation, say a one tenth of E . In this way, the radiation plays a minor role in the total balance of energy and momentum conservation or, equivalently, the hard 4-momenta can be readjusted to balance the radiation emission up to small corrections in the corresponding Feynman amplitudes. In practice, the cutoff allows us to factorize the total phase-space into that for radiation, on one hand, and that for the hard $2 \rightarrow 2$ process on the other, with the latter also including the delta function of 4-momentum conservation. The density matrix (3.16) can thus be related to the physical scattering cross-section.

An upper cut $E_{rad} < E/10$ on the total radiation energy and momentum does not affect the predictions at the double logarithm accuracy. Indeed a simple modification of the real radiation integral (see the discussion around eq. (3.18) computation in Appendix B.1 shows that the effect of this cut on the q momentum of the radiated particle merely entails reduction of the double logarithm from $\log^2 E^2/\lambda$ to $\log^2 E_{rad}^2/\lambda$. The difference is then of order $\log E^2/\lambda \times \log E_{rad}^2/E_{rad}^2$ and falls into the same class as single logarithms as long as E/E_{rad} is not too small, with $1/10$ qualifying.

The hard density matrix (3.16) is a simple generalization of the scattering cross-section in which the conjugated amplitude indices $\bar{\alpha}$ are not equal to the indices α of the non-conjugated amplitude. It is a useful generalization because it obeys charge conservation equations similar to eq. (3.5). Namely, in the regime $\lambda \gg m_W^2$, we have

$$\sum_{i=1^c, 2^c, 3, 4} [(G_i^A)_{\beta}^{\alpha} \mathcal{D}_{\lambda}^{\beta\bar{\alpha}} + (G_{i^c}^A)_{\bar{\beta}}^{\alpha} \mathcal{D}_{\lambda}^{\alpha\bar{\beta}}] \stackrel{\lambda \gg m_W^2}{=} 0, \quad \forall A, \alpha, \bar{\alpha}, \quad (3.17)$$

where the obvious relations $[1^c]^c \equiv 1$ $[2^c]^c \equiv 2$ should be understood. That way the generators acting on the indices $\bar{\beta}$ of the complex conjugated amplitude are those of the corresponding charge conjugated representation. Eq. (3.17) holds only for $\lambda \gg m_W^2$, because in this regime both the virtual and the real emissions are nearly insensitive to EWSB effects as previously explained. For $\lambda \ll m_W^2$, only the electric charge generator is conserved.

The IREE can be obtained like in the previous section by computing the variation of \mathcal{D}_{λ} under $\lambda \rightarrow \lambda + \delta\lambda$, taking now also into account also the effect of the IR cutoff on real emission. The

contribution of virtual loop momentum integrals is thus accompanied by that of integrals over the momentum of real radiation. All integrals have to be performed over the infinitesimal strip $\delta\sigma$ defined in eq. (3.7). Logarithmically enhanced terms only arise from the exchange of virtual or real gauge bosons between different external legs ($i \neq j$), like in Figure 3.2. The effects and the corresponding diagrams can be divided into two classes. The first, in the left panel of Figure 3.2, is given by *primary* radiation diagrams where vector bosons are exchanged between the hard legs. The second, in the right panel, is given by *secondary* radiation diagrams where vector bosons connect to at least one real radiation leg.

We will first consider the effects of primary radiation. The virtual radiation integral gives the result already mentioned in eq. (3.11), and, as we show in Appendix B.1, the result is exactly the same for the real radiation integral. The total variation from primary radiation is then

$$\begin{aligned} \delta\mathcal{D}_\lambda^{\alpha\bar{\alpha}} = & -\frac{1}{16\pi^2} \frac{\delta\lambda}{\lambda} \log \frac{E^2}{\lambda} \sum_{i=1^c, 2^c, 3, 4} \sum_A \left[(G_i^A)^\alpha_\beta \sum_{j \neq i} [(G_j^A)^\beta_\gamma \mathcal{D}_\lambda^{\gamma\bar{\alpha}} + (G_{j^c}^A)^{\bar{\alpha}}_{\bar{\beta}} \mathcal{D}_\lambda^{\beta\bar{\beta}}] \right. \\ & \left. + (G_{i^c}^A)^{\bar{\alpha}}_{\bar{\beta}} \sum_{j \neq i} [(G_j^A)^\alpha_\beta \mathcal{D}_\lambda^{\beta\bar{\beta}} + (G_{j^c}^A)^{\bar{\beta}}_{\bar{\gamma}} \mathcal{D}_\lambda^{\alpha\bar{\gamma}}] \right]. \end{aligned} \quad (3.18)$$

The argument of the first sum, over the four external legs, collects the contributions of all the radiation emitted from the leg “ i ” of the amplitude and of the conjugated amplitude. A factor 1/2 is included to avoid double-counting. Notice that both virtual and real radiation connecting one leg with itself is excluded from the sum, because, as we already mentioned, no enhancement arises from those diagrams.

We can now proceed as in the previous section, and use the charge conservation in eq. (3.17) to perform the sum over j in eq. (3.18). We find the IREE

$$\frac{d\mathcal{D}_\lambda^{\alpha\bar{\alpha}}}{d\log^2(E^2/\lambda)} = -\mathcal{K}_{\beta\bar{\beta}}^{\alpha\bar{\alpha}} \mathcal{D}_\lambda^{\beta\bar{\beta}}, \quad (3.19)$$

with an evolution kernel that is the direct sum of universal terms for each external leg

$$\begin{aligned} \mathcal{K}_{\beta\bar{\beta}}^{\alpha\bar{\alpha}} & \stackrel{\lambda \gg m_w^2}{=} \frac{1}{32\pi^2} \sum_i \left[\left[\sum_A G_i^A G_i^A \right]^\alpha_\beta \delta_{\bar{\beta}}^{\bar{\alpha}} + \delta_\beta^\alpha \left[\sum_A G_{i^c}^A G_{i^c}^A \right]^{\bar{\alpha}}_{\bar{\beta}} + 2 \sum_A (G_i^A)^\alpha_\beta (G_{i^c}^A)^{\bar{\alpha}}_{\bar{\beta}} \right] \\ & = \frac{g^2}{16\pi^2} \sum_i \left[c_i \delta_{\beta_i}^{\alpha_i} \delta_{\bar{\beta}_i}^{\bar{\alpha}_i} + \sum_{A=1,2,3} (T_i^A)^{\alpha_i}_{\beta_i} (T_{i^c}^A)^{\bar{\alpha}_i}_{\bar{\beta}_i} \right] \left[\prod_{j \neq i} \delta_{\beta_j}^{\alpha_j} \delta_{\bar{\beta}_j}^{\bar{\alpha}_j} \right] \\ & = \frac{g^2}{16\pi^2} \sum_i [K_i]_{\beta_i \bar{\beta}_i}^{\alpha_i \bar{\alpha}_i} \left[\prod_{j \neq i} \delta_{\beta_j}^{\alpha_j} \delta_{\bar{\beta}_j}^{\bar{\alpha}_j} \right]. \end{aligned} \quad (3.20)$$

The kernel contains one term, provided by the $SU(2)_L$ Casimir $c_i = t_i(t_i + 1)$, which is proportional to the identity in the color indices of the density matrix tensor, plus a non-diagonal term constructed with the $SU(2)_L$ group generators matrices T_i^A of the external legs. Notice that the contribution of the $U(1)_Y$ hypercharge generator cancels.

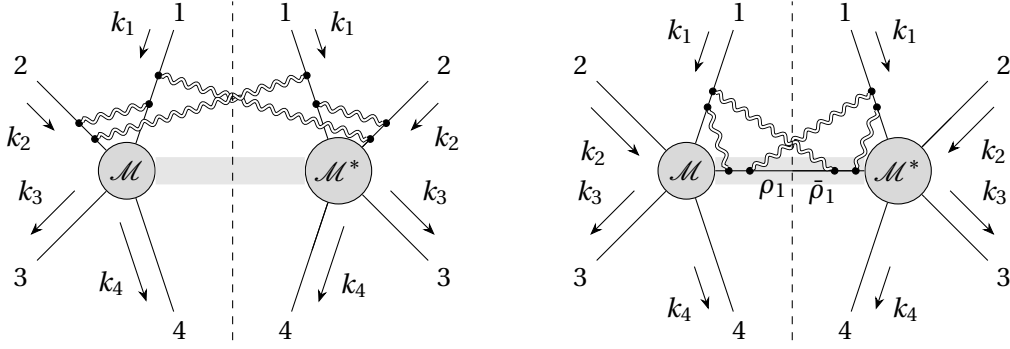


Figure 3.2: Diagrammatic representation of the contributions to the density matrix variation from primary (left panel) and secondary (right panel) radiation. The vector bosons are represented as double lines to indicate that their momenta have to be integrated over the infinitesimal strip (3.7).

There is one peculiarity of eq. (3.20) that is worth emphasizing. The semi-inclusive cross-sections we are interested in are the diagonal entries ($\alpha = \bar{\alpha}$) of the density matrix, with no sum performed over the gauge group index α of the scattering particles.¹⁰ However one can also consider *inclusive* cross-sections, where the sum over the gauge index α_i is performed for one or several external legs. By setting $\bar{\alpha}_i = \alpha_i$ and summing over α_i , the $SU(2)_L$ generators in eq. (3.20) recombine to form the Casimir operator, and the contribution to the evolution kernel from leg “ i ” cancels. We thus find that, at DL accuracy, the cancellation between real and virtual IR effects in inclusive cross-sections occurs on a leg-by-leg basis. Namely, the effects of soft/collinear emissions associated to each individual leg cancel in the cross-section (and in the entire density matrix) for processes that are inclusive over the color of the corresponding particle. This result is stronger than the KLN theorem, which foresees a cancellation only when summing over the color of all legs. The reason for the added strength is that we are here considering radiation that is both soft and collinear. Notice however that fully inclusive observables of practical relevance can only involve summation on the color of the final state particles. This retains the IR effects associated with the colliding particles in the initial state (e.g., two left-handed leptons $\ell_L^+ \ell_L^-$) which are not $SU(2)_L$ singlets. The resulting non-cancellation of IR effects in “fully-inclusive” cross-sections, coincides with the result of Ref. [104].

So far we have ignored the secondary radiation diagrams, depicted in the right panel of Figure 3.2. We show now that their contribution vanishes, giving full justification to eq. (3.20). Secondary radiation diagrams correspond to the effect of the λ cutoff variation on virtual or real vector bosons attached to one of the intermediate “ ρ ” particles in the definition of the density matrix (3.16). Clearly these effects are potentially enhanced only if the intermediate particle is relatively hard, such that a significant separation is present between the IR cutoff λ and the scalar product between the intermediate particle and the external leg momenta. We

¹⁰This is true only in a basis where the gauge indices α_i label the on-shell SM particles, while for the calculation of di-boson cross-sections we work in a different basis. See Section 3.2.3 and Appendix B.2 for details.

thus start considering vector bosons attached to the hardest intermediate particle, with gauge index “ ρ_1 ” as in the figure. The density matrix is inclusive over the color of the intermediate particle. However we can momentarily define an “extended” density matrix $\mathcal{D}_\lambda^{\alpha;\rho_1\bar{\alpha};\bar{\rho}_1}$ with labels ρ_1 and $\bar{\rho}_1$ for the gauge indices of the amplitude and of the conjugate amplitude, as in the figure. The actual density matrix is eventually obtained by first setting $\rho_1 = \bar{\rho}_1$ and then summing. The effect on the extended density matrix variation of all the radiation emitted from ρ_1 and $\bar{\rho}_1$ can be written in a form similar to eq. (3.18) and then simplified using the analog of eq. (3.17) for the extended density matrix. The resulting contribution to the evolution kernel from the intermediate ρ_1 leg is the analog of that from the hard external legs in eq. (3.20). But this contribution cancels out in the evolution of the actual density matrix, which is inclusive over the ρ_1 leg, because of the previously explained leg-by-leg cancellation mechanism. The argument can of course be repeated for the diagrams involving the second hardest intermediate particle, showing, as anticipated, that all the secondary radiation diagrams can be ignored in the calculation of the evolution kernel.

It is straightforward to adapt the previous results to the regime $\lambda \ll m_W^2$, in which only the exchange of photons contributes to the evolution, as discussed in the previous section. By specifying eq. (3.20) to the Abelian $U(1)_Q$ group we immediately find that the kernel vanishes, owing to the well-known cancellation between real and virtual IR effects in QED. For the calculation of the physical ($\lambda \rightarrow 0$) density matrix, and in turn of the semi-inclusive cross-section, we thus only need to solve the IREE with the $\lambda \gg m_W^2$ kernel (3.20), down to $\lambda = m_W^2$.

For $\lambda = E^2$ the hard density matrix (3.16) is well-approximated by its tree-level expression, which serves as the initial condition for the evolution

$$\mathcal{D}_{E^2}^{\alpha\bar{\alpha}} = \mathcal{B}^\alpha(\mathcal{B}^{\bar{\alpha}})^* . \quad (3.21)$$

The kernel is the direct sum of tensors, denoted as K_i in eq. (3.20), each acting on the pair $\alpha_i, \bar{\alpha}_i$ associated to the i -th external particle. Therefore the solution of the IREE reads

$$\mathcal{D}_{\text{si}}^{\alpha\bar{\alpha}} \equiv \mathcal{D}_{m_W^2}^{\alpha\bar{\alpha}} = \left\{ \prod_i \left[\exp \left[-\frac{g^2}{16\pi^2} K_i \log^2(E^2/m_W^2) \right] \right]_{\beta_i \bar{\beta}_i}^{\alpha_i \bar{\alpha}_i} \right\} \mathcal{B}^\beta(\mathcal{B}^{\bar{\beta}})^* , \quad (3.22)$$

where the “_{si}” subscript denotes the density matrix of the semi-inclusive process, with no upper cut on the real radiation hardness. The explicit form of the K_i exponentials in the above equation is reported in eq.s (B.31) and (B.36) for external legs in the doublet and triplet $SU(2)_L$ representations. Applications of eq. (3.22) to specific processes are shown in Sections 3.2.2 and 3.2.3.

We have defined the density matrix (3.16) allowing for the presence of an upper cutoff \mathcal{H} on the real radiation, but this played no role in the previous discussion because this cutoff is effectively absent ($\mathcal{H} \sim E^2$) in our definition of semi-inclusive processes. In exclusive processes we instead set $\mathcal{H} = m_W^2$, namely we veto real radiation particles with hardness above m_W^2 . Obviously, for $\lambda \gg m_W^2$ this upper cut is in contradiction with the IR cutoff in eq. (3.2)

on the radiation phase-space. Therefore in the density matrix for the exclusive process no real radiation is present and in the $\lambda \gg m_W^2$ regime the result simply equals the square of the hard amplitude in eq. (3.3). The evolution up to $\lambda = m_W^2$ can thus be obtained from the hard amplitude evolution (3.14) we obtained in the previous section, or easily re-derived by dropping the terms in eq. (3.18) (namely, the second and the third) that are due to real radiation. The contribution of real radiation is instead restored for $\lambda \ll m_W^2$ and the evolution stops due to the cancellation between virtual and real QED radiation as previously explained. The physical ($\lambda \rightarrow 0$) density matrix for exclusive processes can thus be written in a simple closed form as

$$\mathcal{D}_{\text{ex}}^{\alpha\bar{\alpha}} = \exp \left[- \sum_i \frac{g^2 c_i + g'^2 y_i^2}{16\pi^2} \log^2(E^2/m_W^2) \right] \mathcal{B}^\alpha (\mathcal{B}^{\bar{\alpha}})^* . \quad (3.23)$$

In Sections 3.2.2 and 3.2.3 we employ this formula to compute exclusive di-fermion and di-boson production cross-sections, and discuss the need of supplementing it with fixed-order single-logarithmic terms, from Ref. [127, 128].

Before concluding this section it is worth commenting on the experimental definition of the semi-inclusive and exclusive processes, and on the perspectives for their actual experimental detectability. The semi-inclusive process is characterized by two central (specifically, emitted from 30 to 150 degrees from the beam line) energetic particles of specific EW color and flavor. In particular we will require them to carry a total center of mass energy above 85% of the VHEL E_{cm} , enforcing in this way the upper cut on the total radiation 4-momentum required for the definition of the hard density matrix as discussed below eq. (3.16). The two particles can be accompanied by the radiation of EW bosons, photons, or any other soft particle.

Notice that in our calculation at the DL order we could ignore all the effects of collinear (rather than soft-collinear) radiation, which emerge at the single logarithm. On the other hand, the single logarithms associated with low-virtuality (below m_W) photon splittings are much larger than $\log E^2/m_W^2$. In particular, the emission of real photons that are energetic but collinear to a light charged hard particle (e.g., an electron or a muon) with mass m_ℓ produces terms proportional to $\log E^2/m_\ell^2$. By the KLN theorem these terms will be canceled by the corresponding virtual contributions, but only in suitably-defined observables that recombine the emitted photons in the experimental definition of the hard particle 4-momentum. With a lower threshold of order m_W on the energy of the photons to be recombined, the net effect on our prediction should be of the order of a single EW logarithm $\log E^2/m_W^2$. A more detailed assessment of this aspect, and of the possible interplay between the QED and the EW bosons collinear emissions, requires the inclusion of single logarithms and goes beyond the scope of the present chapter. Similar considerations hold for the collinear emission of QCD gluons to be collected into jets, in the case of colored final states.

Up to the caveats outlined above, there are good perspectives for the actual direct experimental detectability of semi-inclusive cross-sections. The situation is arguably more problematic for the exclusive cross-section. In exclusive final states, we require the presence of the two hard

particles defined as above, plus the absence of any massive vector boson (since $\mathfrak{h}(q) > q^2 = m^2$, as discussed at the end of the previous section), or photons above the hardness upper threshold m_W^2 . However, it is experimentally impossible to impose this radiation veto strictly because the limited coverage of the detector in the forward and backward regions will not allow to tag EW bosons or photons that are collinear to the beam. Furthermore our definition of the exclusive cross-section is problematic in the case of QCD-colored final states. Indeed if the upper cut $\mathfrak{h}(q) < m_W^2$ had to be imposed also on gluon radiation, QCD effects should be included in the exclusive density matrix evolution (but not in the semi-inclusive one, where they cancel because of color inclusivity), resulting in a large QCD Sudakov suppression factor in eq. (3.23). This factor is as small as $\exp[-\alpha_s/(4\pi)(8/3)\log^2 E_{\text{cm}}^2/m_W^2] \sim 0.03$ for di-quark final states at the highest VHEL energy $E_{\text{cm}} = 30$ TeV, entailing a strong suppression of the cross-section. Avoiding this suppression requires a definition of the exclusive cross-section with a higher threshold on the QCD radiation. We will further comment in the Conclusions on the limitations of the exclusive cross-section definition employed in this chapter.

3.2.2 Di-fermion production

The first process we investigate is the production of a highly energetic pair of fermions

$$\ell^+(k_1) \ell^-(k_2) \rightarrow \bar{f}(k_3) g(k_4) + X, \quad (3.24)$$

where f and g can be one of the six quarks, a lepton $\ell' \neq \ell$ or a neutrino $\nu_{\ell'}$. We do not discuss explicitly the final states with the same leptonic flavor as the initial state, $\ell' = \ell$, but these processes will be employed for the muon collider sensitivity projections in Section 3.3. As previously discussed, the final state is characterized (both for exclusive and semi-inclusive processes) by an invariant mass for the (\bar{f}, g) pair that is almost equal to the center of mass energy E_{cm} of the colliding leptons and by central scattering angle $\theta_* \in [30^\circ, 150^\circ]$. Here θ_* is the angle between the incoming ℓ^+ and the final anti-fermion \bar{f} in the lab frame. Notice that θ_* almost coincides with the scattering angle in the center of mass frame of the hard process, because of the tight cut on the invariant mass of the (\bar{f}, g) pair.

In order to resum the DL it is convenient to organize the calculation of the cross-section in terms of amplitudes and density matrices whose external legs are canonical irreducible representations of the EW group. This is trivial to achieve for the di-fermion process because the helicity eigenstates of quarks and leptons in the massless limit do indeed transform as canonical representations (doublets and singlets, with specific hypercharge), reported for completeness in Appendix B.2. Furthermore, since we restrict our attention to inelastic processes $\ell' \neq \ell$, the only sizable helicity amplitudes are those with the same chirality χ_I (χ_O) for the two incoming (outgoing) fermions, corresponding to helicities $\bar{\psi}_{+1/2}\psi_{-1/2}$ for $\chi = L$ and $\bar{\psi}_{-1/2}\psi_{+1/2}$ for $\chi = R$. The dominance of such amplitudes holds in the SM because of the vector-like structure of gauge interaction, and it will be preserved by the 4-fermions new physics contact interaction operators we will study in Section 3.3. We thus have to deal with four polarized cross-sections for each di-fermion production process, labeled by

	3 TeV			10 TeV			30 TeV		
	DL	$e^{\text{DL}-1}$	$\text{SL}(\frac{\pi}{2})$	DL	$e^{\text{DL}-1}$	$\text{SL}(\frac{\pi}{2})$	DL	$e^{\text{DL}-1}$	$\text{SL}(\frac{\pi}{2})$
$\ell_L \rightarrow \ell'_L$	-0.46	-0.37	0.25	-0.82	-0.56	0.33	-1.23	-0.71	0.41
$\ell_L \rightarrow q_L$	-0.44	-0.36	0.25	-0.78	-0.54	0.34	-1.18	-0.69	0.42
$\ell_L \rightarrow e_R$	-0.32	-0.27	0.13	-0.56	-0.43	0.17	-0.85	-0.57	0.21
$\ell_L \rightarrow u_R$	-0.27	-0.24	0.11	-0.48	-0.38	0.15	-0.72	-0.51	0.18
$\ell_L \rightarrow d_R$	-0.24	-0.21	0.10	-0.43	-0.35	0.13	-0.64	-0.47	0.16
$\ell_R \rightarrow \ell'_L$	-0.32	-0.27	0.13	-0.56	-0.43	0.17	-0.85	-0.57	0.21
$\ell_R \rightarrow q_L$	-0.30	-0.26	0.12	-0.53	-0.41	0.16	-0.79	-0.55	0.21
$\ell_R \rightarrow \ell'_R$	-0.17	-0.16	0.07	-0.30	-0.26	0.09	-0.46	-0.37	0.12
$\ell_R \rightarrow u_R$	-0.12	-0.12	0.05	-0.22	-0.20	0.07	-0.33	-0.28	0.08
$\ell_R \rightarrow d_R$	-0.09	-0.09	0.04	-0.17	-0.16	0.05	-0.25	-0.22	0.06

Table 3.1: Double and single logarithmic corrections to the exclusive processes $\ell^+ \ell^- \rightarrow \bar{f} f$. The single-logarithmic corrections are evaluated at $\theta_* = \pi/2$.

$\chi_I \chi_O = LL, LR, RL, RR$. Each such cross-section will be obtained from the diagonal $\alpha = \bar{\alpha}$ entries of the density matrices of Section 3.2.1, times the appropriate phase-space factors.

Exclusive processes

Exclusive cross-sections are readily obtained from eq. (3.23), and take the form

$$\frac{d\sigma_{\text{ex}}}{d\cos\theta_*} = e^{\text{DL}} \frac{d\sigma_B}{d\cos\theta_*}, \quad (3.25)$$

in terms of the corresponding Born-level differential cross-sections. The Double Log exponent DL is of order $g^2/16\pi^2 \log^2(E_{\text{cm}}^2/m_W^2)$, which ranges from 0.14 at $E_{\text{cm}} = 3$ TeV up to 0.25 (0.38) for $E_{\text{cm}} = 10(30)$ TeV, times the sum of the four SU(2) Casimir of the external legs. For LL chirality processes this factor is as large as $4 \times 1/2(1/2 + 1) = 3$, showing that DL resummation is mandatory at VHEL energies $E_{\text{cm}} \geq 10$ TeV, at least for this chirality. Double logs are still considerable for LR and RL chirality, while they get smaller in the RR configuration because $g'^2 \sim g^2/4$. Resummation might instead not be necessary for $E_{\text{cm}} = 3$ TeV. However it will still be needed to include the effects of radiation at fixed order since we aim, eventually, at theoretical predictions with percent-level accuracy.

The DL Sudakov exponents in eq. (3.25) are listed in Table 3.1. The processes are labeled taking into account that electric charge conservation enforces $g = f$ in eq. (3.24), since a charge mismatch cannot be compensated by the emission of charged W bosons, which is forbidden in exclusive processes. The table also reports single logarithm (SL) contributions computed at the fixed one loop order, which we extract from Ref.s [127].¹¹ Specifically, we employ the general formulae of Ref.s [127] to compute the 1-loop log-enhanced cross-section, we subtract

¹¹Two loops NLL results for four-fermion processes are also available in [130, 131].

the corresponding DL and normalize to the Born cross-section. We also subtract the single logarithms from the Renormalization Group evolution, because we decided to compute the Born amplitude with the EW couplings at the hard scale E_{cm} .¹² Notice that the threshold for photon recombination into the hard final state particles matters at the single-logarithmic order. Here we assume a scale of recombination of order m_W , for which the SL terms can be easily obtained by adding a fictitious photon mass $m_\gamma = m_W$ to the calculations of Ref.s [127, 128]. The SL terms obtained in this way can be used for “improved” theoretical predictions

$$\frac{d\sigma_{\text{ex}}^{\text{SL}_1}}{d\cos\theta_*} = e^{\text{DL}}(1 + \text{SL}(\theta_*)) \frac{d\sigma_B}{d\cos\theta_*}, \quad (3.26)$$

that include single logarithms at fixed 1-loop order. We see in Table 3.1 that the SL contributions are relatively large. It is unclear whether they require resummation or if including them at fixed order (definitely higher than 1-loop, if we target 1% accuracy) is sufficient.

Notice that, unlike double logarithms, the single logarithm contributions are not proportional to the Born-level amplitude of the same scattering process. Namely the amplitudes of the neutral-current processes in Table 3.1 receive SL corrections that are proportional to Born charged-current amplitudes. Therefore it should be kept in mind the SL terms in eq. (3.26), which we normalized to the Born cross-section of the process, depend on the ratio between charged and neutral current Born amplitudes. We evaluated the amplitude ratio within the SM to produce the results in Table 3.1. However the amplitude ratio depends on the new physics contact interactions we consider in Section 3.3, entailing a dependence of the SL terms on the new physics parameters. This is not the case for the double logarithms, which are completely universal and insensitive to short-distance physics. The single logarithms also carry a non-trivial dependence on the scattering angle θ_* , as explicitly indicated in eq. (3.26). In Table 3.1 they are evaluated at central angle $\theta_* = \pi/2$, where they are always positive. They can become negative, and typically increase in magnitude, in the forward and backward scattering regions, which we however exclude with the central cut $\theta_* \in [30^\circ, 150^\circ]$. Finally, notice that the SL terms are affected by the sizable mass of the top quark, which we do include in the $t\bar{t}$ production process.

The impact of EW radiation effects on the total (unpolarized) cross-section in the central region, relative to the Born, is displayed in Figure 3.3 as a function of E_{cm} . The production of two light up-type quarks is considered for illustration, but the results for the other final states are similar. The blue line is the one-loop DL prediction without exponentiation, while in red we report the resummed DL prediction in eq. (3.25). The green line (labeled $\text{DL}_1 + \text{SL}_1$) represents the fixed-order one loop DL plus SL, while in black we report the SL-improved prediction in eq. (3.26). The dashed lines are semi-inclusive cross-sections computed below. We notice a significant cancellation between double and single logarithmic terms. However this cancellation is not expected to be structural and to survive at higher orders in perturbation

¹²The calculation is similar to the one performed by two of us in Ref. [132]. We refer the reader to Section 2.3 of [132] for additional details, concerning in particular the inclusion of non-log-enhanced angular-dependent terms.

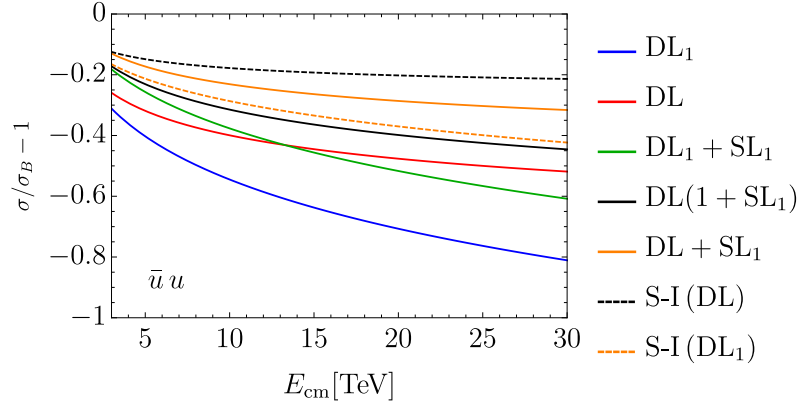


Figure 3.3: Impact of radiative corrections on the production of two up quarks at the VHEL. The solid lines represent different predictions for the exclusive cross-section. The dashed lines are double-logarithm semi-inclusive cross-sections resummed (in black) or at one loop (in orange).

theory.

We do not try to assign theoretical uncertainties to our predictions. However an upper bound can be obtained by considering the orange line in the figure, in which the resummed DL are combined additively with the SL (i.e., as $e^{\text{DL}} + \text{SL}$), rather than multiplicatively. An alternative estimate of the uncertainties could be obtained by varying the scale of the EW couplings employed for the evaluation of the radiation terms DL and SL. Varying this scale from m_W (which we employ for our predictions) to E_{cm} , the relative change of the radiation effects is rather small, typically at the 10% level or less.

Semi-inclusive processes

The semi-inclusive cross-sections are the diagonal $\alpha = \bar{\alpha}$ entries of the semi-inclusive density matrix in eq. (3.22), with the appropriate K_i exponential factors for each external leg. The factors only depend on the $\text{SU}(2)_L$ quantum numbers of the legs and not of their hypercharge (and QCD color). They are provided by eq. (B.31) for L -chirality external external legs (which transform as doublets or conjugate-doublets) and they are trivial for the R -handed singlets. Notice that eq. (B.31) (and the same is true for the triplet exponential factor (B.36)) does not mix diagonal with off-diagonal entries of the density matrix. Namely if we set $\alpha = \bar{\alpha}$ we obtain a tensor that is diagonal in β and $\bar{\beta}$. Therefore the DL resummed cross-sections, collected in a vector $d\vec{\sigma}_{\text{si}}$, are linear combinations of the Born cross-sections $d\vec{\sigma}_B$. We express this relation as

$$\frac{d\vec{\sigma}_{\text{si}}}{d\cos\theta_*} = e^{\text{DL}} \cdot \frac{d\vec{\sigma}_B}{d\cos\theta_*}, \quad (3.27)$$

where the Double-Logarithm terms “DL” are now matrices connecting the Born cross-sections of different processes unlike for exclusive processes (3.25).

For an explicit illustration of the semi-inclusive cross-section calculation, and of the main features of the result, we consider the RL -chirality production processes. In this case, we have

$$d\vec{\sigma}_{\text{si}} = \begin{pmatrix} d\sigma_{\text{si}}(\ell_R^+ \ell_R^- \rightarrow \bar{u}_L u_L) \\ d\sigma_{\text{si}}(\ell_R^+ \ell_R^- \rightarrow \bar{u}_L d_L) \\ d\sigma_{\text{si}}(\ell_R^+ \ell_R^- \rightarrow \bar{d}_L u_L) \\ d\sigma_{\text{si}}(\ell_R^+ \ell_R^- \rightarrow \bar{d}_L d_L) \end{pmatrix}, \quad d\vec{\sigma}_B = \begin{pmatrix} d\sigma_B(\ell_R^+ \ell_R^- \rightarrow \bar{u}_L u_L) \\ 0 \\ 0 \\ d\sigma_B(\ell_R^+ \ell_R^- \rightarrow \bar{d}_L d_L) \end{pmatrix} = d\sigma_B(\ell_R^+ \ell_R^- \rightarrow \bar{q}_L q_L) \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix}, \quad (3.28)$$

where “ u ” and “ d ” denote here the up and down components of a L -handed fermion doublet. The exponentiated DL matrix reads

$$e^{\text{DL}} = \frac{1}{4} e^{-\mathcal{L}} \begin{pmatrix} 4 \cosh^2(\mathcal{L}/2) & 2 \sinh(\mathcal{L}) & 2 \sinh(\mathcal{L}) & 4 \sinh^2(\mathcal{L}/2) \\ 2 \sinh(\mathcal{L}) & 4 \cosh^2(\mathcal{L}/2) & 4 \sinh^2(\mathcal{L}/2) & 2 \sinh(\mathcal{L}) \\ 2 \sinh(\mathcal{L}) & 4 \sinh^2(\mathcal{L}/2) & 4 \cosh^2(\mathcal{L}/2) & 2 \sinh(\mathcal{L}) \\ 4 \sinh^2(\mathcal{L}/2) & 2 \sinh(\mathcal{L}) & 2 \sinh(\mathcal{L}) & 4 \cosh^2(\mathcal{L}/2) \end{pmatrix}, \quad (3.29)$$

where $\mathcal{L} = g^2/16\pi^2 \log^2(E_{cm}^2/m_W^2)$.

We see that DL effects induce a non-vanishing cross-section for charged processes with $g \neq f$ in eq. (3.24), such as $\bar{u}_L d_L$ and $\bar{d}_L u_L$ production. Clearly this stems from the emission of real soft W -bosons, which is allowed in the semi-inclusive final state. Such charged cross-sections are proportional to the Born cross-section for the corresponding neutral ($\bar{u}_L u_L$ or $\bar{d}_L d_L$) processes, and they are not drastically smaller than those because the double-logarithm is sizable at VHEL energies. Therefore they can be measured bringing additional sensitivity to the charge-preserving Born amplitudes and to the corresponding short-distance new physics effects. The interplay with short-distance physics is even more interesting for the LL -chirality process. In that case, $\vec{\sigma}_{\text{si}}$ is a 16-dimensional vector that contains 4 observable ($\ell^+ \ell^-$ -initiated) processes with final states $\bar{u}_L u_L$, $\bar{u}_L d_L$, $\bar{d}_L u_L$ and $\bar{d}_L d_L$. DL is a 16×16 matrix that relates the observable processes to 16 Born amplitudes, among which those (like, e.g., $\bar{\nu}_\ell \ell^- \rightarrow \bar{u}_L d_L$) that are sensitive to new charged current interactions. We can thus probe the latter interactions even with the neutral $\ell^+ \ell^-$ VHEL collisions.

The black dashed lines in Figure 3.3 quantify the impact of the EW radiation effects on the neutral semi-inclusive cross-sections relative to the Born predictions. The effects are smaller than for exclusive cross-sections, as qualitatively expected owing to the partial cancellation between virtual and real radiation. While this suggests that resummation might play a less relevant role in semi-inclusive predictions, we point out that one-loop double logarithms are insufficient for accurate predictions. This is shown in the purple dashed line in the figure, which is obtained by truncating at the one-loop order the exponentiated DL matrix. It would be interesting to study the impact of single logarithms on the predictions. This could be achieved by combining the single radiative logs from Ref.s [127, 128] with the factorized formulas for real emissions in Ref.s [120, 43] (which however would have to be extended to include also the soft radiation region), but is left to future work.

As a final technical note, we remark that the DL matrix is negative semi-defined with a single vanishing eigenvector that corresponds to the “fully-inclusive” cross-section, further averaged over the $SU(2)_L$ color of the initial states. Specifically the vanishing eigenvector of eq. (3.29) is $(1, 1, 1, 1)^t$, which corresponds to the sum of the cross-sections over the $SU(2)_L$ gauge indices of the final states. Therefore in this case the double logarithmic effects cancel on the “fully-inclusive” cross-section, in accordance with the KLN theorem since the right-handed initial leptons are $SU(2)_L$ singlets. Clearly this does not happen for the LL -chirality processes (nor for LR -chirality) and the average over leptons and neutrinos in the initial states would be necessary for the cancellation. The vanishing eigenvalue controls the behavior of the DL exponential at asymptotically high energies. In the case of eq. (3.29), we have

$$e^{\text{DL}} \xrightarrow{E_{\text{cm}} \rightarrow \infty} \frac{1}{4} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}, \quad (3.30)$$

and all the semi-inclusive cross-section listed in eq (3.28) become equal. Notice however this only holds at asymptotic energies, way above the VHEL energies. Cross-sections equality becomes a reasonable (better than order-one) approximation only for if $g^2/16\pi^2 \log^2(E_{\text{cm}}^2/m_W^2)$ is as large as ~ 1.5 , i.e. $E_{\text{cm}} \gtrsim 10000$ TeV.

3.2.3 Di-boson production

We now turn to the production of two energetic vector or Higgs bosons. We are interested in reactions that are not power-like suppressed at high energy, therefore we restrict our attention to “longitudinal” processes entailing the production of zero-helicity W and Z bosons and Higgs, and to “transverse” di-boson processes where the W and the Z (or, the photon) have ± 1 helicities. Indeed the “mixed” longitudinal/transverse production processes are suppressed by m_W/E_{cm} at the amplitude level, as readily understood (see e.g. [123, 43]) by combining the Goldstone Boson Equivalence Theorem with the selection rules associated with the $SU(2)_L \times U(1)_Y$ SM group.

The new physics interactions we consider in Section 3.3 only affect longitudinal di-boson production cross-sections, which thus play the role of the signal in our analysis. We nevertheless also need the transverse cross-sections for an estimate of the background. We discuss the calculation of the (exclusive and semi-inclusive) cross-sections for the two type of processes in turn.

Longitudinal di-boson

We consider the production, out of $\ell^+ \ell^-$, of one of the following hard final states

$$W_0^+ W_0^-, \quad Z_0 h, \quad W_0^\pm Z_0, \quad W_0^\pm h, \quad (3.31)$$

where the subscript “0” refers to the helicity of the massive vectors, and “ h ” denotes the physical Higgs particle. Obviously only the first two final states can be produced in an exclusive process, while the latter ones require the emission of at least one charged W and therefore they only occur at the semi-inclusive level. Notice that the ones listed above are the only hard final states with longitudinal bosons and Higgs that can be produced by soft EW bosons radiation out of sizable Born-level $2 \rightarrow 2$ cross-sections. Therefore they are the only longitudinal di-boson processes that can be considered for precise VHEL measurements in the high-energy regime.

At energies much above m_W , the adequate description of longitudinally-polarized massive vectors is provided by the charged and neutral Goldstone boson scalars π^\pm and π_0 (see Appendix B.2). Together with the Higgs, they form a canonical $SU(2)_L \times U(1)_Y$ doublet H with $1/2$ hypercharge, reported in eq. (B.26). We thus need to consider amplitudes and density matrices associated with the hard processes

$$\begin{aligned} \ell_{-1/2}^+(k_1) \ell_{+1/2}^-(k_1) &\rightarrow \bar{H}(k_3, \alpha_3^{\bar{d}}) H(k_4, \alpha_4^d), \\ \bar{\ell}_{+1/2}(k_1, \alpha_1^{\bar{d}}) \ell_{-1/2}(k_2, \alpha_2^d) &\rightarrow \bar{H}(k_3, \alpha_3^{\bar{d}}) H(k_4, \alpha_4^d), \end{aligned} \quad (3.32)$$

for, respectively, L -handed and R -handed production.¹³ For the gauge group indices we employ the same notation as in eq. (3.3), supplemented by the superscripts d (\bar{d}) to indicate that the indices belong to the doublet (conjugate-doublet) representation. With a slight abuse of notation we are denoting as $\ell_{-1/2} = (v_{\ell, -1/2}, \ell_{-1/2}^-)^t$ the lepton doublet with $-1/2$ helicity and with $\ell_{+1/2}$ the conjugate-doublet with helicity $+1/2$. Notice that final states with two H or two \bar{H} need not to be included because they are power-like suppressed at high energy by hypercharge conservation.

The relevant density matrices are obtained as a straightforward application of the results in Section 3.2.1. The need for employing H and \bar{H} as external states does not pose any additional difficulty (relative to the di-fermion processes) in the evaluation of exclusive cross-sections. That is because the double logs are mere multiplicative factors in front of the Born-level density matrix (3.23). Therefore the exclusive cross-sections still take the form of eq. (3.25) and are proportional to the corresponding Born-level predictions. For the semi-inclusive cross-section, we can proceed as for di-fermions in the determination of the K_i exponential factors, using in particular eq. (B.31) which also holds in the present case because H and \bar{H} are doublets. However in order to apply eq. (3.22) we must first express the $\mathcal{D}_{\text{si}}^{\alpha\bar{\alpha}}$ density matrix, which is written in the isospin basis (H and \bar{H}), in the physical basis of the charge and CP eigenstates h , $Z_0 = \pi_0$ and $W_0^\pm = \pi^\pm$. This is achieved by simply inverting eq. (B.26). The final result can again be expressed in terms of the Born-level cross-sections in the form of eq. (3.27).

The results display the same qualitative features as di-fermions. In particular we observe the same interplay between short-distance physics affecting the neutral- and the charged-current

¹³The production from opposite-chirality leptons is negligible, both in the SM and in the presence of the new contact interactions we investigate in the following section.

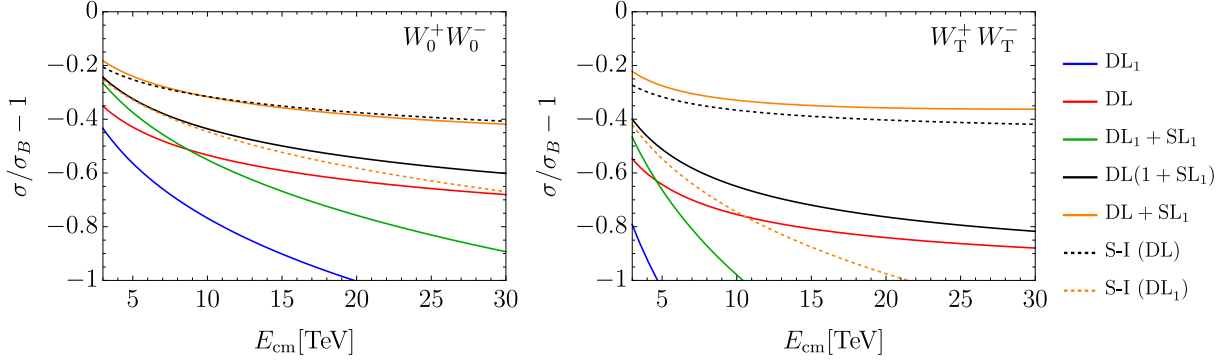


Figure 3.4: Same as Figure 3.3, but for di-boson production. As explained in Section 3.3.2, the cross-sections for $W^+ W^-$ production are integrated in the angular region $\theta_* \in [67^\circ, 150^\circ]$.

Born amplitudes, which we investigate in Section 3.3 in details. Also at the quantitative level, the relative impact of radiation radiation is similar, as expected because $SU(2)_L$ doublets are involved also in these processes. This is shown in the left panel of Figure 3.4, where we show the exclusive and semi-inclusive cross-section predictions for $W_0^+ W_0^-$. The different predictions are obtained as explained in the previous section for the di-fermion processes. Notice in particular the exclusive predictions that include one-loop single logarithms as in eq. (3.26). We employ these predictions for exclusive cross-section in the phenomenological studies of Section 3.2.1.

Transverse di-boson

Vector bosons (W , Z , or γ) with transverse helicity $T = \pm 1$ have zero hypercharge and they decompose as a real triplet plus a singlet under the SM $SU(2)_L$, as in eq. (B.27). Therefore three non-power-suppressed hard processes have to be considered for L -handed production

$$\begin{aligned} \bar{\mu}_{+1/2}(k_1, \alpha_1^{\bar{d}}) \ell_{-1/2}(k_2, \alpha_2^d) &\rightarrow B(k_3) B(k_4), \\ \bar{\mu}_{+1/2}(k_1, \alpha_1^{\bar{d}}) \ell_{-1/2}(k_2, \alpha_2^d) &\rightarrow W(k_3, \alpha_3^t) B(k_4), \\ \bar{\mu}_{+1/2}(k_1, \alpha_1^{\bar{d}}) \ell_{-1/2}(k_2, \alpha_2^d) &\rightarrow W(k_3, \alpha_3^t) W(k_4, \alpha_4^t), \end{aligned} \quad (3.33)$$

while only one is relevant for the production initiated by R -handed leptons¹⁴

$$\ell_{-1/2}^+(k_1) \ell_{+1/2}^-(k_2) \rightarrow B(k_3) B(k_4), \quad (3.34)$$

The “t” superscript in eq. (3.33) refers to the triplet nature of the W indices.

Unlike for di-fermion and longitudinal di-boson, the transverse di-boson cross-sections for L -handed initial leptons are linear combinations of several distinct density matrices with different $SU(2)_L$ quantum numbers. Therefore the exclusive cross-sections are not proportional,

¹⁴The Born process $\ell_{-1/2}^+(k_1) \ell_{+1/2}^-(k_2) \rightarrow W(k_3, \alpha_3^t) W(k_4, \alpha_4^t)$ is power-suppressed in the SM.

unlike in eq. (3.25), to the corresponding Born cross-sections in general. For instance in the $\gamma\gamma$ cross-section the contribution from the WW final state experiences a stronger Sudakov suppression (3.23) than one from the BB (or WB) final state, owing to the higher $SU(2)_L$ Casimir of the final states.

The evaluation of the semi-inclusive cross-sections proceeds as for the longitudinal di-bosons. Namely we derive the cross-sections for the physical states by inverting eq. (B.27) and we compute the double-logarithm exponentials using eq. (B.36) on the $SU(2)$ triplet subspace. Clearly the need of combining different density matrices complicates the calculation, but it does not introduce any novel conceptual aspect. At the quantitative level instead, the situation is significantly different than for di-fermions and longitudinal di-bosons. As shown on the right panel of Figure 3.4, EW radiation effects are much larger due to the larger Casimir $c_t = 2$ of the triplet representation. A sufficiently accurate modeling of these effects will probably require resummation even at the lowest VHEL energy $E_{\text{cm}} = 3$ TeV.

The figure reports the cross-section of the $W_T^+ W_T^-$ final state. This final state, together with $W_T^+ Z_T$, is the only transverse di-boson process we will consider in Section 3.3 (as a background to the corresponding longitudinal processes). Notice however that there are many other transverse di-boson processes (namely ZZ , $Z\gamma$, $\gamma\gamma$, and $W\gamma$) that can be measured at the VHEL. These processes probe heavy new physics in the EW sector. In particular, as shown in Refs. [98, 99, 102], they are sensitive (together with di-fermions) to minimal Dark Matter in large-multiplets. The large effects of EW radiation might have a strong impact on these studies.

3.3 Sensitivity projections

As described in the Introduction, we target effects from short-distance new physics that grow quadratically with the collision energy, to be probed in $\ell^+ \ell^-$ collisions at the highest available energy $E = E_{\text{cm}}$. In this section we consider the dimension-6 EFT operators listed in Table 3.2, and we estimate the sensitivity of muon colliders of energies $E_{\text{cm}} = 3, 10, 14$ or 30 TeV to their Wilson coefficients. We assume a baseline integrated luminosity [61]

$$\widehat{\mathcal{L}} = 10 \text{ ab}^{-1} \left(\frac{E_{\text{CM}}}{10 \text{ TeV}} \right)^2. \quad (3.35)$$

Semi-quantitative comments on the impact of a reduced luminosity target are postponed to the Conclusions. We base our projections on statistically-dominated measurements of exclusive and semi-inclusive cross-sections for the processes listed in Table 3.3. In the table, for each process we label with a check mark the operators that produce a quadratically growing-with-energy correction to the SM cross-section.

The target EFT operators are selected to represent generic manifestations, at energies much below the new physics scale, of the BSM scenarios we investigate in Section 3.3.3. These

are Composite Higgs, Composite Top and a minimal Z' model, which we select as concrete examples of new physics in the Higgs, Top and EW-gauge sectors. Among the many operators that emerge in these scenarios, we focused our attention on those that display energy growth in $2 \rightarrow 2$ scattering processes at the muon collider. We will see in Section 3.3.3 that other operators offer a weaker sensitivity to the same BSM scenarios.

The phenomenological analysis of the various processes listed in Table 3.3 is described in Sections 3.3.1 and 3.3.2, focusing respectively on the effects of the “W&Y” and of the “Di-boson” operators of Table 3.2. In an attempt to mimic realistic experimental results, we include reconstruction (and, in some case, mistag) efficiencies at a level that is comparable with the CLIC detector performances, which we extract, whenever possible, from Refs. [133, 134]. Table 3.3 displays surprisingly low efficiencies for certain processes (e.g., $t\bar{t}$), entailing a considerable degradation of the measurement uncertainty. In Sections 3.3.1 and 3.3.2 we also present our results for the sensitivity of muon colliders to the corresponding set of operators, with the main aim of outlining the impact of the EW radiation effects on the analysis. The operators in the last class, dubbed “3rd family” in Table 3.2, are not discussed explicitly but the sensitivity projection results are reported in Appendix B.3. The relevant final states, $t\bar{t}$, $b\bar{b}$ and $t\bar{b}$ are discussed in Section 3.3.1.

3.3.1 W&Y operators

The first two operators we consider are those associated with the W and Y parameters of LEP EW precision tests [136], namely O_{2W} and O_{2B} defined as in Table 3.2. These operators arise in the so-called *universal* scenarios [136, 137], that is new physics that couples dominantly to the bosonic sector of the SM. Employing O_{2W} and O_{2B} is convenient in the low-energy context of the LEP experiment, however for our purpose it is better to trade them for the current-current operators O'_{2W} and O'_{2B} (see again Table 3.2), using the SM equations of motion. In doing so, we neglect the contribution to the O'_W and O'_B operators, which are expected to have no impact on the sensitivity. In what follows we parameterize the O'_{2W} and O'_{2B} operator coefficients

$$G'_{2W} = -\frac{g^2 W}{2m_W^2}, \quad G'_{2B} = -\frac{g'^2 Y}{2m_W^2}, \quad (3.36)$$

in terms of the dimensionless parameters W and Y.

The relevant scattering processes, listed in Table 3.3, are the production of two energetic fermions in the central region of the detector. Specifically, as explained at the end of Section 3.2, we have in mind the two hard particles whose invariant mass is higher than around 85% of the total collider E_{cm} , and a scattering angle $\theta_* \in [30^\circ, 150^\circ]$. We assume perfect detector sensitivity to massive W and Z bosons of arbitrary low 3-momentum, enabling the measurement of exclusive scattering cross-sections where the emission of massive vectors (and of photons with hardness above m_W^2) is vetoed. The exclusive cross-section measurements are combined with the semi-inclusive cross-sections, where the emission of an arbitrary number (including zero) of massive vectors or hard photons is allowed.

	SILH basis	Warsaw-like basis
W&Y	$O_{2W} = (D_\mu W^{\mu\nu,a})^2$ $O_{2B} = (\partial_\mu B^{\mu\nu})^2$	$O'_{2W} = J_L^{a,\mu} J_{L,\mu}^a \quad J_L^{a,\mu} = \frac{1}{2} \sum_f \bar{f} \gamma^\mu \sigma^a f$ $O'_{2B} = J_Y^\mu J_{Y,\mu} \quad J_Y^\mu = \sum_f Y_f \bar{f} \gamma^\mu f$
Di-boson	$O_W = \frac{ig}{2} (H^\dagger \sigma^a \overleftrightarrow{D}_\mu H) D^\nu W_{\mu\nu}^a$ $O_B = \frac{ig'}{2} (H^\dagger \overleftrightarrow{D}_\mu H) \partial^\nu B_{\mu\nu}$	$O'_W = \frac{g^2}{4} (H^\dagger i \overleftrightarrow{D}_\mu \sigma^a H) (\bar{L}_L \gamma^\mu \sigma^a L_L)$ $O'_B = -\frac{g'^2}{4} (H^\dagger i \overleftrightarrow{D}_\mu H) (\bar{L}_L \gamma^\mu L_L)$ $-\frac{g'^2}{2} (H^\dagger i \overleftrightarrow{D}_\mu H) (\bar{L}_R \gamma^\mu L_R)$
3 rd family	$O_{qD}^{(3)} = (\bar{q} \gamma^\mu \sigma^a q) (D^\nu W_{\mu\nu}^a)$ $O_{qD}^{(1)} = (\bar{q} \gamma^\mu q) (\partial^\nu B_{\mu\nu})$ $O_{tD} = (\bar{t} \gamma^\mu t) (\partial^\nu B_{\mu\nu})$	$O'_{qD}^{(3)} = (\bar{q} \gamma^\mu \sigma^a q) J_{L,\mu}^a$ $O'_{qD}^{(1)} = (\bar{q} \gamma^\mu \sigma^a q) J_{Y,\mu}$ $O'_{tD} = (\bar{t} \gamma^\mu \sigma^a t) J_{Y,\mu}$

Table 3.2: The operators under consideration in their “SILH” [41] form and, after using the equations of motion, expressed as a linear combination of Warsaw [135] operators. Y_f is the hypercharge of the fermionic field f . In the operators involving the 3rd family the fields t and q denote respectively the right-handed and left-handed top quark.

Process	N (Ex)	N (S-I)	Eff.	O'_{2W}	O'_{2B}	O'_W	O'_B	$O'_{qD}^{(3)}$	$O'_{qD}^{(1)}$	O'_{uD}
$e^+ e^-$	6794	9088	100%	✓	✓					
$e \nu_e$	—	2305	100%	✓	✓					
$\mu^+ \mu^-$	206402	254388	100%	✓	✓					
$\mu \nu_\mu$	—	93010	100%	✓	✓					
$\tau^+ \tau^-$	6794	9088	25%	✓	✓					
$\tau \nu_\tau$	—	2305	50%	✓	✓					
$j j$ (Nt)	19205	25725	100%	✓	✓					
$j j$ (Ch)	—	5653	100%	✓	✓					
$c \bar{c}$	9656	12775	25%	✓	✓					
$c j$	—	5653	50%	✓	✓					
$b \bar{b}$	4573	6273	64%	✓	✓			✓	✓	
$t \bar{t}$	9771	11891	5%	✓	✓			✓	✓	✓
$b t$	—	5713	57%	✓	✓			✓	✓	✓
$Z_0 h$	680	858	26%			✓	✓			
$W_0^+ W_0^-$	1200	1456	44%			✓	✓			
$W_T^+ W_T^-$	2775	5027	44%							
$W^\pm h$	—	506	19%			✓	✓			
$W_0^\pm Z_0$	—	399	23%			✓	✓			
$W_T^\pm Z_T$	—	2345	23%							

Table 3.3: The exclusive and semi-inclusive processes employed for the sensitivity projection. The operators that give a growing-with-energy contribution to each operator are labeled with a check mark. The expected number of events (before efficiencies) is for $E_{\text{cm}} = 10$ TeV with the integrated luminosity (3.35).

For each inclusive and semi-inclusive final state, we employ cross-section measurements in 10 equally-spaced bins of $\cos\theta_*$ in the range $[-\sqrt{3}/2, \sqrt{3}/2]$. In processes (e.g., jj , or $b\bar{b}$) where the two final states are effectively indistinguishable, $\cos\theta_*$ is defined to be positive and 5 bins are employed. We assume cross-section measurements with purely statistical uncertainties, which we estimate based on the number of events that are expected in the SM.

Of course in order to combine the exclusive and semi-inclusive cross-sections for the same (neutral) hard final state we must take into account that the exclusive events are also counted in the measurement of the semi-inclusive cross-section. It is thus convenient to consider a cross-section *with radiation*, defined as the difference between the semi-inclusive and the exclusive cross-sections

$$\sigma_{\text{rad}} \equiv \sigma_{\text{si}} - \sigma_{\text{ex}}. \quad (3.37)$$

The measurement of σ_{rad} can be combined with the one of σ_{ex} since they are statistically independent. For charged hard final states there is instead only one type of cross-section, which necessarily involves EW radiation emission by charge conservation. We will refer to the charged cross-section as “semi-inclusive” or “with radiation” interchangeably.

We now discuss the di-fermion processes individually.

- **e^+e^- , $\mu^+\mu^-$ and $\tau^+\tau^-$:** We assume 100% reconstruction efficiency for muon and electrons, and an efficiency of 50% [133] for each τ lepton. Notice that the cross-section for muons is around 30 times larger than for the other leptons. This is mostly due to the t -channel enhancement of the elastic $\mu^+\mu^-$ scattering.
- **$c\bar{c}$ and $b\bar{b}$:** We assume 50% and 80% efficiency for tagging respectively charm and bottom quark jets [133]. We ignore the mis-tag of light jets, as well as c/b misidentification. No information on the charge of the tagged quark is employed.
- **jj :** We consider the production of two light quarks u , d or s , which we suppose to be reconstructed as jets with 100% efficiency. In Table 3.3 we report separately the production of a neutral (Nt) and of a charged (Ch) quark/anti-quark pair, but the two processes are collected into a single 2-jets final state. We also include the contribution from mistagged b and c quarks.
- **$t\bar{t}$:** Based on Ref.s [134, 138], we estimate as 5% the total efficiency for the reconstruction of the $t\bar{t}$ pair. This (somewhat low) efficiency estimate only includes the semi-leptonic $t\bar{t}$ final states, in which the charge of the tagged top quarks can be measured.
- **tb and cj :** We use 50% and 80% tag efficiency for the charm and the bottom, respectively, and $\sqrt{0.05} = 20\%$ efficiency for the top. The charge of the top quark is assumed to be reconstructed.
- **$e\nu_e$, $\mu\nu_\mu$ and $\tau\nu_\tau$:** The efficiency is 100% for muons and electrons, and 50% for the τ . It should be noted that, because of the invisible neutrino, the hard scattering region of this final state can not be selected with a cut on the invariant mass of the two particles. The selection will have instead to be performed on the energy and the transverse momentum

of the observed charged lepton.

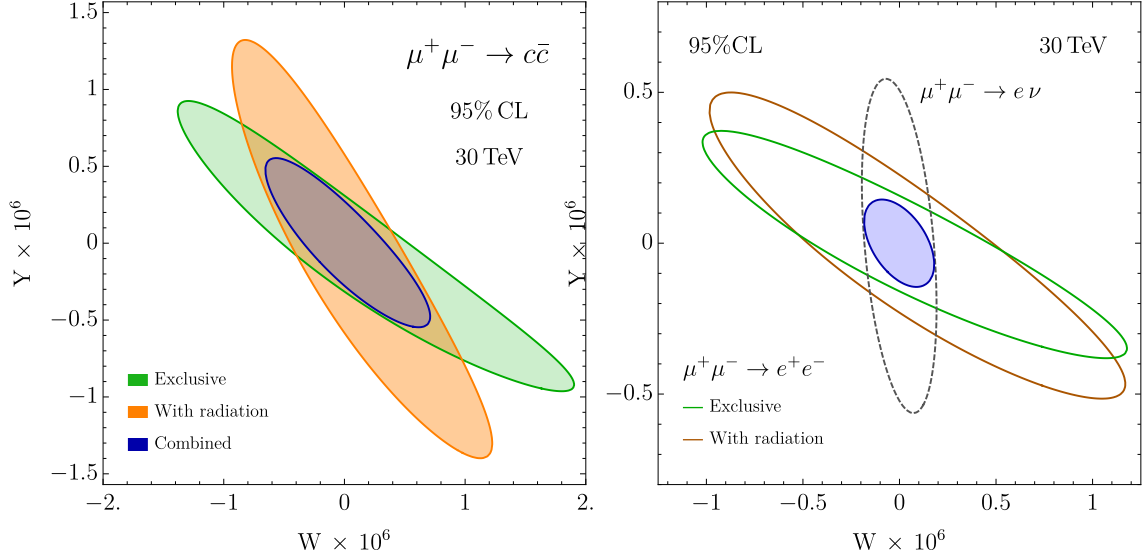


Figure 3.5: 95% CL sensitivities to the W and Y parameters of the 30 TeV muon collider. Exclusive and “with radiation” (i.e., semi-inclusive minus exclusive) cross-section measurements of the $c\bar{c}$ process are considered in the left panel. The right panel shows the impact of e^+e^- (exclusive and “with radiation”) and $e\nu$ (that only exists at the semi-inclusive level) final states.

The different dependence on W and Y of the neutral- and charged-current Born amplitudes entails (see Section 3.2.2) a different dependence on these parameters of the exclusive and semi-inclusive cross-sections. The statistical combination of the two types of cross-sections can thus increase the sensitivity, as illustrated in Figure 3.5. The left panel displays the 95% CL sensitivity of $c\bar{c}$ production to W and Y , comparing the impact of the exclusive cross-section (in green) to that (in orange) of the cross-section with radiation. The two measurements probe different regions of the W and Y parameter space, and a significant sensitivity gain is observed in their combination (in blue). The green and blue lines on the right panel of Figure 3.5 display a similar complementarity pattern for the e^+e^- final state. There also appears an even stronger complementarity with the measurement of the $e\nu$ cross-section, reported as a gray dashed line. The emergence of the $e\nu$ process, as well as the other charged final states in Table 3.3, is entirely due to EW radiation. Nevertheless its (semi-inclusive) cross-section is large, because EW radiation is indeed a prominent phenomenon at $E_{\text{cm}} \simeq 10$ TeV. Furthermore the cross-section displays a peculiar dependence on new physics, producing a sensitivity contour that is different from that of the e^+e^- measurements. The statistical combination of the three measurements (in blue) improves the sensitivity significantly.

The final results of our analysis including all channels are summarized in Figure 3.6 and in Table 3.4. The figure displays the sensitivity contours of exclusive measurements as dotted lines, and the combined impact of charged and of neutral “with radiation” cross-sections, in

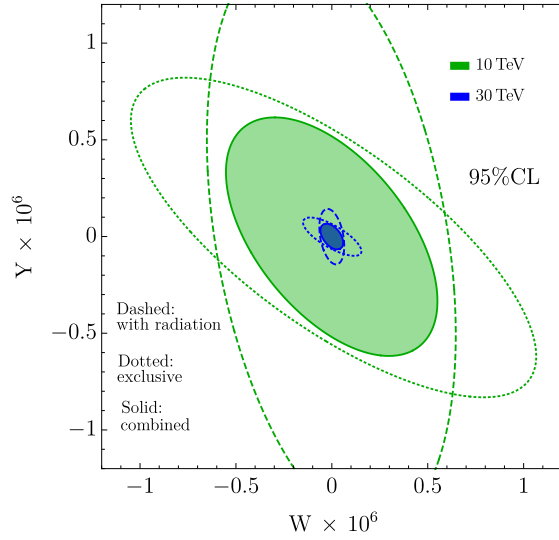


Figure 3.6: 95% CL sensitivities to the W and Y at the 10 and 30 TeV muon collider.

dashed. The combination of all measurements is also shown. The table reports the results for 3, 10, 14 and 30 TeV, comparing the sensitivity of exclusive cross-sections alone with the total combination.

At the High-Luminosity LHC (HL-LHC), it will be possible to probe the W and Y parameters at the level of $4 \cdot 10^{-5}$ and $8 \cdot 10^{-5}$, respectively, at 95% CL [132, 10, 139]. Table 3.4 shows that the 3 TeV muon collider would improve by one order of magnitude or more, and the sensitivity improves quadratically with the muon collider energy. Among the other future collider projects [140], CLIC at 3 TeV has the best sensitivity, of around $4 \cdot 10^{-6}$ for both parameters [133]. This is of course comparable with the 3 TeV muon collider sensitivity, and a factor 10 worst than that of the muon collider at 10 TeV. The comparison with FCC-hh projections is even more favorable to the muon collider.

3.3.2 Diboson operators

The setup for this analysis is similar to that of Ref. [67]. Namely we consider the SILH operators O_W and O_B , we convert them into the current-current interactions O'_W and O'_B as in Table 3.2, and we study their effect on the production of high-energy vector bosons and Higgs. Notice that, by the equivalence theorem, O'_W and O'_B only significantly affect the production of longitudinally polarized vector bosons. We are therefore here studying the production of high-energy longitudinally vector bosons and Higgs, with the production of transversely polarized vector bosons playing merely the role of background. Since the effects are quadratically enhanced by the energy, such high-energy di-boson processes are by far the best probe of these operators at the muon collider [67].

We thus consider, among those in Table 3.3, the following final states

	Exclusive-only [95% CL]			Combined [95% CL]		
	$W \times 10^7$	$Y \times 10^7$	$\rho_{W,Y}$	$W \times 10^7$	$Y \times 10^7$	$\rho_{W,Y}$
3 TeV	[-53, 53]	[-48, 48]	-0.72	[-41, 41]	[-46, 46]	-0.60
10 TeV	[-5.71, 5.71]	[-4.47, 4.47]	-0.74	[-3.71, 3.71]	[-4.16, 4.16]	-0.54
14 TeV	[-3.11, 3.11]	[-2.31, 2.31]	-0.74	[-1.90, 1.90]	[-2.13, 2.13]	-0.52
30 TeV	[-0.80, 0.80]	[-0.52, 0.52]	-0.75	[-0.42, 0.42]	[-0.47, 0.47]	-0.48

Table 3.4: Single-operator 95% CL reach and correlation for the W&Y parameters at different muon collider energies including only exclusive cross-sections and combining all measurements. Since the likelihood is dominated by the linear terms in the new physics parameters, the single parameter reach plus the correlation characterizes our results completely.

- **Zh** : Following Ref. [67], we consider an efficiency of 26% for tagging the two hard and central final state particles, with a selection that reduces the background to a manageable level. Notice that this final state is dominated by the longitudinal helicity channel $Z_0 h$.
- **$W^+ W^-$** : Again like in [67], we assume a 44% efficiency for the detection of the two W bosons in the semi-leptonic decay channel, where the charge of the W 's can be reconstructed. Transverse WW production plays here the role of background.
- **Wh** : We consider an efficiency of 19%, having in mind the leptonic W decay, and $h \rightarrow b\bar{b}$. Like for Zh , there is no relevant background from transverse production.
- **WZ** : We apply an efficiency of 23%, which corresponds to the leptonic W and the hadronic Z decay. The background from transverse WZ production is considerable, and is taken into account.

In our analysis we do not consider the possibility of employing the decay angles of the bosons to extract information on their polarization. Therefore the transverse di-bosons processes $W_T^+ W_T^-$ and $W_T Z_T$ are effectively irreducible backgrounds to the corresponding longitudinal processes, and the scattering angle θ_* is the only discriminating variable. An increased lower cut on θ_* benefits the sensitivity, as it suppresses the t -channel enhancement of the transverse background processes. After optimization we find, like in Ref. [67], that a good signal sensitivity is obtained by the measurement of fiducial WW and WZ cross-sections in the range

$$\theta_* \in [67^\circ, 150^\circ]. \quad (3.38)$$

The possibility of binning θ_* has been considered, but found not to improve the sensitivity. Our analysis will thus be only based on the measurement of the fiducial WW and WZ cross-sections in the above region, and of the Zh and Wh cross-sections for $\theta_* \in [30^\circ, 150^\circ]$. As in the previous section, both exclusive and semi-inclusive cross-sections will be employed for the neutral processes WW and Zh , plus the semi-inclusive charged cross-sections for Wh and WZ .

Single Operator (Exclusive-only) [95% CL]				
		$C_B[10^{-4} \text{ TeV}^{-2}]$		$C_W[10^{-4} \text{ TeV}^{-2}]$
	Linear	Quadratic	Linear	Quadratic
3 TeV	[-170, 170]	[-189, 157]	[-77.4, 77.4]	[-81, 74.4]
10 TeV	[-15.3, 15.3]	[-17, 14.2]	[-8.18, 8.18]	[-8.62, 7.82]
14 TeV	[-7.86, 7.86]	[-8.69, 7.25]	[-4.40, 4.40]	[-4.65, 4.20]
30 TeV	[-1.73, 1.73]	[-1.92, 1.6]	[-1.1, 1.1]	[-1.16, 1.04]

Single Operator (Combined) [95% CL]				
		$C_B[10^{-4} \text{ TeV}^{-2}]$		$C_W[10^{-4} \text{ TeV}^{-2}]$
	Linear	Quadratic	Linear	Quadratic
3 TeV	[-153, 153]	[-169, 142]	[-65.8, 65.8]	[-68.2, 63.6]
10 TeV	[-12.8, 12.8]	[-13.9, 11.9]	[-6.14, 6.14]	[-6.37, 5.93]
14 TeV	[-6.40, 6.40]	[-6.95, 5.99]	[-3.17, 3.17]	[-3.29, 3.06]
30 TeV	[-1.34, 1.34]	[-1.44, 1.25]	[-0.71, 0.71]	[-0.737, 0.686]

Marginalized (Exclusive-only) [95% CL]				
		$C_B[10^{-4} \text{ TeV}^{-2}]$		$C_W[10^{-4} \text{ TeV}^{-2}]$
	Linear	Quadratic	Linear	Quadratic
3 TeV	[-478, 478]	[-352, 596]	[-217, 217]	[-583, 125]
10 TeV	[-53.2, 53.2]	[-35.2, 50]	[-28.4, 28.4]	[-53.5, 14.2]
14 TeV	[-29.4, 29.4]	[-18.6, 25]	[-16.5, 16.5]	[-27.5, 7.82]
30 TeV	[-7.98, 7.98]	[-4.45, 5.19]	[-5.04, 5.04]	[-6.16, 2.05]

Marginalized (Combined) [95% CL]				
		$C_B[10^{-4} \text{ TeV}^{-2}]$		$C_W[10^{-4} \text{ TeV}^{-2}]$
	Linear	Quadratic	Linear	Quadratic
3 TeV	[-442, 442]	[-341, 535]	[-189, 189]	[-426, 115]
10 TeV	[-44, 44]	[-33.4, 43.4]	[-21.1, 21.1]	[-35.1, 12.3]
14 TeV	[-23.1, 23.1]	[-17.6, 21.6]	[-11.4, 11.4]	[-17.6, 6.6]
30 TeV	[-5.24, 5.24]	[-4.12, 4.43]	[-2.79, 2.79]	[-3.70, 1.62]

Table 3.5: Single operator and marginalized 95% reach on C_B and C_W , at different muon collider energies. The sensitivity of exclusive cross-section measurements alone is shown separately from the combination of all the measurements. The significant degradation of the marginalized bounds relative to the single-operators ones, and the strong sensitivity to the quadratic terms at the marginalized level, is due to the approximately flat direction displayed in Figure 3.7

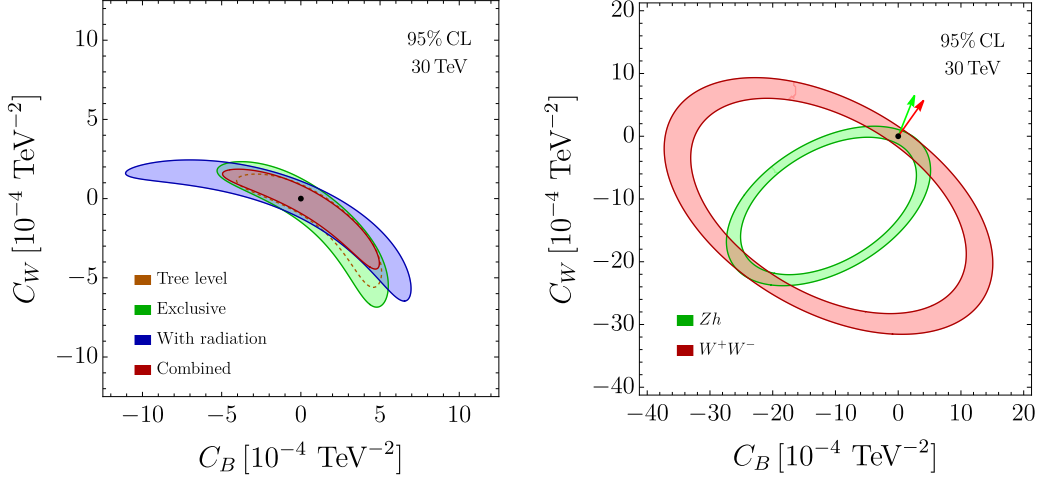


Figure 3.7: Left: 95% sensitivity contours in the (C_B, C_W) plane at the 30 TeV muon collider. A second allowed region, not shown in the figure, can be eliminated by other measurements [67]. Right: Zh and WW likelihood contours at tree-level. Notice that the ellipses for WW and Zh are tangent in two points, one being the SM, the other being the point where the amplitudes have the same magnitude as in the SM but opposite sign.

The results of our analysis are reported in Table 3.5 and on the left panel of Figure 3.7, in terms of the dimensionful coefficients (C_B and C_W) of the O'_B and O'_W operators of Table 3.2. Our findings are quantitatively similar to the ones of Ref. [67]. We can thus refer to that article for the (very favorable) assessment of the muon collider sensitivity to C_B and C_W in comparison with current knowledge and with other future colliders. We devote the rest of this section to discuss the approximate flat direction of the likelihood in the (C_B, C_W) plane, which we observe in Figure 3.7 (left panel).

The flat direction entails a strong degradation of the marginalized sensitivity, as in Table 3.5. Furthermore this degradation brings the marginalized C_B and C_W limits to large values, in a region where the likelihood is considerably affected by the contributions to the cross-sections of the terms that are quadratic in the new physics parameters. In theories like Composite Higgs where $C_{B,W} \sim 1/m_*^2$, this fact implies that the marginalized limits correspond to a new physics scale m_* not much above the collider energy. In fact, looking at Table 3.5 we notice that the 30 TeV C_B reach corresponds to $m_* = 43$ TeV. Thus, if new physics happened to sit along the flat direction in Figure 3.7, diboson processes would fail to extend the muon collider sensitivity well above the direct mass-reach. We do not have reasons to expect new physics to lie in that direction. Actually in certain Composite Higgs models one expects it to lie in the nearly orthogonal direction $C_B = C_W$ [67]. However the presence of the flat direction is an obstruction to the broad exploration of new physics and to the characterization of a putative discovery. It is thus worth explaining its origin and discussing strategies to eliminate it.

The origin of the flat direction in the tree-level sensitivity contour (showed dashed, on the left panel of Figure 3.7) is readily understood analytically, by considering the gradients “ ∇ ”

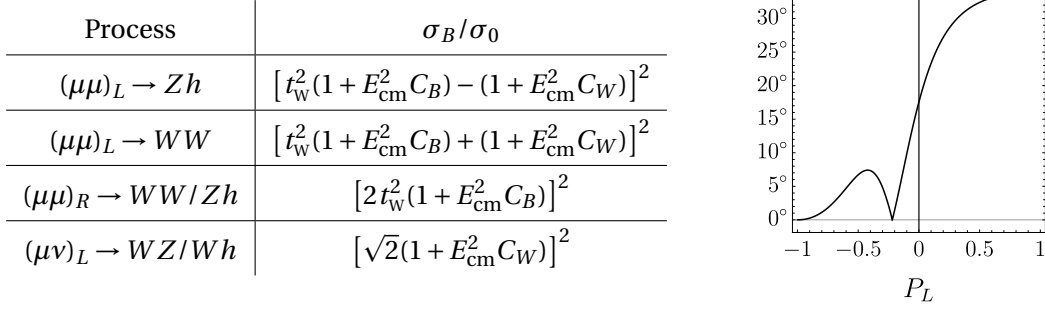


Figure 3.8: Left: Born-level cross-sections, with t_W the tangent of the Weinberg angle, normalized to a common σ_0 (whose expression is irrelevant). Right: the angle between the Zh and WW cross-section gradients as a function of the beam polarization fraction.

of the Born-level cross-sections in the (C_B, C_W) plane, around the SM point $(0, 0)$. Using the results for WW and Zh shown in Figure 3.8 and rescaling the gradients by the common factor $2E_{\text{cm}}^2\sigma_0$ one finds

$$\nabla_L^{Zh} = (1 - t_W^2)\{-t_W^2, +1\}, \quad \nabla_L^{WW} = (1 + t_W^2)\{+t_W^2, +1\}, \quad \nabla_R^{Zh} = \nabla_R^{WW} = 4t_W^4\{1, 0\}, \quad (3.39)$$

where sup- and sub-scripts refer respectively to the final states and to the chirality of the incoming fermions. Notice that the Zh and WW gradients for right-handed initial states are perfectly aligned, so that this contribution to the cross sections has a flat direction (orthogonal to the gradient). The degeneracy can only be lifted by the left-handed contribution to the cross sections. However, given the small value of $t_W^2 \simeq 0.3$, the gradients ∇_L^{Zh} and ∇_L^{WW} also form a relatively small angle, $\sim 30^\circ$. They are thus not very effective in lifting the flat direction when considering the total $(L + R)$ contribution to the WW and Zh cross-section. Indeed, the angle between $\nabla_L^{Zh} + \nabla_R^{Zh}$ and $\nabla_L^{WW} + \nabla_R^{WW}$ is in the end only $\sim 17^\circ$ and thus the flat directions of the two cross-section measurements essentially coincide, as the right panel of Figure 3.7 shows. The combined likelihood is consequently also flat, in the same direction.

As evident in eq. (3.39), the L -gradients form a large angle with the R -gradient. Therefore, if one could use polarized beams, the degeneracy would be eliminated by measuring the contribution of each chirality. Considering a polarization fraction $-P_L$ for the muon, and $+P_L$ for the anti-muon beam, the cross-section gradients read (we indicate by ∇_R the identical ∇_R^{Zh} , ∇_R^{WW})

$$\nabla_{P_L}^{Zh} = \frac{(1 + P_L)^2}{4} \nabla_L^{Zh} + \frac{(1 - P_L)^2}{4} \nabla_R, \quad \nabla_{P_L}^{WW} = \frac{(1 + P_L)^2}{4} \nabla_L^{WW} + \frac{(1 - P_L)^2}{4} \nabla_R. \quad (3.40)$$

The angle between the two gradients steeply increases for positive P_L , as indicated by the plot in the right panel of Figure 3.8. Correspondingly, even a modest amount of polarization has a considerable impact on the sensitivity. The left panel of Figure 3.9 displays our sensitivity projections in a scheme where the VHEL integrated luminosity is equally split between positive

and negative $P_L = \pm 30\%$. The likelihood contour (in green) corresponding to $P_L = +30\%$ is significantly smaller than that (in blue) for $P_L = -30\%$, owing to the lifting of the flat direction achieved for positive P_L . On the other hand, the measurements at $P_L = -30\%$ probe a direction complementary to that probed at $P_L = +30\%$. The combination of the two measurements thus benefits the sensitivity. The impact of beam polarization was emphasized already in Ref. [67]. Here we confirm that result, using more accurate predictions and including the entire set of exclusive and semi-inclusive cross-section measurements previously described.

Up to this point, we discussed the flat direction in the un-polarized likelihood (left panel of Figure 3.7) by employing the tree-level cross-sections. When considering also EW radiation, the predictions are significantly affected, but the flat direction is not fully eliminated. For the exclusive Zh and WW cross-sections this is easily understood, since virtual radiation suppresses the L -processes more than the R ones, owing to the larger Sudakov for incoming left-handed muons. The exclusive Zh and WW cross-sections gradients are thus even more aligned than the gradients of the corresponding tree-level predictions. The semi-inclusive cross-sections for Zh and WW production are also quite aligned, among them and with the exclusive cross-sections. This was expected because the partial cancellation between real and virtual logarithms make semi-inclusive cross-sections not vastly different from the tree-level ones.

On the contrary, the measurement of the charged processes WZ and Wh could have been expected to eliminate or mitigate the flat direction, because they are strongly sensitive to the Born cross-section of the charged scattering $(\mu\nu)_L \rightarrow WZ/Wh$ (see the left panel of Figure 3.8). The associated gradient

$$\nabla_L^{\text{ch}} = 2\{0, +1\}, \quad (3.41)$$

points in a different direction than ∇_L^{Zh} , ∇_L^{WW} and ∇_R . Therefore the gradient of $\sigma^{Wh/WZ}$ could in principle point in a direction that is completely different from that of the (nearly parallel) gradients of the Zh and WW cross-sections. However, by expanding at the first order in $\mathcal{L} = g^2/16\pi^2 \log^2(E_{\text{cm}}^2/m_W^2)$, the unpolarized (longitudinal) WZ and Wh cross-sections are approximately equal and read

$$\sigma^{WZ} \simeq \sigma^{Wh} \simeq \frac{1}{4} \mathcal{L} \cdot (\sigma_B^{Zh} + \sigma_B^{WW} + \sigma_B^{\text{ch}}), \quad (3.42)$$

where σ_B^{ch} is the charged Born cross-section reported on the left panel of Figure 3.8 (times 1/4, from the polarization average) and $\sigma_B^{Zh, WW}$ are the Born cross-sections of the neutral processes. Therefore, the charged cross-section gradient ∇_L^{ch} must compete with the (nearly parallel) gradients of Zh and WW , and its size happens to be insufficient to produce a large misalignment angle between the $\sigma^{Wh/WZ}$ and $\sigma^{Zh/WW}$ gradients.

The situation would be improved, if we could tailor an observable in which the σ_B^{Zh} and σ_B^{WW} contributions in eq. (3.42) are eliminated or reduced. Notice, for that purpose, that the Zh and WW terms in eq. (3.42) can be interpreted as due to one hard $\mu\mu$ neutral-current scattering, followed by the radiation of one charged W boson from the final legs of the hard process.

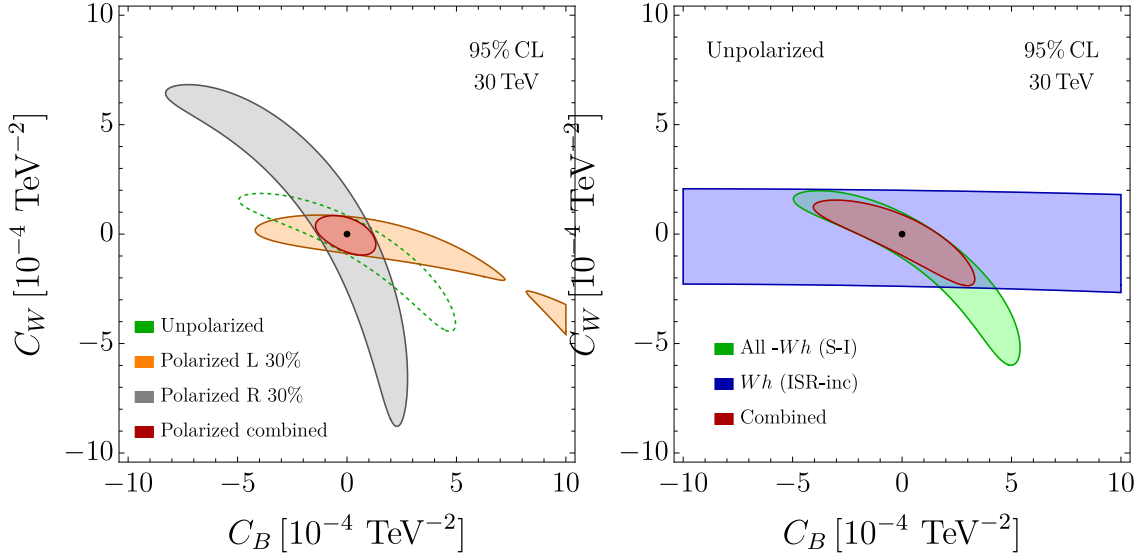


Figure 3.9: Left Panel: 95% CL contours in for $P_L = \pm 30\%$ beam polarization. Right Panel: the impact of the ISR-inclusive Wh cross-section measurement.

The W is thus preferentially collinear to the final states. The σ_B^{ch} term comes instead from the radiation of a W from the initial state, collinear to the beam axis, followed by a hard $\mu\nu$ scattering.¹⁵ This suggests to consider alternative Wh and WZ cross-sections that exclude final state radiation (FSR) while being inclusive on initial state radiation (ISR). FSR consists of soft radiation collinear to the hard particles in the final state, which is precisely the source of the first two terms in eq. (3.42). Excluding FSR, the resulting “ISR-inclusive” cross-sections are expected to be roughly

$$\sigma_{\text{ISR-inc}}^{WZ/Wh} \simeq \frac{1}{4} \mathcal{L} \sigma_B^{\text{ch}}. \quad (3.43)$$

This observable should thus be mostly sensitive to C_W , and its measurement should produce a nearly horizontal band in the (C_B, C_W) plane, thus eliminating the flat direction.

Unfortunately we are unable to produce resummed predictions for the ISR-inclusive cross-sections with the IREE methodology. We can however illustrate the impact of such measurements using tree-level MadGraph [46] predictions with the SMEFT@NLO model [141], focusing in particular on the Wh channel. Specifically, we simulate the process

$$\mu^+ \mu^- \rightarrow W^+ W^- h, \quad (3.44)$$

at $E_{\text{cm}} = 30 \text{ TeV}$, with the following selection cuts. First, we identify as “hard” the W boson that forms, together with the Higgs, the pair with the highest invariant mass. Secondly we ask this mass to be above $0.85 \cdot E_{\text{cm}} = 25.5 \text{ TeV}$ and the hard W and h to be within the central region

¹⁵This interpretation would straightforwardly correspond to Feynman diagrams in a physical gauge, where DLs are associated to emissions from individual legs. We already remarked that in covariant gauges instead they arise from the interference between emission from strictly different legs.

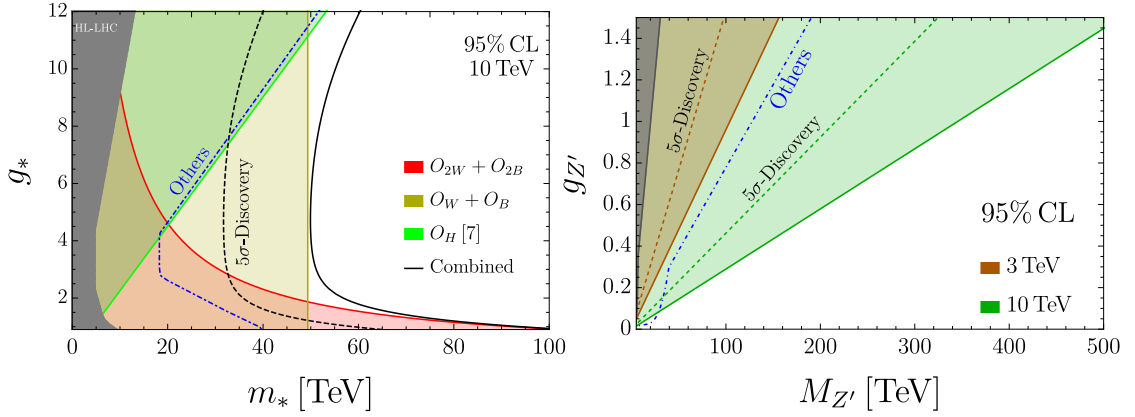


Figure 3.10: Left Panel: 95% exclusion reach on the Composite Higgs coupling-mass parameter space. The reach for O_H is taken from [67]. Right Panel: sensitivity projections for a Y -universal Z' model. The gray band and the blue dash-dot line represent respectively the region probed by the HL-LHC program and the sensitivity projections for all other future collider projects [142].

$\theta_* \in [30^\circ, 150^\circ]$. These selections enforce the occurrence of a hard scattering, and correspond to our definition of a “semi-inclusive” process. We further restrict to the “ISR-inclusive” region by asking the other (“soft”) W to be parallel to the beam, in a cone of 20° . Since the emission of at least one soft W is required for Wh production, the latter cut effectively corresponds to a veto on central EW radiation.¹⁶

The above estimate of the ISR-inclusive cross-section produces the blue band on the right panel of Figure 3.9. As expected, the band is nearly horizontal. In the figure we also display, in green, the 95% CL contour of the likelihood including all the measurements discussed in the present section, apart from the measurement of the semi-inclusive Wh cross-section which is correlated with the ISR-inclusive measurement. The combination of the two contours, shown in red, strongly mitigates the flat direction issue. Notice however that our tree-level estimate of the ISR-inclusive cross-section could be subject to large errors, and resummed predictions should be employed for a conclusive assessment of the sensitivity gain.

3.3.3 BSM sensitivity

Composite Higgs

As a first concrete scenario of new physics we consider Composite Higgs [143, 144, 145]. In this scenario, the Higgs is a composite Pseudo-Nambu-Goldstone boson emerging from some strong dynamics at a scale m_* . In principle the underlying dynamics could arise from gauge interactions, like in QCD. However the only concrete realistic constructions, accounting for the

¹⁶The attempt made in Ref. [67] to exploit the $WW h$ process did not impose the crucial angular cut that defines the ISR-inclusive region.

origin of both the fermion masses and the scale m_* itself, have been obtained in the context of warped compactifications. In these constructions, compositeness occurs in a holographic sense. Within the Composite Higgs scenario, the size of the Wilson coefficients in the resulting low energy EFT, can be estimated, under simple but robust dynamical assumptions, in terms of the mass scale m_* and overall coupling strength g_* of the underlying strong dynamics [41]. Furthermore, simple considerations suggest $g_* \lesssim 4\pi$, while the existence of $\mathcal{O}(1)$ couplings within the SM implies $g_* \gtrsim 1$. The Composite Higgs power-counting rules predict the Wilson coefficients of the operators in the left column of Table 3.2 to scale as

$$C_{2W} = -\frac{c_{2W}}{2} \frac{g^2}{g_*^2 m_*^2}, \quad C_{2B} = -\frac{c_{2B}}{2} \frac{g'^2}{g_*^2 m_*^2}, \quad C_W = c_W \frac{1}{m_*^2}, \quad C_B = c_B \frac{1}{m_*^2}, \quad (3.45)$$

where the dimensionless coefficients c_{2W} , c_{2B} , c_W , c_B are expected to be of order 1. Even though it does not affect the processes studied in this chapter, an important role is also played by

$$O_H = \partial_\mu (H^\dagger H) \partial^\mu (H^\dagger H), \quad C_H = \frac{c_H}{2} \frac{g_*^2}{m_*^2}. \quad (3.46)$$

In our sensitivity projections we will report the corresponding bounds, as obtained in [67] by studying the process $\mu^+ \mu^- \rightarrow hh\nu\nu$ at tree level. Other probes of C_H at the muon collider, from Higgs coupling measurements, are superior or competitive at the lower energy muon colliders [67], but they are not considered in the sensitivity plots.

Using the above scalings, and setting all the c coefficients to 1, we can translate the bounds of Section 3.3 for a 10 TeV muon collider into sensitivity estimates in the plane (m_*, g_*) , as in Figure 3.10. In the same plot we also report the HL-LHC sensitivity projections, the envelope of the 95% CL sensitivity contours of all the future collider projects that have been considered for the 2020 update of the European Strategy for Particle Physics [142]. The advantage of the muon collider is evident. Results at muon colliders with different energies, with an integrated luminosity scaling as in eq. (3.35), are reported in Appendix B.4.

Composite top.

The results for purely bosonic operators we just discussed apply robustly to basically all composite Higgs scenarios. Operators involving fermions are more sensitive to the assumptions on the flavor dynamics, but one convenient option is offered by the mechanism of partial compositeness [147], under which the elementary fermions mix linearly with heavy partners in the strong sector. Due to its large Yukawa coupling, the top quark is expected to have a large mixing with its partners and therefore precise measurements involving the third family represent an appealing opportunity to probe new physics.

At the muon collider the most relevant effects are expected in $t\bar{t}$ and $b\bar{b}$ production.¹⁷ Indeed the dimension-6 operators in the last block of Table 3.2 gives rise to contributions that grow

¹⁷See [148] for a similar analysis for CLIC.

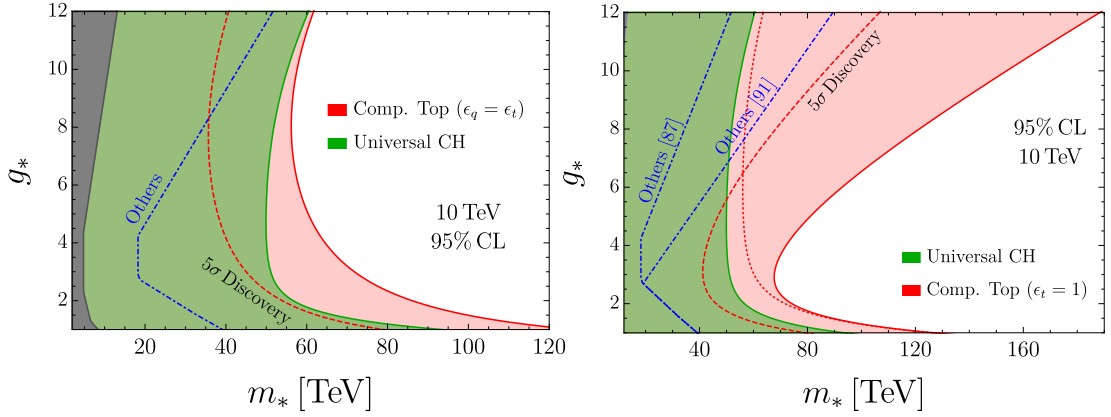


Figure 3.11: 95% exclusion reach for the two partial compositeness scenarios under consideration. The green shapes represent the combined bound from the flavor universal measurements, while the red contours also includes the di-top and di-bottom constraints. The projected sensitivity of other future collider projects and the gray band of HL-LHC are taken from Ref. [142]. The right panel ($\epsilon_t = 1$) also includes the stronger CLIC sensitivity estimated in Ref. [146].

with E_{cm} and which can be exploited at the large energy of the muon collider. In a model-independent approach one can parametrize the “amount of compositeness” of respectively the 3rd quark family left-handed doublet and right-handed up-type singlet by ϵ_q and ϵ_t . These quantities range from 0 to 1. Given the universal coupling strength g_* of the strong sector the resulting top Yukawa coupling scale as [41]

$$y_t \sim \epsilon_q \epsilon_t g_* . \quad (3.47)$$

The relevant Wilson coefficients are then expected to scale as (see [138] for a short review)

$$C_{qD}^{(3)} = c_{qD}^{(3)} \frac{g \epsilon_q^2}{m_*^2}, \quad C_{qD}^{(1)} = c_{qD}^{(1)} \frac{g' \epsilon_q^2}{m_*^2}, \quad C_{tD} = c_{tD} \frac{g' \epsilon_t^2}{m_*^2}, \quad (3.48)$$

where the c_i are, as usual, expected to be order 1 coefficient. For concreteness we focus on two benchmark scenarios, where we fix ϵ_t and ϵ_q and leave g_* and m_* free. In the first scenario, the right-handed top quark is assumed to be fully composite, corresponding to $\epsilon_t = 1$ and $\epsilon_q = y_t/g_*$. In the second, the two top chiralities are assumed equally composite, that is $\epsilon_q = \epsilon_t = \sqrt{y_t/g_*}$.

Notice that the contribution of the operator

$$O_{tt} \equiv \frac{1}{2} (\bar{t}_R \gamma^\mu t_R) (\bar{t}_R \gamma_\mu t_R), \quad (3.49)$$

to the Wilson coefficients of the O_{tD} , through Renormalization Group (RG) evolution, is not negligible in the scenario of total right-handed top quark compositeness [146]. Using the

power-counting estimate

$$C_{tt} = \epsilon_t^4 \frac{g_*^2}{m_*^2} c_{tt}, \quad (3.50)$$

we obtain a correction [146] to the C_{tD} coefficient at a scale $\mu = E_{\text{cm}}$

$$C_{tD}(\mu) = C_{tD}(m_*) + C_{tt}(m_*) \frac{32}{9} \frac{g'}{16\pi^2} \log\left(\frac{m_*^2}{\mu^2}\right) = \epsilon_t^2 \frac{g'}{m_*^2} \left(c_{tD} + c_{tt} \frac{32}{9} \frac{\epsilon_t^2 g_*^2}{16\pi^2} \log\left(\frac{m_*^2}{\mu^2}\right) \right). \quad (3.51)$$

This correction is sizable if $\epsilon_t \sim 1$, especially for large g_* , because the sensitivity of the muon collider extends to a scale m_* that is significantly larger than E_{cm} .

There are in principle three more operators O_{Ht} , $O_{Hq}^{(1)}$ and $O_{Hq}^{(3)}$ (defined as in the “Warsaw” basis [135]) that mix significantly with those in eq. (3.48) through RG evolution. However, their effects can only be important in the case where $\epsilon_q \sim 1$, which we do not contemplate in our analysis. We will therefore neglect the RG effects of the latter three operators and consider only those of O_{tt} .

Our results are summarized in Figure 3.11, where we report the projected exclusion reach in the g_* and m_* plane in the two scenarios under consideration for $E_{\text{cm}} = 10$ TeV. Additional results can be found in Appendix B.4. Starting from the scenario of equal compositeness (left panel) we notice that at g_* the additional hypothesis of top compositeness extends the muon collider potential to probe the scale of Higgs compositeness m_* . The effect is even stronger for fully composite t_R (right panel), which shows that di-top measurements can cover up to $m_* \sim 150$ TeV for $g_* \gtrsim 8$. We should point out, however, that this result depends on the exact $\mathcal{O}(1)$ value of the c_{tt} , c_{tD} coefficients in eq. (3.51). This dependence is illustrated in Figure B.4, where we set $c_{tD} = 1$ and we vary the value of c_{tt} .

Finally we remark that a detailed analysis of the composite Higgs scenario with partial compositeness would require specific hypotheses on the flavor dynamics and a detailed inspection of the flavor observables. Depending on those hypotheses, principally the maximality or minimality of the underlying flavor symmetry, the resulting flavor constraints on the new physics scale m_* can vary dramatically. While a comprehensive analysis clearly exceeds the purposes of this work, a perspective can be gained by considering available studies. As shown in Ref. [149], under the strongest assumptions, that is for a symmetry structure offering the best protection from unwanted effects, flavour and CP observables could start exploring the range $m_* = \mathcal{O}(10)$ TeV in the next decade or so, given the availability of better measurements and assuming better theoretical predictions. This is roughly the same range explored by a 3 TeV muon collider. Moreover the $m_* \sim 50$ TeV reach of a 10 TeV muon collider vastly surpasses any conceivable improvement of flavour constraints, and competes with the more stringent flavour bounds obtained by making more generic assumptions on the flavor dynamics. Notice also that the present lepton flavor universality anomalies in B-decays, at least the seemingly more prominent ones in neutral currents, suggest a new physics scale in the ~ 30 TeV range, which could be explored both directly and indirectly by the muon collider.

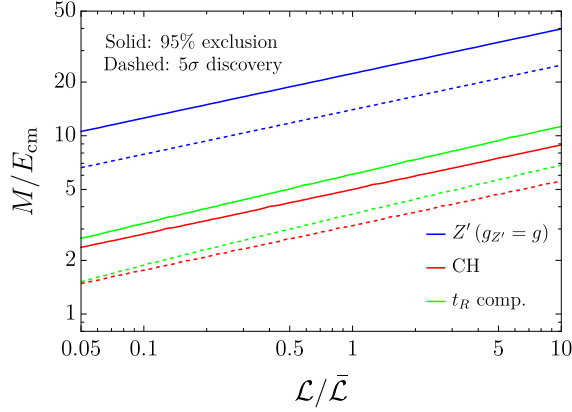


Figure 3.12: Reach on the new physics scale at 2σ (continuous) and at 5σ (dashed), relative to the collider energy, as a function of the integrated luminosity normalized to eq. (3.35). The red lines are for Universal manifestations of Higgs compositeness, while the green ones include the effect of Top compositeness in the $\epsilon_t = 1, \epsilon_q = y_t/g_*$ scenario. The blue lines are for the Y -universal Z' for a fixed coupling $g_{Z'} = g$.

Y -Universal Z' model

The Y -universal Z' model represents a simple extension of the SM, featuring an additional heavy gauge boson, of mass $M_{Z'}$, on top of the SM particles.¹⁸ In this benchmark scenario the new vector charges are aligned with the SM hypercharge with coupling $g_{Z'}$. Requiring the width of the Z' to not exceed $0.3M_{Z'}$ sets the perturbative limit on the coupling to be $g_{Z'} \lesssim 1.5$. At energies below $M_{Z'}$, integrating out the Z' only generates the O'_{2B} operator of Table 3.2. The Wilson coefficient of the operators corresponds, by eq. (3.36), to

$$Y = \left(\frac{g_{Z'} m_W}{g' M_{Z'}} \right)^2. \quad (3.52)$$

The sensitivity projections are reported in the right panel of Figure 3.10. The orange and green regions are the ones probed by muon colliders at 3 and 10 TeV energy, respectively. The gray band represents the expected exclusion reach from HL-LHC, while the blue line indicates the combined sensitivity from other future collider projects (dominantly FCC, and the 3 TeV stage of CLIC). The 3 TeV muon collider sensitivity is obviously similar to the one of CLIC. A 10 TeV machine would greatly improve this result probing up to 500 TeV for large (but still perturbative) coupling. The dashed lines represent the discovery reach, showing that already at 3 TeV there are vast opportunities for indirect discovery, well above the region that the HL-LHC might exclude. Results at higher muon collider energies are reported in Figure B.2.

¹⁸See [142] for details. Concrete BSM scenarios featuring additional Z' s can be found, for instance, in Ref. [150].

3.4 Conclusions and outlook

We have studied the interplay between two classes of novel phenomena, which can be observed at future lepton colliders with very large center of mass energy. The first class consists of hard scattering processes induced by new physics at around 100 TeV. The second class consists of the long-distance effects of EW radiation. Both phenomena play a relevant role at lepton colliders with $E_{\text{cm}} \sim 10$ TeV. In particular they are relevant at muon colliders, which are the main target of the present work.

The interplay manifests itself in two ways. The first one is simply that EW radiation effects on the SM predictions are large (see Sections 3.2.2 and 3.2.3) and require to be included and resummed with high accuracy in order to isolate the putative BSM contribution to the measurements. EW radiation thus plays for muon colliders a similar role as QCD radiation for the LHC, with the difference (discussed in the Introduction) that its effects can not be mitigated by the choice of suitable (inclusive) observables. Therefore they are actually even more important for muon colliders than QCD is for the LHC.

The second and possibly more interesting aspect of the interplay is given by the influence on the pattern of EW radiation operated by the presence of new physics in the hard scattering amplitude. This makes observables that require or that exclude the presence of radiation display a different dependence on the new physics parameters, and the sensitivity profits from their combined measurements as illustrated in Sections 3.3.1 and 3.3.2, and in Appendix B.3.

Our sensitivity projections rely on putative measurements of exclusive and semi-inclusive cross-sections. Both classes of processes are characterized by the occurrence of a hard scattering, with two particles in the final state carrying almost all the available energy. The emission of additional EW bosons and hard photons is vetoed in the exclusive case and allowed in the semi-inclusive one. We computed the resummed semi-inclusive cross-sections in double logarithm (DL) approximation by extending the IREE methodology [105], as described in Section 3.2.1. The exclusive cross-sections were computed at DL, but also including single logarithms at 1-loop, which we found to be sizable.

The studies performed in this paper should be improved and extended in many directions. Better predictions will be definitely needed in order to approach the percent-level accuracy target that is needed to fully exploit the statistical precision potential of a muon collider. Moreover, given the magnitude of the radiation effects we observed, it is possible that more accurate predictions will considerably affect some of our sensitivity projections. A first step in that direction, which we leave for future work, is the inclusion of single logarithms at fixed leading order in the semi-inclusive predictions. That could be achieved by combining one loop virtual logarithms with a factorized treatment of real emission. That calculation, would possibly also help clarify the connection between soft-collinear effects (studied in this paper) and the PDF/Fragmentation Function treatment of EW radiation. It is not unconceivable that the same approach could be extended at the two loop order. In parallel, the impact of resummation beyond DL should be assessed. The SCET methodology currently offers the

most promising approaches.

Another priority is to investigate further classes of cross-sections, sensitive to different EW radiation patterns. Our results indicate that investigation should be done on the basis of the structure of short distance new physics. At the end of Section 3.3.2 we took one step in that direction, showing that the approximately flat direction in the (C_B, C_W) likelihood contours is mitigated, in the absence of polarized beams, by considering an “ISR-inclusive” cross-section. This third cross-section type is inclusive on radiation collinear to the beam, but vetoes centrally emitted radiation. We could not compute the ISR-inclusive cross-section at DL with our IREE methodology and limited ourself to a tree-level estimate. A straightforward direction for progress would be to perform that calculation and verify if and how it impacts our findings.

The definition and study of cross-sections should be also based on experimental considerations. We already pointed out that exclusive cross-sections are problematic in that regard. Indeed imposing the radiation veto requires experimental sensitivity to EW radiation that is emitted in all directions, including the forward and backward regions along the beam line. The angular coverage of the muon collider detector is still to be quantified, however we expect that it will be insufficient for the measurement of exclusive cross-sections.

In view of the above, it is important to emphasize that our sensitivity projections have been verified to not change radically when exclusive cross-section measurements are not available. This conclusion is not in contradiction with (and cannot be inferred from) our sensitivity plots, where (see e.g. Figure 3.6) we observe a strong complementarity between “exclusive” observables and observables “with radiation”. Indeed for neutral processes the latter observables consists of the difference between the semi-inclusive and the exclusive cross-sections. Therefore the impact of eliminating the exclusive measurements can not be estimated by suppressing the “exclusive” cross section measurement in the computation of the likelihood. The proper estimate is obtained by employing the semi-inclusive neutral processes without subtraction, combined with the charged measurements, produces a combined reach that is not much inferior to the one that exploits the exclusive measurements. In essence, the main sensitivity gain due to radiation stems from the emergence of the charged processes and from their complementarity with the neutral ones. The complementarity between neutral measurements with different degrees of radiation inclusiveness (e.g., exclusive versus semi-inclusive) plays instead a less relevant role in our results. This same qualitative behavior can be observed in the comparison between the neutral and charged lepton production processes on the right panel of Figure 3.5.

On the other hand, the complementarity between exclusive and semi-inclusive measurements exists and can benefit the sensitivity as we illustrated on the left panel of Figure 3.5. It plays a marginal role in the combined fit to the limited number of EFT operators we considered in this paper. It could however be relevant in a more global exploration and characterization of putative new physics. One way to recover sensitivity, if exclusive measurements were indeed

unavailable, could be to exploit the ISR-inclusive cross-sections. These are easier to measure because they do not require sensitivity to radiation in the forward and backward regions. This aspect should be investigated.

Our phenomenological results strengthen and extend previous estimates of the muon collider sensitivity to heavy new physics. We have considered a variety of BSM scenarios with new physics coupled to the SM with electroweak strength. We found that a $E_{\text{cm}} = 10$ TeV muon collider can probe up to a scale ranging from 50 to 200 TeV. The reach improves linearly with E_{cm} . These figures are significantly above the potential (direct and indirect) sensitivity of other future collider projects, and above the direct sensitivity reach of the muon collider itself, which is obviously bounded by the collider energy.

The indirect sensitivity to scales that are well above the direct reach is a great addition to the physics case of a muon collider, whose relevance would not be diminished, but on the contrary augmented, by the occurrence of a direct discovery. Indeed, direct hints for new particles observed at the muon collider will turn into a full-fledged discovery of new physics only after unveiling the underlying theoretical description of their dynamics. The possibility of probing this dynamics well above the particle's mass will play a decisive role in this context. Furthermore, the direct manifestation of new physics might be hard to detect. Perhaps, indirect probes will provide the first hint of its existence, to be eventually confirmed by targeted direct searches. Finally, indirect searches for BSM phenomena based on precise measurements guarantee a sound output of the project. The connection with the phenomenon of EW radiation, which is interesting per se, adds scientific value to the program.

Before concluding, we discuss the impact of the integrated luminosity on our results. We employed the baseline luminosity in eq. (3.35), which corresponds to 90 ab^{-1} for a 30 TeV muon collider. Since the possibility of reducing the 90 ab^{-1} target is under discussion, it is worth assessing the impact of a lower integrated luminosity on our conclusions. An important aspect is associated with the actual experimental feasibility of the relevant measurements. While a conclusive assessment will require dedicated studies, the expected number of events in Table 3.3¹⁹ and the corresponding efficiencies show that, with a factor 10 reduction in luminosity, some of the processes we employed would be left with a handful of observed events, possibly preventing the corresponding measurements. If the reduction in luminosity were less extreme, the sensitivity to the scale of new physics would simply deteriorate as the fourth root of the luminosity, as shown in Figure 3.12 for some of the BSM scenarios we studied in Section 3.3.3. The figure displays the exclusion and discovery reach on the new physics scale normalized to the collider energy $E_{\text{cm}} = 30$ TeV. This is the right figure of merit, since the goal is to extend the muon collider sensitivity above the direct reach. The reduction by a factor of $10^{1/4} = 1.8$ due to a factor 10 luminosity reduction partially undermines this goal, especially for what concerns the generic manifestations of the Composite Higgs scenario.

¹⁹The table is for $E_{\text{cm}} = 10$ TeV, however with the scaling in eq. (3.35) the results depend weakly on the energy.

A Formulation of the Quadratic Classifier

A.1 The general Quadratic Classifier

Any quadratic-order real polynomial of $n-1$ variables c_i , $i = 1, \dots, n-1$, with arbitrary constant, linear and quadratic terms, can be written as a quadratic form in the n -dimensional variable

$$v(c) = (1, c_1, \dots, c_{n-1})^T. \quad (\text{A.1})$$

Namely, we write the polynomial as

$$P(c) = v^T(c) A v(c), \quad (\text{A.2})$$

with A a generic n -dimensional real symmetric square matrix.

If $P(c)$ is non-negative for any value of c , it is easy to show that the matrix A must be positive semi-definite. Being real, symmetric and positive semi-definite, it is possible to use the Cholesky decomposition for A , and write it as

$$A = L^T L, \quad (\text{A.3})$$

where L is a upper-triangular (i.e., $L_{ij} = 0$ for $j < i$) real matrix. Therefore the most general positive quadratic order polynomial reads

$$P(c) = v^T(c) L^T L v(c) = \sum_{i=1}^n \left(\sum_{j=1}^n L_{ij} v_j(c) \right)^2 = \sum_{i=1}^n \left(L_{i1} + \sum_{j=2}^n L_{ij} c_{j-1} \right)^2, \quad (\text{A.4})$$

which is manifestly non-negative because it is the sum of square terms. Moreover for $c = 0$, since $L_{i1} = L_{11} \delta_{i1}$, we have $P(0) = L_{11}^2$. The Cholesky decomposition is unique up to sign flips of the rows of L . Rather than resolving this ambiguity, for instance by choosing the diagonal entries of L to be positive, we adopt eq. (A.4) without further constraints as the most general (though redundant) parametrization of $P(c)$.

Appendix A. Formulation of the Quadratic Classifier

The EFT differential cross section is a positive quadratic polynomial in the Wilson Coefficient c_i at each phase-space point x , and it reduces to the SM cross section for $c = 0$. It must therefore take the form

$$d\sigma_0(x; c) = d\sigma_1(x) \sum_{i=1}^n \left[\delta_{i1} + \sum_{j=2}^n \lambda(x)_{ij} c_{j-1} \right]^2, \quad (\text{A.5})$$

with $\lambda(x)$ an upper-triangular matrix of real functions. If only one c parameter is present (i.e., $n = 2$), this reduces to eq. (2.1) with the identifications

$$\lambda(x)_{12} = \alpha(x) \quad \lambda(x)_{22} = \beta(x). \quad (\text{A.6})$$

The Quadratic Classifier that generalizes eq. (2.3) is thus defined as

$$f(x, c) \equiv \frac{1}{1 + \sum_{i=1}^n \left[\delta_{i1} + \sum_{j=2}^n n(x)_{ij} c_{j-1} \right]^2}, \quad (\text{A.7})$$

in terms of an upper-triangular matrix $n(x)$ of real-output Neural Networks.

A.2 Minimization of the parametrized loss

In the Large Sample limit, the loss function in eq. (2.4) becomes

$$L[n(\cdot)] \stackrel{\text{LS}}{=} \sum_{c \in \mathcal{C}} \left\{ \int d\sigma_0(x; c) [f(x, c)]^2 + \int d\sigma_1(x) [1 - f(x, c)]^2 \right\}, \quad (\text{A.8})$$

with the Quadratic Classifier f defined in eq. (A.7). By simple algebraic manipulations, this can be rewritten as

$$L[n(\cdot)] \stackrel{\text{LS}}{=} \sum_{c \in \mathcal{C}} \left\{ \int \frac{d\sigma_1(x) d\sigma_0(x; c)}{d\sigma_1(x) + d\sigma_0(x; c)} + \int [d\sigma_1(x) + d\sigma_0(x; c)] \left[f(x, c) - \frac{1}{1 + r(x, c)} \right]^2 \right\}, \quad (\text{A.9})$$

with $r(x, c) = d\sigma_0(x; c)/d\sigma_1(x)$. The first integral is independent of f and thus it is irrelevant for the minimization of the loss. The second one is the integral of a non-negative function of x which attains its global minimum (i.e., it vanishes) if and only if

$$f(x, c) = f_{\min}(x, c) = \frac{1}{1 + r(x, c)}, \quad \forall c \in \mathcal{C}. \quad (\text{A.10})$$

By using eq. (A.5), and comparing with eq. (A.7), we immediately conclude that the configuration $n(x)_{ij} = \lambda(x)_{ij}$ is a global minimum of the loss and that this minimum is unique provided the set \mathcal{C} contains at least two distinct non-vanishing elements. More precisely, this holds only up to sign ambiguities, associated with those of the Cholesky decomposition. However this is irrelevant because the ambiguity cancels out in f , and in turn it cancels out in the reconstructed distribution ratio $\hat{r}(x, c) = 1/\hat{f}(x, c) - 1$.

We have shown that the Quadratic Classifier reconstructs the distribution ratio exactly (in the Large Sample limit and for infinitely complex Neural Network) at the global minimum of the loss, and that this minimum is unique. Notice however that we could not show that the Large Sample limit loss does not possess additional local minimums, as it is instead readily proven for the standard classifier of Section 1.4.2 by variational calculus.

B Resummation of Radiation at the Muon Collider

B.1 Radiation Integrals

The contribution of virtual radiation to the amplitude variation in eq. (3.10) is proportional to the integral

$$I \equiv \frac{-i}{(2\pi)^4} \int \delta\sigma d^4q \frac{1}{q^2 - m_V^2 + i\epsilon} \frac{(k_i \cdot k_j)}{(q \cdot k_i)(q \cdot k_j)}, \quad (\text{B.1})$$

where we included a mass $m_V \sim m_W$ for the virtual vector in order to verify explicitly that the integral is log-enhanced only in the $\lambda \gg m_W^2$ regime, where the IR cutoff is much above the EW scale. We now proceed (following [151]) to the evaluation of I assuming, for simplicity, exactly massless hard 4-momenta $k_i^2 = k_j^2 = 0$. The integral is Lorentz-invariant, therefore it can only depend on the scalar product $(k_i \cdot k_j)$, that we set to

$$(k_i \cdot k_j) = \frac{1}{2}(k_i + k_j)^2 = \frac{1}{2}E^2, \quad (\text{B.2})$$

in what follows.

The calculation is conveniently performed in Sudakov coordinates [152]. Namely we parametrize the loop momentum q as

$$q = uk_i + vk_j + q_\perp^1 \zeta_1 + q_\perp^2 \zeta_2, \quad (\text{B.3})$$

where $(\zeta_1)^2 = (\zeta_2)^2 = -1$ and $\zeta_{1,2} \cdot k_i = \zeta_{1,2} \cdot k_j = 0$. In these coordinates

$$q^2 = uvE^2 - |q_\perp|^2, \quad (\text{B.4})$$

and the infinitesimal strip $\delta\sigma$ (3.7) that defines the integration region is expressed as

$$|uv|E^2 \in [\lambda, \lambda + \delta\lambda]. \quad (\text{B.5})$$

Appendix B. Resummation of Radiation at the Muon Collider

After the change of variables, the integral reads

$$I = \frac{i}{(2\pi)^4} \int_{\delta\sigma} \frac{dudv}{uv} \int \frac{d^2 q_\perp}{|q_\perp|^2 - uvE^2 + m_V^2 - i\epsilon}. \quad (\text{B.6})$$

The $d^2 q_\perp$ integral must be performed up to an upper cutoff that justifies the usage of the Eikonal approximation formula in eq. (3.8) for the gauge boson vertices. In particular we notice that the actual denominators of the virtual legs in the diagram are not $(k_{i,j} \cdot q)$ as in the Eikonal formula, but rather $(k_{i,j} \cdot q) \pm q^2/2$, with plus or minus for incoming and outgoing particles, respectively. Neglecting $q^2/2$ is justified only if $|q_\perp|^2/2$ is not as large as to compete with the minimum among $|k_i \cdot q| = |u|E^2/2$ and $|k_j \cdot q| = |v|E^2/2$. Therefore the q_\perp integral should be cutoff at

$$|q_\perp|^2 < \Lambda^2, \quad \text{with} \quad \Lambda^2 = E^2 \min[|u|, |v|]. \quad (\text{B.7})$$

Similarly, the term proportional to uv in q^2 should not be large compared to $(k_{i,j} \cdot q)$, therefore the u and v integrals are also bounded, in the region

$$|u| < 1, \quad |v| < 1, \quad (\text{B.8})$$

supplemented by eq. (B.5).

The integration boundaries of u and v are invariant under $u \rightarrow -u$ and under $v \rightarrow -v$ reflections. We can thus perform the integral (B.6) in the first quadrant of the (u, v) plane, provided we duly symmetrize the integrand. Furthermore we notice that first-quadrant integration region (restricted to $u < 1$ and $v < 1$, owing to eq. (B.8)) is conveniently described by the coordinates τ and y , defined by

$$u = \sqrt{\tau} e^y, \quad v = \sqrt{\tau} e^{-y}. \quad (\text{B.9})$$

Indeed in these coordinates the condition (B.5) merely sets the value of τ to

$$\tau = \frac{\lambda}{E^2}, \quad \text{with} \quad \delta\tau = \frac{\delta\lambda}{E^2}, \quad (\text{B.10})$$

while the upper bound (B.7) on the $|q_\perp|$ integral reads

$$\Lambda^2 = E^2 \sqrt{\tau} e^{-|y|} = E \sqrt{\lambda} e^{-|y|}. \quad (\text{B.11})$$

We can thus express the integral as

$$I = \frac{2i}{(2\pi)^3} \frac{\delta\lambda}{\lambda} \int_0^{y_M} dy \int_0^{E\sqrt{\lambda}e^{-y}} \frac{d|q_\perp|^2}{|q_\perp|^2 - \lambda + m_V^2 - i\epsilon} \left[\frac{1}{|q_\perp|^2 - \lambda + m_V^2 - i\epsilon} - \frac{1}{|q_\perp|^2 + \lambda + m_V^2 - i\epsilon} \right]. \quad (\text{B.12})$$

where, since Λ^2 is symmetric under $y \rightarrow -y$, we could take the y integral from 0 to

$$y_M = -\frac{1}{2} \log \tau = \log \frac{E}{\sqrt{\lambda}}, \quad (\text{B.13})$$

and multiply by a factor 2.

The $d|q_\perp|^2$ integral in eq. (B.12) is readily performed. It is convenient to separate the terms that emerge from the upper $|q_\perp|^2$ integration extreme from the one of the lower extreme $|q_\perp|^2 = 0$, obtaining two contributions to I , I_U and I_L . We will readily see that the former contribution is suppressed, therefore

$$I \simeq I_L = \frac{1}{8\pi^2} \frac{\delta\lambda}{\lambda} \log \frac{E^2}{\lambda} \frac{i}{\pi} \log \left[\frac{\lambda + m_V^2 - i\epsilon}{-\lambda + m_V^2 - i\epsilon} \right]. \quad (\text{B.14})$$

The logarithm gives quite different results in the two regimes $\lambda \gg m_W^2 \sim m_V^2$ and $\lambda \ll m_W^2 \sim m_V^2$. In the second one, the argument of the logarithm has positive real part, almost equal to 1 up to m_V^2/λ power-corrections. In the first regime, the argument has negative real part and the log equals $+i\pi$, plus λ/m_V^2 corrections. Namely

$$I^{\lambda \gg m_W^2} = -\frac{1}{8\pi^2} \frac{\delta\lambda}{\lambda} \log \frac{E^2}{\lambda} [1 + \mathcal{O}(m_V^2/\lambda)], \quad I^{\lambda \ll m_W^2} = -\frac{1}{8\pi^2} \frac{\delta\lambda}{\lambda} \log \frac{E^2}{\lambda} \cdot \mathcal{O}(\lambda/m_V^2), \quad (\text{B.15})$$

from which we recover eq. (3.11). More precisely, notice that the integral for each pair of external legs ij is controlled by the specific scale $E^2 = 2(k_i \cdot k_j)$. In eq. (3.11) we set all these scales equal up to corrections that are not log enhanced, but of order one. This in turn corresponds to order $1/\log$ corrections to the evolution kernel and to single-logarithm corrections to the solutions of the IREE.

The contribution to I from the upper $|q_\perp|^2$ integration extreme is suppressed. To see this it is convenient to employ the integration variable $\rho = \exp(y - y_M)$, obtaining

$$I_U = \frac{2i}{(2\pi)^3} \frac{\delta\lambda}{\lambda} \int_{\sqrt{\lambda}/E}^1 \frac{d\rho}{\rho} \frac{1}{\rho} \log \left[\frac{1 - \rho(1 - m_V^2/\lambda)}{1 + \rho(1 + m_V^2/\lambda)} \right], \quad (\text{B.16})$$

where we could drop the $-i\epsilon$ because the argument of the logarithm has positive real part in the entire range of integration. The ρ integral is finite for $\sqrt{\lambda}/E \rightarrow 0$, therefore it does not produce log-enhanced contributions. In particular the integral gives $-\pi^2/4$ for $\lambda \gg m_W^2 \sim m_V^2$ and it is power-suppressed in the opposite regime.

We now compute the contribution of real radiation to the density matrix variation, which we employ in eq. (3.18). The relevant integral that controls the contribution from a real radiation diagram like those in Figure 3.2, reads

$$I_R \equiv -\frac{1}{(2\pi)^3} \int_{\delta\sigma} \frac{d^3q}{2q^0} \frac{(k_i \cdot k_j)}{(k_i \cdot q)(k_j \cdot q)} = -\frac{1}{(2\pi)^3} \int_{\delta\sigma} d^4q \delta(q^2 - m_V^2) \theta(q^0) \frac{(k_i \cdot k_j)}{(k_i \cdot q)(k_j \cdot q)}, \quad (\text{B.17})$$

where we employed the eikonal formula in eq. (3.8), but ignored the “ G ” factors that are included separately in eq. (3.18). Notice the presence of a minus sign, which is due to three factors of -1 . The first minus is due to the fact that applying the eikonal formula to the conjugate amplitude gives the complex conjugate of the generator matrix “ G_i ” of the corresponding leg, while eq. (3.18) is expressed in terms of the generators $G_{i^c} = -G_i^*$ of the conjugate representation. The second minus sign emerges from the sum over the polarizations of the real vector boson, which gives $-\eta_{\mu\nu}$. The third minus is because the contribution to the variation is minus the integral over the strip $\delta\sigma$, since the λ cutoff is a lower bound on the hardness.

In Sudakov coordinates (B.3), and setting $(k_i \cdot k_j) = E^2/2$, the integral becomes

$$I_R = -\frac{1}{(2\pi)^2} \int_{\delta\sigma} \frac{dudv}{uv} \theta(u+v) \int d|q_\perp|^2 \delta(uvE^2 - |q_\perp|^2 - m_V^2). \quad (\text{B.18})$$

The integration extremes of all the variables are like those of the virtual integral, including the condition (B.5) that defines the infinitesimal integration strip $\delta\sigma$. The delta function in eq. (B.18) has support only if $uv > 0$, and the theta function further restricts the integral to the first quadrant of the (u, v) plane. We can thus employ the τ and y coordinates in eq. (B.9) and readily obtain

$$I_R = -\frac{1}{8\pi^2} \frac{\delta\lambda}{\lambda} \log \frac{E^2}{\lambda} \cdot \theta(\lambda - m_V^2). \quad (\text{B.19})$$

Evidently, the theta function condition is not satisfied for $\lambda \ll m_W^2 \sim m_V^2$, therefore the integral vanishes in this regime. In the other regime

$$I_R \stackrel{\lambda \gg m_W^2}{=} -\frac{1}{8\pi^2} \frac{\delta\lambda}{\lambda} \log \frac{E^2}{\lambda}, \quad (\text{B.20})$$

which is equal to the virtual radiation integral, as anticipated in the main text. Notice that the pre-factor of eq. (3.18) contains an additional $1/2$, due to the fact that the real radiation contribution is effectively counted twice in the equation by the two terms proportional to $(G_i^A)(G_{j^c}^A)$ and to $(G_{i^c}^A)(G_j^A)$, which are equal after summing over i and j .

B.2 High-energy EW multiplets

The EW symmetry is effectively unbroken at energies much above the EW scale. Therefore in this regime it is convenient to describe the SM particles in terms of representations of the (linearly-realized) SM group $SU(2)_L \times U(1)_Y$, with generators

$$\mathcal{T}^A = \{\mathcal{T}^a, \mathcal{Y}\}, \quad a = 1, 2, 3. \quad (\text{B.21})$$

The generators act on the single particle states as

$$\mathcal{T}^A |p(k, \alpha)\rangle = |p(k, \beta)\rangle (T_r^A)^\beta_\alpha, \quad (\text{B.22})$$

Bosons	$SU(2)_L$	\mathcal{Y}	Leptons	$SU(2)_L$	\mathcal{Y}
W	\mathbf{t}	0	$(\nu_{\ell,-1/2}, \ell_{-1/2}^-)$	\mathbf{d}	$-1/2$
B	\mathbf{s}	0	$(\bar{\nu}_{\ell,+1/2}, \ell_{+1/2}^+)^t$	$\bar{\mathbf{d}}$	$-1/2$
H	\mathbf{d}	$+1/2$	$\ell_{+1/2}^-$	\mathbf{s}	-1
\bar{H}	$\bar{\mathbf{d}}$	$-1/2$	$\ell_{-1/2}^+$	\mathbf{s}	$+1$

Quarks	$SU(2)_L$	\mathcal{Y}
$(u_{-1/2}, d_{-1/2})^t$	\mathbf{d}	$+1/6$
$(\bar{u}_{+1/2}, \bar{d}_{+1/2})^t$	$\bar{\mathbf{d}}$	$-1/6$
$u_{+1/2}$	\mathbf{s}	$+2/3$
$\bar{u}_{-1/2}$	\mathbf{s}	$-2/3$
$d_{+1/2}$	\mathbf{s}	$-1/3$
$\bar{d}_{-1/2}$	\mathbf{s}	$+1/3$

 Table B.1: $SU(2)_L \times U(1)_Y$ quantum numbers of the SM particles in the high-energy regime.

with generator matrices T_r^A that define the representation “r” of the particle multiplet. The field Φ_r^α that interpolates the multiplet from the vacuum, namely

$$\langle 0 | \Phi_r^\alpha(0) | p(k, \beta) \rangle \propto \delta_\beta^\alpha, \quad (\text{B.23})$$

transforms with the same generator matrix, i.e.

$$[\Phi_r^\alpha(x), \mathcal{T}^A] = (T_r^A)^\alpha_\beta \Phi_r^\alpha(x). \quad (\text{B.24})$$

The $SU(2)_L$ representations of the SM particles and the corresponding $U(1)_Y$ charges are listed in Table B.1.

All the fermionic particles with helicity $-1/2$ transforms as doublets (i.e., $r = \mathbf{d}$), the anti-particles with helicity $+1/2$ transform in the conjugate ($r = \bar{\mathbf{d}}$) of the doublet representation, while all the others are singlets. Obviously this is true only in the high energy limit where the fermions are effectively massless and the helicity corresponds the chirality of the corresponding interpolating fields. The doublet representation matrices are the standard

$$T_{\mathbf{d}}^a = \left\{ \begin{pmatrix} 0 & +1/2 \\ +1/2 & 0 \end{pmatrix}, \begin{pmatrix} 0 & -i/2 \\ +i/2 & 0 \end{pmatrix}, \begin{pmatrix} +1/2 & 0 \\ 0 & -1/2 \end{pmatrix} \right\}, \quad T_{\bar{\mathbf{d}}}^a = -(T_{\mathbf{d}}^a)^* = -(T_{\mathbf{d}}^a)^t. \quad (\text{B.25})$$

The EW multiplets of bosonic particles are less well-known, but equally straightforward to work out employing the standard Goldstone Boson Equivalence Theorem, or better its stronger formulation in Ref.s [43, 153]. The point is that for massive W^\pm and Z vector bosons with 0 (longitudinal) helicity one can employ interpolating fields that are a specific combination of

Appendix B. Resummation of Radiation at the Muon Collider

the regular gauge fields W_μ^\pm and Z_μ and of the Goldstone boson scalar fields π^\pm and π_Z . The longitudinal states are thus a linear combination of the quantum fluctuation modes associated to this two different type of fields, and the scattering amplitudes with external longitudinal states are a linear combination of Feynman diagrams where the external states are represented either as gauge fields or as Goldstone fields. This is a convenient formalism in the high energy limit because the polarization vector associated with gauge external lines vanishes as m_W/E (unlike the regular longitudinal polarization, that diverges as E/m_W), and only the Goldstone diagrams survive. In essence this means that the Goldstones, and not the gauge, are the adequate interpolating fields for the longitudinal particles at high energy. Therefore the EW quantum numbers of the longitudinal particles are the ones of the Goldstones, and not of the gauge fields. The Goldstone bosons, together with the Higgs, form a doublet with $+1/2$ hypercharge, and the corresponding conjugate doublet

$$H = \left(\pi^+, \frac{h + i\pi_Z}{\sqrt{2}} \right)^t, \quad \bar{H} = \left(\pi^-, \frac{h - i\pi_Z}{\sqrt{2}} \right)^t. \quad (\text{B.26})$$

Vector bosons with transverse helicity $T = \pm 1$ are instead well-described by gauge fields even at high energy. Therefore they decompose into a triplet plus a singlet EW multiplet, which are readily obtained by undoing the Weinberg rotation

$$W^a = \left(\frac{W_T^+ + W_T^-}{\sqrt{2}}, i \frac{W_T^+ - W_T^-}{\sqrt{2}}, c_W Z_T + s_W \gamma_T \right)^t, \quad B = -s_W Z_T + c_W \gamma_T. \quad (\text{B.27})$$

Explicitly, the triplet generator matrices are

$$T_t^a = \left\{ \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & i \\ 0 & 0 & 0 \\ -i & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \right\}. \quad (\text{B.28})$$

We now proceed to the evaluation of the K_i exponentials in the explicit formula for the semi-inclusive density matrix (3.22), for external legs in the doublet (or anti-doublet) or in the triplet canonical representations defined by eq. (B.25) and (B.28). The K tensors in eq. (3.20) are

$$[K_d]^{\alpha\bar{\alpha}}_{\beta\bar{\beta}} = [K_{\bar{d}}]^{\alpha\bar{\alpha}}_{\beta\bar{\beta}} = c_d \delta_\beta^\alpha \delta_{\bar{\beta}}^{\bar{\alpha}} + \sum_a (T_d^a)^\alpha_\beta (T_{\bar{d}}^a)^{\bar{\alpha}}_{\bar{\beta}} = \delta_\beta^\alpha \delta_{\bar{\beta}}^{\bar{\alpha}} - \frac{1}{2} \delta^{\alpha\bar{\alpha}} \delta_{\beta\bar{\beta}}, \quad (\text{B.29})$$

$$[K_t]^{\alpha\bar{\alpha}}_{\beta\bar{\beta}} = c_t \delta_\beta^\alpha \delta_{\bar{\beta}}^{\bar{\alpha}} + \sum_a (T_t^a)^\alpha_\beta (T_t^a)^{\bar{\alpha}}_{\bar{\beta}} = 2 \delta_\beta^\alpha \delta_{\bar{\beta}}^{\bar{\alpha}} + \delta_\beta^\alpha \delta_{\bar{\beta}}^{\bar{\alpha}} - \delta^{\alpha\bar{\alpha}} \delta_{\beta\bar{\beta}}. \quad (\text{B.30})$$

The exponential of eq. (B.29) is trivial and we readily obtain

$$\left[\exp(-\mathcal{L} K_d) \right]^{\alpha\bar{\alpha}}_{\beta\bar{\beta}} = e^{-\mathcal{L} \delta_\beta^\alpha \delta_{\bar{\beta}}^{\bar{\alpha}}} + e^{-\mathcal{L}/2} \sinh(\mathcal{L}/2) \delta^{\alpha\bar{\alpha}} \delta_{\beta\bar{\beta}}, \quad (\text{B.31})$$

where we defined

$$\mathcal{L} = \frac{g^2}{16\pi^2} \log^2(E^2/m_W^2). \quad (\text{B.32})$$

The exponential of eq. (B.30) is slightly more involved. First of all notice that the three tensor structures in the right-hand-side of eq. (B.30) commute so we can exponentiate all of them independently. In particular we get

$$\exp\left(-2\mathcal{L}\delta_\beta^\alpha\delta_{\bar\beta}^{\bar\alpha}\right) = e^{-2\mathcal{L}}\delta_\beta^\alpha\delta_{\bar\beta}^{\bar\alpha} \quad (\text{B.33})$$

$$\exp\left(-\mathcal{L}\delta_\beta^\alpha\delta_{\bar\beta}^{\bar\alpha}\right) = \cosh(\mathcal{L})\delta_\beta^\alpha\delta_{\bar\beta}^{\bar\alpha} - \sinh(\mathcal{L})\delta_\beta^\alpha\delta_{\bar\beta}^{\bar\alpha}, \quad (\text{B.34})$$

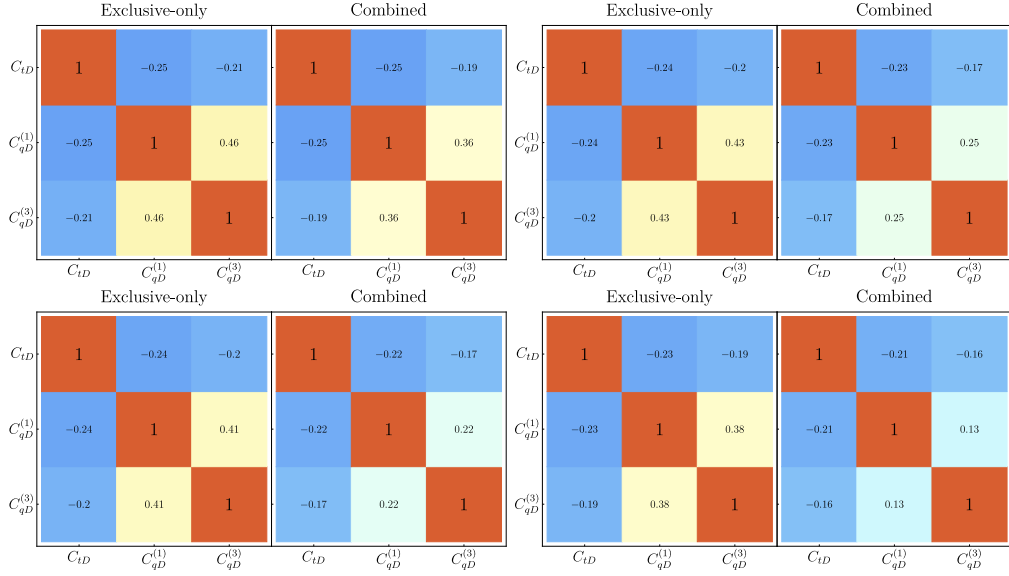
$$\exp\left(\mathcal{L}\delta^{\alpha\bar\alpha}\delta_{\beta\bar\beta}\right) = \delta_\beta^\alpha\delta_{\bar\beta}^{\bar\alpha} + \frac{1}{3}\left(e^{3\mathcal{L}} - 1\right)\delta^{\alpha\bar\alpha}\delta_{\beta\bar\beta}. \quad (\text{B.35})$$

Taking the product of the factors just found, we get

$$\left[\exp\left(-\mathcal{L} K_t\right)\right]_{\beta\bar\beta}^{\alpha\bar\alpha} = e^{-2\mathcal{L}}\left(\cosh(\mathcal{L})\delta_\beta^\alpha\delta_{\bar\beta}^{\bar\alpha} - \sinh(\mathcal{L})\delta_\beta^\alpha\delta_{\bar\beta}^{\bar\alpha}\right) + \frac{2}{3}e^{-\frac{3}{2}\mathcal{L}}\sinh(3\mathcal{L}/2)\delta^{\alpha\bar\alpha}\delta_{\beta\bar\beta}. \quad (\text{B.36})$$

B.3 3rd family operators

The sensitivity to the 3rd family operators in Table 3.2 are summarized in this section. In Figure B.1 we report the two-dimensional contours in the $(C_{tD}, C_{qD}^{(3)})$ and $(C_{qd}^{(1)}, C_{qD}^{(3)})$ planes, with the third operator set to zero. We notice that the “with radiation” cross-section measurements (see the main text) is mostly effective to probe $C_{qD}^{(3)}$ producing a significant sensitivity improvement on the combined bound in this direction. The effect is milder in the orthogonal directions. The likelihood is dominated by the linear term in the new physics parameters so all our result can be expressed in terms of the single operator reaches (at 95% CL) of and the correlations matrices in Table B.2. In the table we report the sensitivity of exclusive cross-section measurements alone, and the combination of all the measurements.



Top left: $E_{\text{cm}} = 3$ TeV; top right: $E_{\text{cm}} = 10$ TeV; bottom left: $E_{\text{cm}} = 14$ TeV; bottom right: $E_{\text{cm}} = 30$ TeV

B.4 Summary plots

In this appendix we collect additional results skipped in the main text. In particular in Figure B.2 we report the sensitivity projections for the Y -universal Z' model, in the $(g_{Z'}, M_{Z'})$ plane for the different collider energies. In Figure B.3 we collect the sensitivity projections for the composite Higgs model in the (m_*, g_*) plane for $E_{\text{cm}} = 3, 14, 30$ TeV. Projections including composite top measurements can be found in Figure B.5. Finally, Figure B.4 shows the dependence of the bound on the value of the c_{tt} coefficients, as explained in the main text.

	Exclusive-only [95% CL]			Combined [95% CL]		
	C_{tD}	$C_{qD}^{(1)}$	$C_{qD}^{(3)}$	C_{tD}	$C_{qD}^{(1)}$	$C_{qD}^{(3)}$
3 TeV	[-24.4, 24.4]	[-9.47, 9.47]	[-6.68, 6.68]	[-23.1, 23.1]	[-8.59, 8.59]	[-5.45, 5.45]
10 TeV	[-2.24, 2.24]	[-0.97, 0.97]	[-0.71, 0.71]	[-2.04, 2.04]	[-0.81, 0.81]	[-0.49, 0.49]
14 TeV	[-1.15, 1.15]	[-0.52, 0.52]	[-0.38, 0.38]	[-1.03, 1.03]	[-0.42, 0.42]	[-0.25, 0.25]
30 TeV	[-0.26, 0.26]	[-0.13, 0.13]	[-0.10, 0.10]	[-0.22, 0.22]	[-0.10, 0.10]	[-0.05, 0.05]

Table B.2: Single-operator 95% CL reach and correlation matrices for the Wilson coefficient $C_{qD}^{(3)}$, $C_{qD}^{(1)}$ and C_{tD} of the operators of Table 3.2 at different collider energies. All results include exclusive cross-sections or combined measurements. The Wilson coefficient are expressed in 10^{-4} TeV^{-2} . Since the likelihood is dominated by the linear terms in the new physics parameters, the single parameter reach and the correlation characterize our results completely.

Appendix B. Resummation of Radiation at the Muon Collider

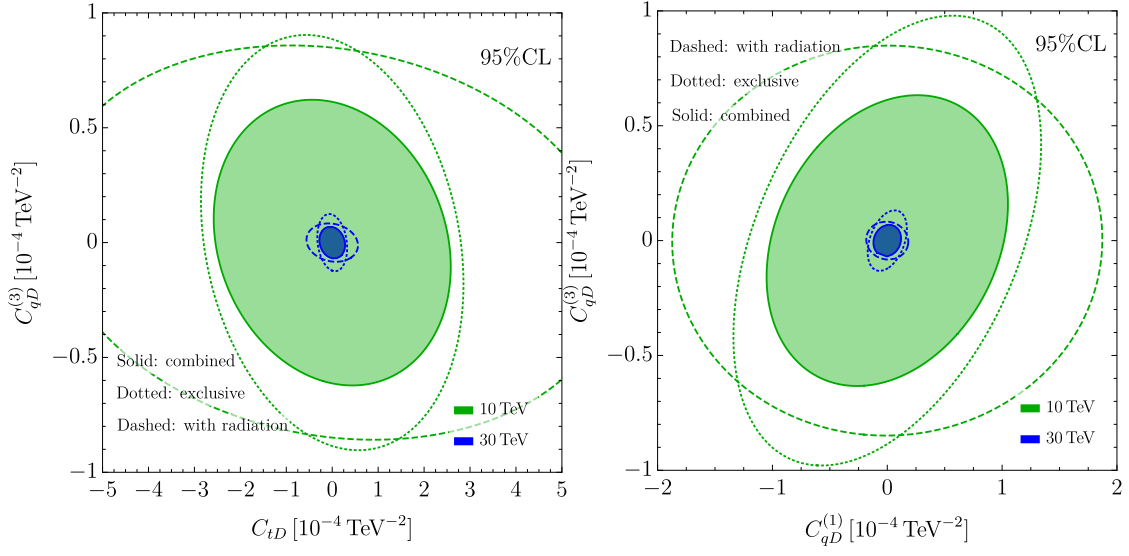


Figure B.1: 95% CL contours in the $(C_{tD}, C_{qD}^{(3)})$ (left) and $(C_{qD}^{(1)}, C_{qD}^{(3)})$ (right) planes at the 10 and 30 TeV muon colliders.

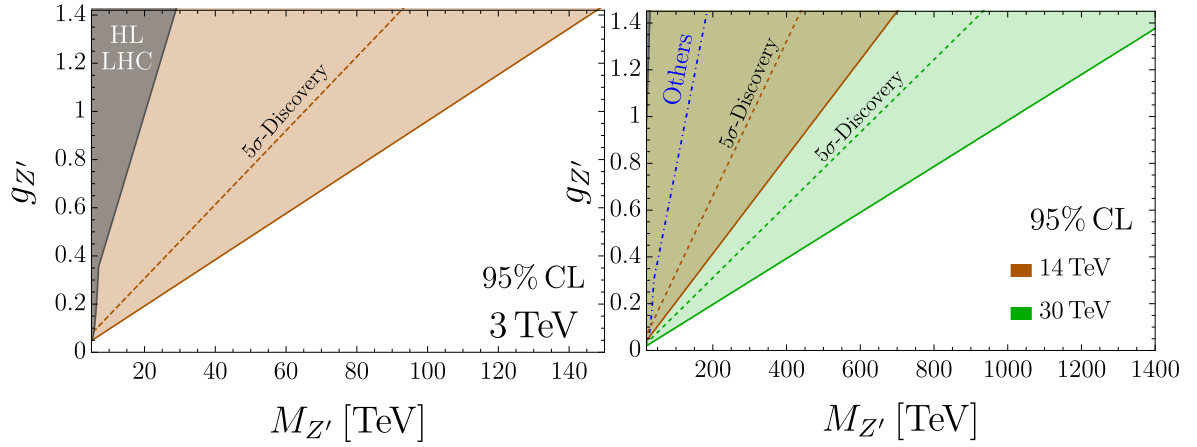


Figure B.2: The same as the right panel of Figure 3.10 for various collider energies.

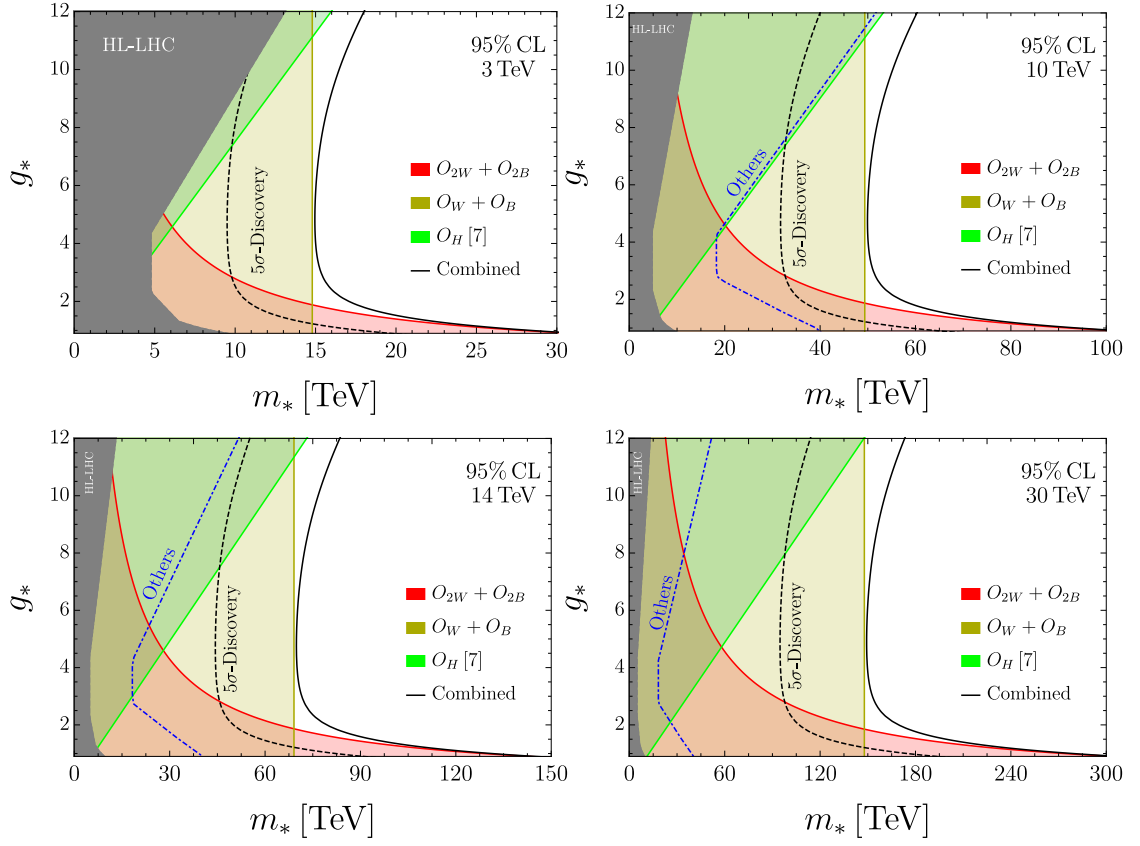
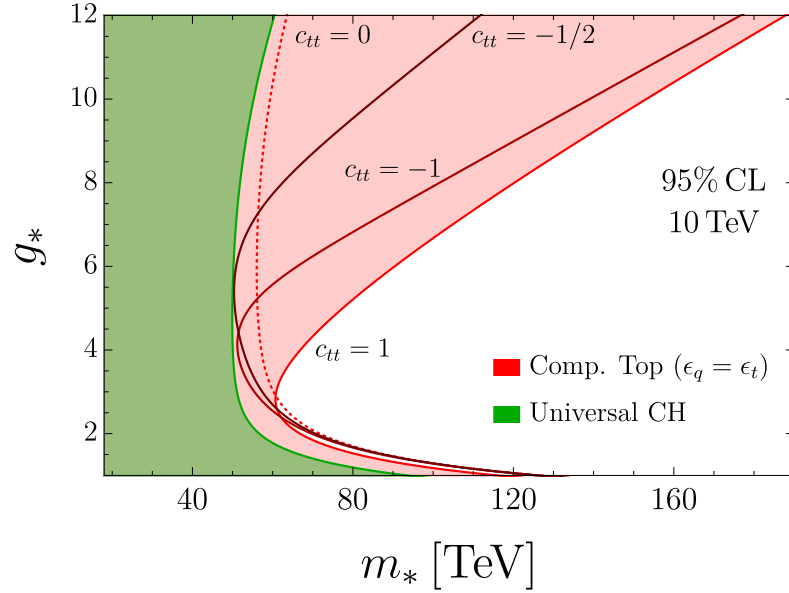


Figure B.3: The same as the left panel of Figure 3.10 for various collider energies.


 Figure B.4: The same as the right panel of Figure 3.11 for different values of c_{tt} to show the model dependence of the result.

Appendix B. Resummation of Radiation at the Muon Collider

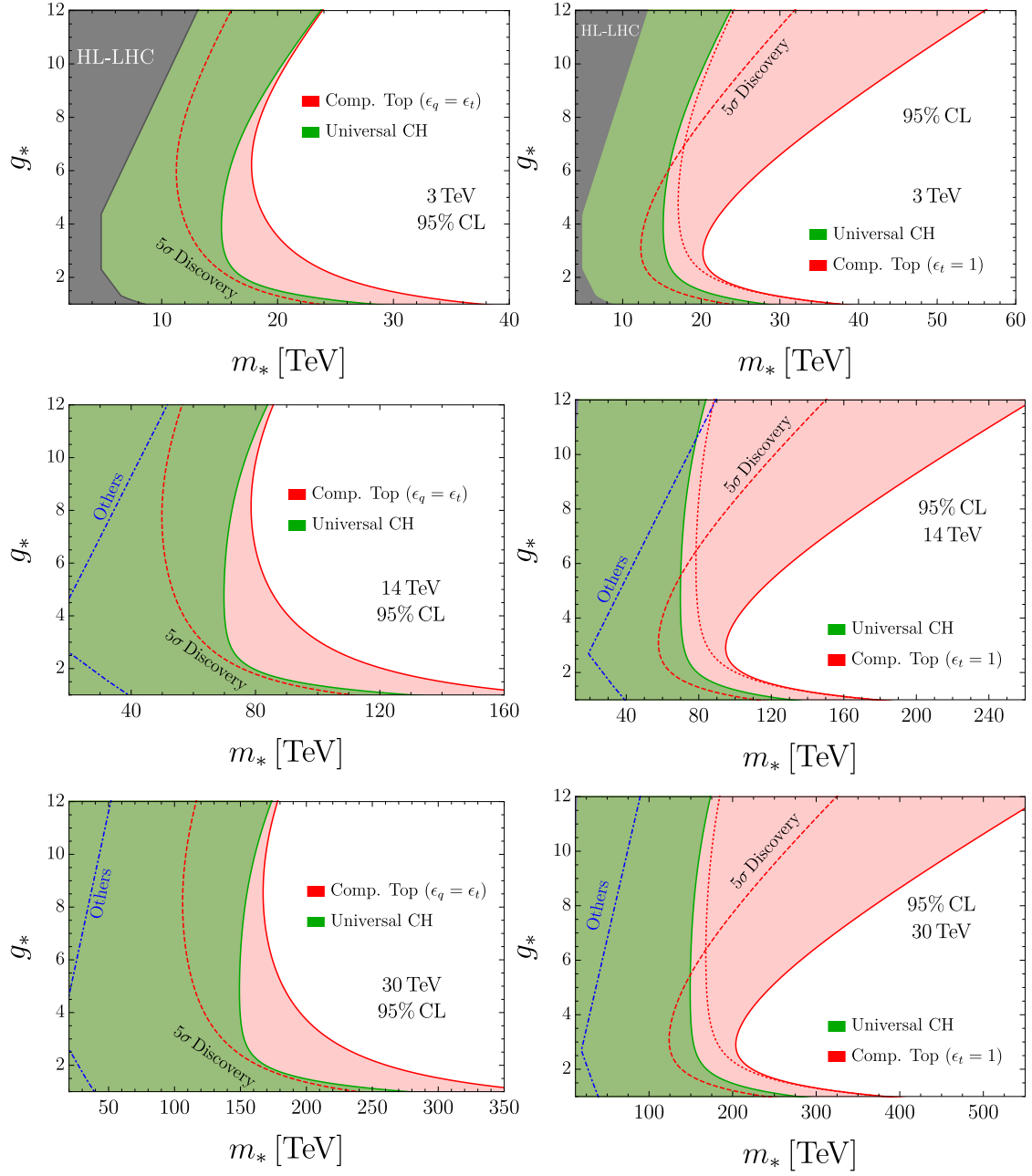


Figure B.5: The same as Figure 3.11 for various collider energies. The blue line on the equally-composite (left panels) projections are taken from [142], while on the right-handed composite top scenario are taken from [146].

Bibliography

- [1] Fred L. Wilson. Fermi's Theory of Beta Decay. *Am. J. Phys.*, 36(12):1150–1160, 1968. doi: 10.1119/1.1974382.
- [2] E. Fermi. An attempt of a theory of beta radiation. 1. *Z. Phys.*, 88:161–177, 1934. doi: 10.1007/BF01351864.
- [3] C. Patrignani et al. Review of Particle Physics. *Chin. Phys. C*, 40(10):100001, 2016. doi: 10.1088/1674-1137/40/10/100001.
- [4] A. Astbury. FOUND: W AND Z. THE DISCOVERY OF THE W AND Z INTERMEDIATE VECTOR BOSONS AT CERN'S PROTON ANTI-PROTON COLLIDER. *Phys. Bull.*, 34: 434–436, 1983.
- [5] Giuliano Panico and Andrea Wulzer. *The Composite Nambu-Goldstone Higgs*, volume 913. Springer, 2016. doi: 10.1007/978-3-319-22617-0.
- [6] Gian Francesco Giudice, Christophe Grojean, Alex Pomarol, and Riccardo Rattazzi. The strongly-interacting light higgs. *Journal of High Energy Physics*, 2007(06):045–045, 6 2007. doi: 10.1088/1126-6708/2007/06/045. URL <https://doi.org/10.1088/1126-6708/2007/06/045>.
- [7] B. Grzadkowski, M. Iskrzynski, M. Misiak, and J. Rosiek. Dimension-Six Terms in the Standard Model Lagrangian. *JHEP*, 10:085, 2010. doi: 10.1007/JHEP10(2010)085.
- [8] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt. Parton distributions for the LHC. *Eur. Phys. J. C*, 63:189–285, 2009. doi: 10.1140/epjc/s10052-009-1072-5.
- [9] Roberto Franceschini, Giuliano Panico, Alex Pomarol, Francesco Riva, and Andrea Wulzer. Electroweak Precision Tests in High-Energy Diboson Processes. *JHEP*, 02:111, 2018. doi: 10.1007/JHEP02(2018)111.
- [10] Marco Farina, Giuliano Panico, Duccio Pappadopulo, Joshua T. Ruderman, Riccardo Torre, and Andrea Wulzer. Energy helps accuracy: electroweak precision tests at hadron colliders. *Phys. Lett. B*, 772:210–215, 2017. doi: 10.1016/j.physletb.2017.06.043.
- [11] Giuliano Panico, Francesco Riva, and Andrea Wulzer. Diboson Interference Resurrection. *Phys. Lett.*, B776:473–480, 2018. doi: 10.1016/j.physletb.2017.11.068.

- [12] Chiara Aime et al. Muon Collider Physics Summary. 3 2022.
- [13] J. de Blas et al. The CLIC Potential for New Physics, 12 2018.
- [14] Daniel Schulte, Jean-Pierre Delahaye, Marcella Diemoz, Ken Long, Bruno Mansoulié, Nadia Patrone, Lenny Rivkin, Alexander M. Skrinsky, and Andrea Wulzer. Muon Collider. A Path to the Future? *PoS, EPS-HEP2019:004*, 2020. doi: 10.22323/1.364.0004.
- [15] Siyu Chen, Alfredo Glioti, Riccardo Rattazzi, Lorenzo Ricci, and Andrea Wulzer. Learning from radiation at a very high energy lepton collider. *JHEP*, 05:180, 2022. doi: 10.1007/JHEP05(2022)180.
- [16] Johann Brehmer, Kyle Cranmer, Gilles Louppe, and Juan Pavez. A Guide to Constraining Effective Field Theories with Machine Learning. *Phys. Rev. D*, 98(5):052004, 2018. doi: 10.1103/PhysRevD.98.052004.
- [17] James S. Gainer, Joseph Lykken, Konstantin T. Matchev, Stephen Mrenna, and Myeonghun Park. The Matrix Element Method: Past, Present, and Future. In *Community Summer Study 2013: Snowmass on the Mississippi*, 7 2013.
- [18] J. Alwall, A. Freitas, and O. Mattelaer. Matrix element method and qcd radiation. *Phys. Rev. D*, 83:074010, 4 2011. doi: 10.1103/PhysRevD.83.074010. URL <https://link.aps.org/doi/10.1103/PhysRevD.83.074010>.
- [19] Davison E. Soper and Michael Spannowsky. Finding physics signals with shower deconstruction. *Phys. Rev. D*, 84:074002, 2011. doi: 10.1103/PhysRevD.84.074002.
- [20] Davison E. Soper and Michael Spannowsky. Finding top quarks with shower deconstruction. *Phys. Rev. D*, 87:054012, 2013. doi: 10.1103/PhysRevD.87.054012.
- [21] John M. Campbell, R. Keith Ellis, Walter T. Giele, and Ciaran Williams. Finding the Higgs boson in decays to $Z\gamma$ using the matrix element method at Next-to-Leading Order. *Phys. Rev. D*, 87(7):073005, 2013. doi: 10.1103/PhysRevD.87.073005.
- [22] John M. Campbell, Walter T. Giele, and Ciaran Williams. Extending the Matrix Element Method to Next-to-Leading Order. In *47th Rencontres de Moriond on QCD and High Energy Interactions*, pages 319–322, 5 2012.
- [23] Jessie Shelton. Jet Substructure. In *Theoretical Advanced Study Institute in Elementary Particle Physics: Searching for New Physics at Small and Large Scales*, pages 303–340, 2013. doi: 10.1142/9789814525220_0007.
- [24] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel. Backpropagation applied to handwritten zip code recognition. *Neural Computation*, 1(4):541–551, 1989. doi: 10.1162/neco.1989.1.4.541.

- [25] Konstantina Kourou, Themis P Exarchos, Konstantinos P Exarchos, Michalis V Karamouzis, and Dimitrios I Fotiadis. Machine learning applications in cancer prognosis and prediction. *Computational and structural biotechnology journal*, 13:8–17, 2015.
- [26] Anastasia Giachanou and Fabio Crestani. Like it or not: A survey of twitter sentiment analysis methods. *ACM Computing Surveys (CSUR)*, 49(2):1–41, 2016.
- [27] Yoshua Bengio et al. Learning deep architectures for ai. *Foundations and trends® in Machine Learning*, 2(1):1–127, 2009.
- [28] Kurt Hornik. Approximation capabilities of multilayer feedforward networks. *Neural networks*, 4(2):251–257, 1991.
- [29] Shuiwang Ji, Wei Xu, Ming Yang, and Kai Yu. 3d convolutional neural networks for human action recognition. *IEEE transactions on pattern analysis and machine intelligence*, 35(1):221–231, 2012.
- [30] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 2012.
- [31] Ross Girshick. Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 1440–1448, 2015.
- [32] IBM Cloud Education. What are neural networks?, 2020. URL <https://www.ibm.com/cloud/learn/neural-networks>.
- [33] David Saad. Online algorithms and stochastic approximations. *Online Learning*, 5(3):6, 1998.
- [34] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- [35] Geoffrey E Hinton, Nitish Srivastava, Alex Krizhevsky, Ilya Sutskever, and Ruslan R Salakhutdinov. Improving neural networks by preventing co-adaptation of feature detectors. *arXiv preprint arXiv:1207.0580*, 2012.
- [36] M. Anthony and P. L. Bartlett. *Neural Network Learning: Theoretical Foundations*. Cambridge University Press., 1999. ISBN 978-0-51-162421-6.
- [37] Markus Stoye, Johann Brehmer, Gilles Louppe, Juan Pavez, and Kyle Cranmer. Likelihood-free inference with an improved cross-entropy estimator, 8 2018.
- [38] Raffaele Tito D’Agnolo and Andrea Wolz. Learning new physics from a machine. *Phys. Rev. D*, 99:015014, 1 2019. doi: 10.1103/PhysRevD.99.015014. URL <https://link.aps.org/doi/10.1103/PhysRevD.99.015014>.

Bibliography

- [39] Pierre Baldi, Kyle Cranmer, Taylor Faucett, Peter Sadowski, and Daniel Whiteson. Parameterized neural networks for high-energy physics. *Eur. Phys. J. C*, 76(5):235, 2016. doi: 10.1140/epjc/s10052-016-4099-4.
- [40] Johann Brehmer, Kyle Cranmer, Gilles Louppe, and Juan Pavez. A Guide to Constraining Effective Field Theories with Machine Learning. *Phys. Rev. D*, 98(5):052004, 2018. doi: 10.1103/PhysRevD.98.052004.
- [41] G.F. Giudice, C. Grojean, A. Pomarol, and R. Rattazzi. The Strongly-Interacting Light Higgs. *JHEP*, 06:045, 2007. doi: 10.1088/1126-6708/2007/06/045.
- [42] Da Liu, Alex Pomarol, Riccardo Rattazzi, and Francesco Riva. Patterns of Strong Coupling for LHC Searches. *JHEP*, 11:141, 2016. doi: 10.1007/JHEP11(2016)141.
- [43] Gabriel Cuomo, Luca Vecchi, and Andrea Wulzer. Goldstone Equivalence and High Energy Electroweak Physics. *SciPost Phys.*, 8(5):078, 2020. doi: 10.21468/SciPostPhys.8.5.078.
- [44] A. Kusina et al. nCTEQ15 - Global analysis of nuclear parton distributions with uncertainties. *PoS, DIS2015:041*, 2015. doi: 10.22323/1.247.0041.
- [45] D.B. Clark, E. Godat, and F.I. Olness. ManeParse : A Mathematica reader for Parton Distribution Functions. *Comput. Phys. Commun.*, 216:126–137, 2017. doi: 10.1016/j.cpc.2017.03.004.
- [46] 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. doi: 10.1007/JHEP07(2014)079.
- [47] Celine Degrande, Gauthier Durieux, Fabio Maltoni, Ken Mimasu, A. Vasquez, Eleni Vryonidou, and Cen Zhang. no. URL <https://feynrules.irmp.ucl.ac.be/wiki/SMEFTatNLO>.
- [48] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. PYTHIA 6.4 Physics and Manual. *JHEP*, 05:026, 2006. doi: 10.1088/1126-6708/2006/05/026.
- [49] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. A Brief Introduction to PYTHIA 8.1. *Comput. Phys. Commun.*, 178:852–867, 2007. doi: 10.1016/j.cpc.2008.01.036.
- [50] Eric Conte, Benjamin Fuks, and Guillaume Serret. MadAnalysis 5, A User-Friendly Framework for Collider Phenomenology. *Comput. Phys. Commun.*, 184:222–256, 2013. doi: 10.1016/j.cpc.2012.09.009.
- [51] U. Baur, Tao Han, and J. Ohnemus. Amplitude zeros in W^+Z production. *Phys. Rev. Lett.*, 72:3941–3944, 1994. doi: 10.1103/PhysRevLett.72.3941.

- [52] Lance J. Dixon and Yael Shadmi. Testing gluon selfinteractions in three jet events at hadron colliders. *Nucl. Phys. B*, 423:3–32, 1994. doi: 10.1016/0550-3213(94)90563-0. [Erratum: Nucl.Phys.B 452, 724–724 (1995)].
- [53] Jerzy Neyman and Egon Sharpe Pearson. On the Problem of the Most Efficient Tests of Statistical Hypotheses. *Phil. Trans. Roy. Soc. Lond. A*, 231(694-706):289–337, 1933. doi: 10.1098/rsta.1933.0009.
- [54] no. Poisson distribution. Encyclopedia of Mathematics. URL http://encyclopediaofmath.org/index.php?title=Poisson_distribution&oldid=48216.
- [55] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019. URL <http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf>.
- [56] Johannes Bellm, Simon Plätzer, Peter Richardson, Andrzej Siódmok, and Stephen Webster. Reweighting Parton Showers. *Phys. Rev. D*, 94(3):034028, 2016. doi: 10.1103/PhysRevD.94.034028.
- [57] Stephan Duell, Florian Bernlochner, Zoltan Ligeti, Michele Papucci, and Dean Robinson. HAMMER: Reweighting tool for simulated data samples. *PoS, ICHEP2016*:1074, 2017. doi: 10.22323/1.282.1074.
- [58] Stefano Frixione and Bryan R. Webber. Matching NLO QCD computations and parton shower simulations. *JHEP*, 06:029, 2002. doi: 10.1088/1126-6708/2002/06/029.
- [59] Olivier Mattelaer. On the maximal use of Monte Carlo samples: re-weighting events at NLO accuracy. *Eur. Phys. J. C*, 76(12):674, 2016. doi: 10.1140/epjc/s10052-016-4533-7.
- [60] R. Frederix, S. Frixione, S. Prestel, and P. Torrielli. On the reduction of negative weights in MC@NLO-type matching procedures. *JHEP*, 07:238, 2020. doi: 10.1007/JHEP07(2020)238.
- [61] Jean Pierre Delahaye, Marcella Diemoz, Ken Long, Bruno Mansoulié, Nadia Pastrone, Lenny Rivkin, Daniel Schulte, Alexander Skrinsky, and Andrea Wulzer. Muon Colliders, 1 2019.
- [62] The international muon collider collaboration. <https://muoncollider.web.cern.ch>.
- [63] Dario Buttazzo, Diego Redigolo, Filippo Sala, and Andrea Tesi. Fusing Vectors into Scalars at High Energy Lepton Colliders. *JHEP*, 11:144, 2018. doi: 10.1007/JHEP11(2018)144.

Bibliography

- [64] Roberto Franceschini. High(est) energy lepton colliders. *Int. J. Mod. Phys. A*, 36(22): 2142016, 2021. doi: 10.1142/S0217751X21420161.
- [65] Rodolfo Capdevilla, Federico Meloni, Rosa Simoniello, and Jose Zurita. Hunting wino and higgsino dark matter at the muon collider with disappearing tracks. *JHEP*, 06:133, 2021. doi: 10.1007/JHEP06(2021)133.
- [66] Hind Al Ali et al. The Muon Smasher’s Guide, 3 2021.
- [67] Dario Buttazzo, Roberto Franceschini, and Andrea Wulzer. Two Paths Towards Precision at a Very High Energy Lepton Collider. *JHEP*, 05:219, 2021. doi: 10.1007/JHEP05(2021)219.
- [68] Tao Han, Da Liu, Ian Low, and Xing Wang. Electroweak couplings of the Higgs boson at a multi-TeV muon collider. *Phys. Rev. D*, 103(1):013002, 2021. doi: 10.1103/PhysRevD.103.013002.
- [69] Nazar Bartosik et al. Preliminary Report on the Study of Beam-Induced Background Effects at a Muon Collider, 5 2019.
- [70] Rodolfo Capdevilla, David Curtin, Yonatan Kahn, and Gordan Krnjaic. Discovering the physics of $(g - 2)_\mu$ at future muon colliders. *Phys. Rev. D*, 103(7):075028, 2021. doi: 10.1103/PhysRevD.103.075028.
- [71] Maximilian Ruhdorfer, Ennio Salvioni, and Andreas Weiler. A Global View of the Off-Shell Higgs Portal. *SciPost Phys.*, 8:027, 2020. doi: 10.21468/SciPostPhys.8.2.027.
- [72] Mauro Chiesa, Fabio Maltoni, Luca Mantani, Barbara Mele, Fulvio Piccinini, and Xiaoran Zhao. Measuring the quartic Higgs self-coupling at a multi-TeV muon collider. *JHEP*, 09: 098, 2020. doi: 10.1007/JHEP09(2020)098.
- [73] Antonio Costantini, Federico De Lillo, Fabio Maltoni, Luca Mantani, Olivier Mattelaer, Richard Ruiz, and Xiaoran Zhao. Vector boson fusion at multi-TeV muon colliders. *JHEP*, 09:080, 2020. doi: 10.1007/JHEP09(2020)080.
- [74] Rodolfo Capdevilla, David Curtin, Yonatan Kahn, and Gordan Krnjaic. A No-Lose Theorem for Discovering the New Physics of $(g - 2)_\mu$ at Muon Colliders, 1 2021.
- [75] Nazar Bartosik et al. Detector and Physics Performance at a Muon Collider. *JINST*, 15 (05):P05001, 2020. doi: 10.1088/1748-0221/15/05/P05001.
- [76] Wen Yin and Masahiro Yamaguchi. Muon $g - 2$ at multi-TeV muon collider, 12 2020.
- [77] Jan Kalinowski, Tania Robens, Dorota Sokolowska, and Aleksander Filip Zarnecki. IDM Benchmarks for the LHC and Future Colliders. *Symmetry*, 13(6):991, 2021. doi: 10.3390/sym13060991.

-
- [78] Guo-yuan Huang, Farinaldo S. Queiroz, and Werner Rodejohann. Gauged $L_\mu - L_\tau$ at a muon collider. *Phys. Rev. D*, 103(9):095005, 2021. doi: 10.1103/PhysRevD.103.095005.
- [79] Wei Liu and Ke-Pan Xie. Probing electroweak phase transition with multi-TeV muon colliders and gravitational waves. *JHEP*, 04:015, 2021. doi: 10.1007/JHEP04(2021)015.
- [80] Tao Han, Shuailong Li, Shufang Su, Wei Su, and Yongcheng Wu. Heavy Higgs bosons in 2HDM at a muon collider. *Phys. Rev. D*, 104(5):055029, 2021. doi: 10.1103/PhysRevD.104.055029.
- [81] Salvatore Bottaro, Alessandro Strumia, and Natascia Vignaroli. Minimal Dark Matter bound states at future colliders. *JHEP*, 06:143, 2021. doi: 10.1007/JHEP06(2021)143.
- [82] Tong Li, Michael A. Schmidt, Chang-Yuan Yao, and Man Yuan. Charged lepton flavor violation in light of the muon magnetic moment anomaly and colliders. *Eur. Phys. J. C*, 81(09):811, 2021. doi: 10.1140/epjc/s10052-021-09569-9.
- [83] Pouya Asadi, Rodolfo Capdevilla, Cari Cesarotti, and Samuel Homiller. Searching for leptoquarks at future muon colliders. *JHEP*, 10:182, 2021. doi: 10.1007/JHEP10(2021)182.
- [84] M. Sahin and A. Caliskan. Excited muon production at muon colliders via contact interaction, 5 2021.
- [85] Junmou Chen, Chih-Ting Lu, and Yongcheng Wu. Measuring Higgs boson self-couplings with $2 \rightarrow 3$ VBS processes. *JHEP*, 10:099, 2021. doi: 10.1007/JHEP10(2021)099.
- [86] Gholamhossein Haghighat and Mojtaba Mohammadi Najafabadi. Search for lepton-flavor-violating ALPs at a future muon collider and utilization of polarization-induced effects, 6 2021.
- [87] Salvatore Bottaro, Dario Buttazzo, Marco Costa, Roberto Franceschini, Paolo Panci, Diego Redigolo, and Ludovico Vittorio. Closing the window on WIMP Dark Matter, 7 2021.
- [88] Priyotosh Bandyopadhyay, Anirban Karan, and Chandrima Sen. Discerning Signatures of Seesaw Models and Complementarity of Leptonic Colliders, 11 2020.
- [89] Chandrima Sen, Priyotosh Bandyopadhyay, Saunak Dutta, and Aleesha KT. Displaced Higgs production in Type-III Seesaw at the LHC/FCC, MATHUSLA and Muon collider, 7 2021.
- [90] Tao Han, Wolfgang Kilian, Nils Kreher, Yang Ma, Jürgen Reuter, Tobias Striegl, and Keping Xie. Precision Test of the Muon-Higgs Coupling at a High-energy Muon Collider, 8 2021.
- [91] Priyotosh Bandyopadhyay, Anirban Karan, and Rusa Mandal. Distinguishing signatures of scalar leptoquarks at hadron and muon colliders, 8 2021.

Bibliography

- [92] Radovan Dermisek, Keith Hermanek, and Navin McGinnis. Di-Higgs and tri-Higgs boson signals of muon $g-2$ at a muon collider. *Phys. Rev. D*, 104(9):L091301, 2021. doi: 10.1103/PhysRevD.104.L091301.
- [93] Sitian Qian, Congqiao Li, Qiang Li, Fanqiang Meng, Jie Xiao, Tianyi Yang, Meng Lu, and Zhengyun You. Searching for heavy leptoquarks at a muon collider, 9 2021.
- [94] Richard Ruiz, Antonio Costantini, Fabio Maltoni, and Olivier Mattelaer. The Effective Vector Boson Approximation in High-Energy Muon Collisions, 11 2021.
- [95] Mauro Chiesa, Barbara Mele, and Fulvio Piccinini. Multi Higgs production via photon fusion at future multi-TeV muon colliders, 9 2021.
- [96] Wei Liu, Ke-Pan Xie, and Zihan Yi. Testing leptogenesis at the LHC and future muon colliders: a Z' scenario, 9 2021.
- [97] Dario Buttazzo and Paride Paradisi. Probing the muon $g - 2$ anomaly with the Higgs boson at a muon collider. *Phys. Rev. D*, 104(7):075021, 2021. doi: 10.1103/PhysRevD.104.075021.
- [98] Luca Di Luzio, Ramona Gröber, and Giuliano Panico. Probing new electroweak states via precision measurements at the LHC and future colliders. *JHEP*, 01:011, 2019. doi: 10.1007/JHEP01(2019)011.
- [99] Tao Han, Zhen Liu, Lian-Tao Wang, and Xing Wang. WIMPs at High Energy Muon Colliders. *Phys. Rev. D*, 103(7):075004, 2021. doi: 10.1103/PhysRevD.103.075004.
- [100] Junmou Chen, Tong Li, Chih-Ting Lu, Yongcheng Wu, and Chang-Yuan Yao. The measurement of Higgs self-couplings through $2 \rightarrow 3$ VBS in future muon colliders, 12 2021.
- [101] Rodolfo Capdevilla, David Curtin, Yonatan Kahn, and Gordan Krnjaic. Systematically Testing Singlet Models for $(g - 2)_\mu$, 12 2021.
- [102] Roberto Franceschini and Xiaoran Zhao. to appear.
- [103] Marcello Ciafaloni, Paolo Ciafaloni, and Denis Comelli. Bloch-Nordsieck violating electroweak corrections to inclusive TeV scale hard processes. *Phys. Rev. Lett.*, 84: 4810–4813, 2000. doi: 10.1103/PhysRevLett.84.4810.
- [104] Marcello Ciafaloni, Paolo Ciafaloni, and Denis Comelli. Electroweak Bloch-Nordsieck violation at the TeV scale: 'Strong' weak interactions? *Nucl. Phys. B*, 589:359–380, 2000. doi: 10.1016/S0550-3213(00)00508-3.
- [105] Victor S. Fadin, L. N. Lipatov, Alan D. Martin, and M. Melles. Resummation of double logarithms in electroweak high-energy processes. *Phys. Rev. D*, 61:094002, 2000. doi: 10.1103/PhysRevD.61.094002.

-
- [106] Michael Melles. Subleading Sudakov logarithms in electroweak high-energy processes to all orders. *Phys. Rev. D*, 63:034003, 2001. doi: 10.1103/PhysRevD.63.034003.
- [107] Jui-yu Chiu, Frank Golf, Randall Kelley, and Aneesh V. Manohar. Electroweak Sudakov corrections using effective field theory. *Phys. Rev. Lett.*, 100:021802, 2008. doi: 10.1103/PhysRevLett.100.021802.
- [108] Jui-yu Chiu, Frank Golf, Randall Kelley, and Aneesh V. Manohar. Electroweak Corrections in High Energy Processes using Effective Field Theory. *Phys. Rev. D*, 77:053004, 2008. doi: 10.1103/PhysRevD.77.053004.
- [109] Jui-yu Chiu, Andreas Fuhrer, Randall Kelley, and Aneesh V. Manohar. Factorization Structure of Gauge Theory Amplitudes and Application to Hard Scattering Processes at the LHC. *Phys. Rev. D*, 80:094013, 2009. doi: 10.1103/PhysRevD.80.094013.
- [110] Jui-yu Chiu, Andreas Fuhrer, Randall Kelley, and Aneesh V. Manohar. Soft and Collinear Functions for the Standard Model. *Phys. Rev. D*, 81:014023, 2010. doi: 10.1103/PhysRevD.81.014023.
- [111] Aneesh Manohar, Brian Shotwell, Christian Bauer, and Sascha Turczyk. Non-cancellation of electroweak logarithms in high-energy scattering. *Phys. Lett. B*, 740: 179–187, 2015. doi: 10.1016/j.physletb.2014.11.050.
- [112] Aneesh V. Manohar and Wouter J. Waalewijn. Electroweak Logarithms in Inclusive Cross Sections. *JHEP*, 08:137, 2018. doi: 10.1007/JHEP08(2018)137.
- [113] Christian W. Bauer, Nicolas Ferland, and Bryan R. Webber. Combining initial-state resummation with fixed-order calculations of electroweak corrections. *JHEP*, 04:125, 2018. doi: 10.1007/JHEP04(2018)125.
- [114] Marcello Ciafaloni, Paolo Ciafaloni, and Denis Comelli. Towards collinear evolution equations in electroweak theory. *Phys. Rev. Lett.*, 88:102001, 2002. doi: 10.1103/PhysRevLett.88.102001.
- [115] Paolo Ciafaloni and Denis Comelli. Electroweak evolution equations. *JHEP*, 11:022, 2005. doi: 10.1088/1126-6708/2005/11/022.
- [116] Bartosz Fornal, Aneesh V. Manohar, and Wouter J. Waalewijn. Electroweak Gauge Boson Parton Distribution Functions. *JHEP*, 05:106, 2018. doi: 10.1007/JHEP05(2018)106.
- [117] Christian W. Bauer, Nicolas Ferland, and Bryan R. Webber. Standard Model Parton Distributions at Very High Energies. *JHEP*, 08:036, 2017. doi: 10.1007/JHEP08(2017)036.
- [118] Christian W. Bauer, Davide Provasoli, and Bryan R. Webber. Standard Model Fragmentation Functions at Very High Energies. *JHEP*, 11:030, 2018. doi: 10.1007/JHEP11(2018)030.

Bibliography

- [119] Tao Han, Yang Ma, and Keping Xie. High energy leptonic collisions and electroweak parton distribution functions. *Phys. Rev. D*, 103(3):L031301, 2021. doi: 10.1103/PhysRevD.103.L031301.
- [120] Junmou Chen, Tao Han, and Brock Tweedie. Electroweak Splitting Functions and High Energy Showering. *JHEP*, 11:093, 2017. doi: 10.1007/JHEP11(2017)093.
- [121] Sally Dawson. The Effective W Approximation. *Nucl. Phys. B*, 249:42–60, 1985. doi: 10.1016/0550-3213(85)90038-0.
- [122] Zoltan Kunszt and Davison E. Soper. On the Validity of the Effective W Approximation. *Nucl. Phys. B*, 296:253–289, 1988. doi: 10.1016/0550-3213(88)90673-6.
- [123] Pascal Borel, Roberto Franceschini, Riccardo Rattazzi, and Andrea Wulzer. Probing the Scattering of Equivalent Electroweak Bosons. *JHEP*, 06:122, 2012. doi: 10.1007/JHEP06(2012)122.
- [124] T. Kinoshita. Mass singularities of Feynman amplitudes. *J. Math. Phys.*, 3:650–677, 1962. doi: 10.1063/1.1724268.
- [125] T. D. Lee and M. Nauenberg. Degenerate systems and mass singularities. *Phys. Rev.*, 133:B1549–B1562, Mar 1964. doi: 10.1103/PhysRev.133.B1549. URL <https://link.aps.org/doi/10.1103/PhysRev.133.B1549>.
- [126] Paolo Ciafaloni and Denis Comelli. The Importance of weak bosons emission at LHC. *JHEP*, 09:055, 2006. doi: 10.1088/1126-6708/2006/09/055.
- [127] Ansgar Denner and Stefano Pozzorini. One loop leading logarithms in electroweak radiative corrections. 1. Results. *Eur. Phys. J. C*, 18:461–480, 2001. doi: 10.1007/s100520100551.
- [128] Ansgar Denner and Stefano Pozzorini. One loop leading logarithms in electroweak radiative corrections. 2. Factorization of collinear singularities. *Eur. Phys. J. C*, 21:63–79, 2001. doi: 10.1007/s100520100721.
- [129] Marcello Ciafaloni, Paolo Ciafaloni, and Denis Comelli. Anomalous Sudakov Form Factors. *JHEP*, 03:072, 2010. doi: 10.1007/JHEP03(2010)072.
- [130] Ansgar Denner, B. Jantzen, and S. Pozzorini. Two-loop electroweak next-to-leading logarithmic corrections to massless fermionic processes. *Nucl. Phys. B*, 761:1–62, 2007. doi: 10.1016/j.nuclphysb.2006.10.014.
- [131] Johann H. Kuhn, S. Moch, A. A. Penin, and Vladimir A. Smirnov. Next-to-next-to-leading logarithms in four fermion electroweak processes at high-energy. *Nucl. Phys. B*, 616:286–306, 2001. doi: 10.1016/S0550-3213(01)00454-0. [Erratum: *Nucl.Phys.B* 648, 455–456 (2003)].
- [132] Lorenzo Ricci, Riccardo Torre, and Andrea Wulzer. On the $W&Y$ interpretation of high-energy Drell-Yan measurements. *JHEP*, 02:144, 2021. doi: 10.1007/JHEP02(2021)144.

-
- [133] R. Franceschini et al. The CLIC Potential for New Physics. 3/2018, 12 2018. doi: 10.23731/CYRM-2018-003.
- [134] H. Abramowicz et al. Top-Quark Physics at the CLIC Electron-Positron Linear Collider. *JHEP*, 11:003, 2019. doi: 10.1007/JHEP11(2019)003.
- [135] B. Grzadkowski, M. Iskrzynski, M. Misiak, and J. Rosiek. Dimension-Six Terms in the Standard Model Lagrangian. *JHEP*, 10:085, 2010. doi: 10.1007/JHEP10(2010)085.
- [136] Riccardo Barbieri, Alex Pomarol, Riccardo Rattazzi, and Alessandro Strumia. Electroweak symmetry breaking after LEP-1 and LEP-2. *Nucl. Phys. B*, 703:127–146, 2004. doi: 10.1016/j.nuclphysb.2004.10.014.
- [137] James D. Wells and Zhengkang Zhang. Effective theories of universal theories. *JHEP*, 01:123, 2016. doi: 10.1007/JHEP01(2016)123.
- [138] Gauthier Durieux and Oleksii Matsedonskyi. The top-quark window on compositeness at future lepton colliders. *JHEP*, 01:072, 2019. doi: 10.1007/JHEP01(2019)072.
- [139] Giuliano Panico, Lorenzo Ricci, and Andrea Wulzer. High-energy EFT probes with fully differential Drell-Yan measurements. *JHEP*, 07:086, 2021. doi: 10.1007/JHEP07(2021)086.
- [140] J. de Blas et al. Higgs Boson Studies at Future Particle Colliders. *JHEP*, 01:139, 2020. doi: 10.1007/JHEP01(2020)139.
- [141] Céline Degrande, Gauthier Durieux, Fabio Maltoni, Ken Mimasu, Eleni Vryonidou, and Cen Zhang. Automated one-loop computations in the standard model effective field theory. *Phys. Rev. D*, 103(9):096024, 2021. doi: 10.1103/PhysRevD.103.096024.
- [142] Richard Keith Ellis et al. Physics Briefing Book: Input for the European Strategy for Particle Physics Update 2020. 10 2019.
- [143] David B. Kaplan and Howard Georgi. $SU(2) \times U(1)$ Breaking by Vacuum Misalignment. *Phys. Lett. B*, 136:183–186, 1984. doi: 10.1016/0370-2693(84)91177-8.
- [144] Kaustubh Agashe, Roberto Contino, and Alex Pomarol. The Minimal composite Higgs model. *Nucl. Phys. B*, 719:165–187, 2005. doi: 10.1016/j.nuclphysb.2005.04.035.
- [145] Giuliano Panico and Andrea Wulzer. *The Composite Nambu-Goldstone Higgs*, volume 913. Springer, 2016. doi: 10.1007/978-3-319-22617-0.
- [146] Giovanni Banelli, Ennio Salvioni, Javi Serra, Tobias Theil, and Andreas Weiler. The Present and Future of Four Top Operators. *JHEP*, 02:043, 2021. doi: 10.1007/JHEP02(2021)043.
- [147] David B. Kaplan. Flavor at SSC energies: A New mechanism for dynamically generated fermion masses. *Nucl. Phys. B*, 365:259–278, 1991. doi: 10.1016/S0550-3213(05)80021-5.

Bibliography

- [148] Gauthier Durieux, Martín Perelló, Marcel Vos, and Cen Zhang. Global and optimal probes for the top-quark effective field theory at future lepton colliders. *JHEP*, 10:168, 2018. doi: 10.1007/JHEP10(2018)168.
- [149] Riccardo Barbieri. Flavour and Higgs compositeness: present and “near” future. 10 2019.
- [150] Thomas Appelquist, Bogdan A. Dobrescu, and Adam R. Hopper. Nonexotic Neutral Gauge Bosons. *Phys. Rev. D*, 68:035012, 2003. doi: 10.1103/PhysRevD.68.035012.
- [151] L. D. Landau and E. M. Lifshitz. *Quantum Electrodynamics*, volume 4 of *Course of Theoretical Physics*. Butterworth-Heinemann, 1982. ISBN 978-0-7506-3371-0.
- [152] V V Sudakov. Vertex parts at very high energies in quantum electrodynamics. *Zh. Eksp. Teor. Fiz.*, 30:87–95, 1956. URL <http://cds.cern.ch/record/478063>.
- [153] Andrea Wulzer. An Equivalent Gauge and the Equivalence Theorem. *Nucl. Phys. B*, 885: 97–126, 2014. doi: 10.1016/j.nuclphysb.2014.05.021.

Education

EPFL, PhD; Lausanne, Switzerland

2018—2022

Particle Physics Phenomenology

Supervisor: Prof. Andrea Wulzer

University of Oxford, Master of Science; Oxford, United Kingdom

2017—2018

Mathematical and Theoretical Physics

National University of Singapore, Bachelor of Science; Singapore

2013 —2017

Physics and Mathematics

Publications

Chen, S., Glioti, A., Panico, G., & Wulzer, A. (2021). Parametrized classifiers for optimal EFT sensitivity. *Journal of High Energy Physics*, 2021(5), 1-39.

Chen, S., Glioti, A., Rattazzi, R., Ricci, L., & Wulzer, A. (2022). Learning from radiation at a very high energy lepton collider. *Journal of High Energy Physics*, 2022(5), 1-58.

Aimè, C., Apyan, A., Mahmoud, M. A., Bartosik, N., Batsch, F., Bertolin, A., ... & Zurita, J. (2022). Muon Collider Physics Summary. *arXiv preprint arXiv:2203.07256*.

De Blas, J., Buttazzo, D., Capdevilla, R., Curtin, D., Franceschini, R., Maltoni, F., ... & Zurita, J. (2022). The physics case of a 3 TeV muon collider stage. *arXiv preprint arXiv:2203.07261*.

Chen, S., Huang, G., Piccioli, G., & Zdeborová, L. (2022). Planted X Y model: Thermodynamics and inference. *Physical Review E* 106.5 (2022): 054115.

Talks

IML Machine Learning Working Group, Jul 2021

Higgs Hunting 2022, Sep 2022