Università di Genova

Scuola di Scienze Matematiche, Fisiche e Naturali

Corso di Dottorato in Fisica e Nanoscienze

Drell–Yan Phenomenology in the SMEFT and Statistical Validation of Machine Learning Generators

SUPERVISORS: Riccardo Torre
Simone Marzani

CANDIDATE:
Samuele Grossi

XXXVIII ciclo

Final examination:
May 22, 2026.

Università degli Studi di Genova,
Dipartimento di Fisica,
Genova, Italia.

Thesis submitted to Università di Genova for the degree of **Dottore di Ricerca in Fisica e Nanoscienze.**

External Referees:

Prof.	**Tilman Plehn**	Heidelberg University
Prof.	**Andrea Wulzer**	ICREA, Barcelona and Barcelona, IFAE

Examination Committee:

Prof.	**Riccardo Umberto Valente**	INFN Milano LASA
Prof.	**Paolo Francavilla**	Università di Pisa
Prof.	**Giovanni Ridolfi**	Università di Genova

Abstract

The Standard Model of particle physics encodes our current knowledge of fundamental particles and their interactions. Although it successfully describes a wide range of experimental observations, it fails to address several open questions, such as the matter–antimatter asymmetry, the origin of neutrino masses, and the nature of dark matter. Moreover, it presents a conceptual issue in the naturalness problem, which implies an extreme fine-tuning of parameters to justify theoretically the measured value of the Higgs boson mass. The prevailing view within the particle physics community is that the Standard Model should be extended by new, heavy degrees of freedom that lie beyond the current experimental reach and that could provide answers to these open questions.

The current lack of direct observations of new particles implies that the best approach to explore physics at high energy scales is through an indirect search. The Standard Model Effective Field Theory (SMEFT) offers a model agnostic framework for such searches, and has therefore attracted interest within the High Energy Physics community. In part I of this thesis, we employ the SMEFT framework to investigate possible New Physics scenarios through the study of the Drell-Yan process in the context of the LHC. We analyze a wide range of operators, including both leading and suppressed contributions, using different observables and binning strategies.

In part II, we discuss the impact of Machine Learning techniques on particle physics simulations, explaining why these approaches will become necessary to respect the precision requirements posed by experimental improvements, particularly with the advent of the High-Luminosity LHC. Our work focuses on the validation of data provided by Machine Learning-based generative models, with the goal of developing robust statistical procedures to assess their reliability and to compare different models.

Declaration

This thesis represents the research work conducted during my PhD at the University of Genova. It includes the original results published in the following papers:

The discussion presented in Ch. 2 is based on the results published in [1]:

- S. Grossi, R. Torre, "More variables or more bins? Impact on the EFT interpretation of Drell-Yan measurements", *European Physics Journal C* 84 (2024) 7, 713; e-Print: 2404.10569 [hep-ph].

Chapter 3 includes results presented in [2]:

- S. Grossi, X. Li, L. Rolla, R. Torre, "Isolating chirality-breaking SMEFT operators with Drell-Yan angular analysis", e-Print: 2512.05018 [hep-ph].

Chapter 5 is based on the results presented in [3]:

- S. Grossi, M. Letizia, R. Torre, "Refereeing the referees: evaluating two-sample tests for validating generators in precision sciences", Machine Learning Science and Technology, 6 (2025) 1, 015052; e-Print: 2409.16336 [stat],

Chapter 6 incorporates results presented in [4]:

- S. Grossi, M. Letizia, R. Torre, "Comparing Generative Models with the New Physics Learning Machine", Nuclear Physics B, 1024 (2026), 117349; e-Print: 2508.02275 [stat].

Introduction

The Standard Model (SM) of particle physics is the theory describing our knowledge of the fundamental particles and their interactions. Its predictions have been rigorously tested over decades of experiments, with several remarkable confirmations, such as the discovery of the Higgs boson. Nonetheless, some evidences suggest that it can not be the definitive theory of nature. For instance, it does not incorporate gravity, predicts massless neutrinos, fails to account for the matter-antimatter asymmetry observed in the universe, and offers no explanation for dark matter or dark energy. Also, the theory itself exhibits some phenomenological and conceptual limitations, such as the naturalness problem, which is closely related to the huge separation between the scales of the different forces. Indeed, the QCD scale lies around $\Lambda_{\rm QCD} \sim 200\,\text{MeV}$, the EW scale is set by the Fermi constant, and is around 100 GeV, while the scale of gravity, as set by the Newton constant (or, equivalently, the Planck mass), is around $m_{\rm P} \sim 10^{19}$ GeV. This huge separation of scales implies an extreme fine-tuning of parameters, necessary to cancel high energy corrections to the Higgs boson mass. For these reasons, the prevailing view among particle physicists is that the SM represents a low-energy manifestation of a more fundamental ultraviolet-complete theory.

A significant part of the high-energy physics community is therefore dedicated to the search for physics Beyond the Standard Model (BSM), often called New Physics (NP). Given the experimental limitations in reaching higher and higher energies, relying solely on direct searches is not sustainable and indirect searches have become the most promising way to investigate NP. Furthermore, adopting an agnostic perspective, without committing to a specific model, can be a valuable strategy to avoid overlooking potential scenarios. In this context, the Standard Model Effective Field Theory (SMEFT) provides a powerful framework for conducting model independent, indirect searches, and is widely used by theorists for this purpose. In the SMEFT framework, the NP contributions manifest as small deviations from SM predictions, making precision a critical requirement. Hence, this approach becomes even more promising in light of the increased statistical power that will be available during the High-Luminosity (HL) phase of the Large Hadron Collider (LHC).

This thesis is divided into two parts, each addressing a different topic.

In the first part, we analyze the di-lepton production at the LHC, referred to as Drell-Yan (DY) process, within the context of the SMEFT. Chapter 1 begins with an introduction to the SM, highlighting the open questions it fails to address, and a brief overview of the Effective Field Theory (EFT) approach. We then introduce the SMEFT, explaining its relevance for model-independent searches for NP at the LHC. Finally, we describe the DY process in the SMEFT context, which serves as the starting point for the subsequent analyses. In Chapter 2 we use this process to study the effects the most relevant dimension-6 operators, which interfere with the SM and whose effects grow with the energy. We also compare different binning schemes, in order to identify the most promising analysis strategy. In Chapter 3, we exploit again the DY process, this time focusing on a specific angular observable that allows to study some dimension-6 operators that would be otherwise suppressed. These operators do not interfere with the SM; the first class is more relevant

around the Z peak, while the second exhibit energy-growing effects and become more significant in the high-energy tails of kinematic distributions.

In the second part of the thesis, we investigate methods for determining whether two independent datasets are drawn from the same probability density function. This is a crucial task in modern high-energy physics (HEP), due to the growing use of machine learning (ML) based generators as a potential alternative to traditional Monte Carlo simulators. These generators provide synthetic data more efficiently, making them particularly appealing in light of the increased precision required with the advent of the HL-LHC, which in turn demands a substantially larger amount of simulated data. Their validation is therefore an important step for the HEP community. In Chapter 4, we discuss the motivations for which the particle physics community is moving towards ML-based simulators and illustrate how they can improve current simulation techniques. We also emphasize the necessity of validating these models and introduce the statistical tool of two-sample testing, which is employed throughout the subsequent Chapters. In Chapter 5, we propose a set of metrics based on 1D distances that can be used to perform two-sample tests for the validation of generative models. We introduce a general framework that includes different configurations and enables a systematic comparison of the performance of the various metrics across multiple scenarios. Finally, in Chapter 6, we consider a classifier trained to learn the likelihood ratio as a test-statistic. Considering the same framework introduced in Chapter 5, we are able to assess its performance more clearly and determine, across different cases, whether its use is advantageous.

Contents

Part I

The Drell-Yan process in the Standard Model Effective Field Theory

Chapter 1

The Standard Model Effective Field Theory at the Large Hadron Collider

In this Chapter, we introduce the Standard Model Effective Field Theory (SMEFT) as a tool to search for possible signals of New Physics (NP) at the Large Hadron Collider (LHC), the largest particle collider available to date, used to perform high energy physics experiments at CERN. The conceptual framework highlighted here will serve as the foundation for the entire first part of this thesis.

The Chapter is organized as follows. In Section 1.1 we introduce the basic concepts of Effective Field Theory (EFT). Our purpose is not to provide a formal or exhaustive treatment, but rather to convey the key concepts. More formal and in depth discussions about the topic can be found in the literature [5–12]. In Section 1.2 we discuss the distinction between masses and scales, an aspect that is not immediately apparent when adopting the usual natural units. This is crucial in order to understand the range of validity of an EFT. In Section 1.3 we provide a brief recap of the main features of the Standard Model (SM) and point out its limitations. These limitations motivate the search for physics Beyond the Standard Model (BSM). In Section 1.4, we describe how the EFT framework can be employed to extend the SM in a systematic way, and how this approach can be exploited in the context of the LHC to identify potential NP effects. Finally, in Section 1.5, we introduce the Drell-Yan (DY) process, one of the most studied channels in particle physics, and discuss how it can be analysed within the SMEFT framework. This process will constitute the central topic of Chapters 2 and 3.

1.1 Effective Field Theory basics

The intuitive idea behind Effective Field Theory (EFT) is conceptually simple: up to energy scales where no new degrees of freedom are present, it is possible to perform accurate calculations and obtain reliable predictions even without the complete knowledge of the underlying fundamental theory. This principle is something that we experience in everyday life, where most engineering

applications rely solely on classical physics, without taking quantum mechanics, special relativity or even general relativity into account. In the following, we will illustrate how this idea can be implemented in a mathematically consistent way within the framework of an interacting quantum field theory.

The correlation functions of a quantum field theory generally depend on complicated functions of the kinematic variables of the incoming and outgoing particles. These functions exhibit non-analytic behaviors, such as cuts and poles, associated with kinematic configurations that allow the production of physical intermediate states. In particular, such non-analyticities arise from the singularities of propagators, which typically take the form $\propto 1/(p^2 - m^2)$, and become divergent when the intermediate particle can go on-shell.

However, when the kinematics are far from this region, i.e. when the energy is much lower than the mass of the intermediate particle, the contribution of such a propagator to the correlation function becomes simple and can be well approximated by a Taylor expansion.

A typical example is the β decay of a muon, where the energies involved in the process are far below the mass of the W boson. Therefore, the propagator can be expanded as:

$$\frac{1}{p^2 - m_W^2} \simeq -\frac{1}{m_W^2} - \frac{p^2}{m_W^4} + \dots \ , \tag{1.1}$$

and retaining only the terms relevant at those energies:

$$\frac{1}{p^2 - m_W^2} \simeq -\frac{1}{m_W^2} + \mathcal{O}\left(\frac{p^2}{m_W^2}\right), \tag{1.2}$$

is a good approximation.

An Effective Field Theory (EFT), is a quantum field theory in itself, describing physics in an energy regime where certain degrees of freedom can be neglected. In the previous example, for instance, the W boson does not need to be treated as a dynamical field at the typical energies of the β decay. Rather than approximating at the level of amplitudes, one can construct a lagrangian that directly reproduces those low-energy amplitudes. The EFT lagrangian includes only the light degrees of freedom and is organized as a Taylor expansion in a small parameter (typically, the ratio of the characteristic energy of the process to a higher ultraviolet (UV) scale Λ). This expansion defines what is known as *power counting*, which provides a systematic way to rank the terms in the lagrangian according to their relevance at low energies. To get an idea of why, it is useful to look at the action. Denoting by ϕ_L the light degrees of freedom, relevant at scales below Λ and with ϕ_H the heavy states, negligible at those scales, one can write:

$$\int \mathcal{D}\phi_L \mathcal{D}\phi_H e^{iS(\phi_L,\phi_H)} = \int \mathcal{D}\phi_L e^{iS_\Lambda(\phi_L)}\,, \tag{1.3}$$

with

$$e^{iS_\Lambda(\phi_L)} = \int \mathcal{D}\phi_H e^{iS(\phi_L,\phi_H)}\,. \tag{1.4}$$

The effective action $S_\Lambda(\phi_L)$ can be expanded in terms of local operators $\mathcal{O}_i$:

$$S_\Lambda(\phi_L) = \int d^D x \sum_i g_i \mathcal{O}_i \,, \tag{1.5}$$

where the operators in the sum are all those allowed by the symmetries of the problem. In natural unit, we denote by d_i the mass dimensions (or canonical dimensions) of the operator $\mathcal{O}_i$: $[\mathcal{O}_i] = d_i$. Consequently, the dimension of the coupling is: $[g_i] = D - d_i$.

Let us consider, for simplicity, a scalar theory, where $[\phi] = D/2 - 1$. If an operator is made up by M scalar fields and N derivatives, then:

$$d_i = M(D/2 - 1) + N \,. \tag{1.6}$$

We define the dimensionless coupling $\lambda_i = \Lambda^{d_i - D} g_i$. Each term in eq. (1.5) then scales with energy as:

$$\int d^D x \sum_i g_i \mathcal{O}_i \sim \lambda_i \left(\frac{E}{\Lambda}\right)^{d_i - D} \,. \tag{1.7}$$

Therefore, at low energies, terms with $d_i > D$ are increasingly suppressed. From the perspective of relevance at low energies, that is in the infrared (IR), operators can be categorized as follows:

$d_i > D$	irrelevant
$d_i = D$	marginal
$d_i < D$	relevant

Table 1.1: EFT operators classification based on their dimensions and how they behave at low energy.

Before explaining why this framework is useful in the context of high-energy particle physics, it is helpful to clarify what exactly is meant by a *scale* in this context.

1.2 Dimensional Analysis, masses and scales

The purpose of the following subsections is to clarify the distinction between what is referred to as a *mass* and what is referred to as a *scale*, two physical quantities that can not be immediately distinguished by just looking at their units of measurements in the conventional natural units framework. To highlight their difference explicitly, we will temporarily introduce an alternative system of units.

1.2.1 Dimensional Analysis in QFT

In quantum field theory, it is common to adopt natural units. In this system, the speed of light c and the reduced Planck constant $\hbar$ (Planck constant divided by 2π) are set to unity:

$$c = \hbar = 1\,. \tag{1.8}$$

With this choice, all physical quantities can be expressed in terms of a single unit. Typically, by convention, energy is taken as the fundamental unit.

Assessing the dimension of the building blocks of the lagrangian is an important task, and it is even more relevant in the framework of EFT. In the following, we will restrict our discussion to the case of a four-dimensional spacetime, which will be the main setting for this thesis, but the reasoning we are adopting can be extended to an arbitrary number of dimensions.

Starting from the Plank relation:

$$E = h\nu\,, \tag{1.9}$$

we deduce that, in natural units, where $h = 1$, since $[\nu] = [T^{-1}]$, time has the dimension of the inverse of an energy: $[T] = E^{-1}$. By convention, this is often denoted simply as $[T] = -1$. Then, considering the relation:

$$E^2 = p^2 + m^2\,, \tag{1.10}$$

we obtain that $[L] = -1$. This directly implies that derivatives has dimension of energy: $[\partial_\mu] = 1$.

We now turn to the action which, in natural units, must be dimensionless, given that it appears as an exponent when computing path integrals. Action can be written as:

$$S = \int \mathcal{L} d^4x\,, \tag{1.11}$$

where, as usual, $\mathcal{L}$ is the lagrangian density. From the previous consideration it follows that $[\mathcal{L}] = 4$. Each term in the lagrangian must therefore have mass dimension 4. Starting from this observation, we can deduce the dimensions of the building blocks, such as scalar fields, fermion fields, and gauge fields.

Looking at the kinetic terms for bosons:

$$\mathcal{L}_{kin,\phi} = \frac{1}{2}\partial_\mu\phi\partial^\mu\phi\,, \tag{1.12}$$

we deduce that $[\phi] = 1$.

Similarly for fermions:

$$\mathcal{L}_{kin,\psi} = \overline{\psi} i \not{\partial} \psi\,, \tag{1.13}$$

we obtain $[\psi] = 3/2$.

Finally, from the field strengths definitions we can obtain the dimensions of gauge boson field:

$$[F_{\mu\nu}F^{\mu\nu}] = 4 \implies [F_{\mu\nu}] = 2 \implies [A_\mu] = 1 \tag{1.14}$$

In the context of EFT, all operators compatible with the symmetries of the theory are included in the lagrangian, regardless of their dimensions. From the previous discussion, this implies that operators must be accompanied by dimensionful constants, so that their contributions have overall mass dimension 4 in the lagrangian. As a simple example, we write down the lagrangian for a scalar boson, symmetric under the transformation $\phi \leftrightarrow -\phi$:

$$\mathcal{L} = \frac{1}{2}\partial_\mu\phi\partial^\mu\phi + \frac{1}{2}m^2\phi^2 + \frac{\lambda}{4!}\phi^4 + \sum_{n=1}^{\infty}\left(c_n\frac{\phi^{(4+2n)}}{\Lambda^{2n}} + d_n\frac{(\partial\phi)^2\phi^{2n}}{\Lambda^{2n}} + ...\right). \tag{1.15}$$

From what we said, the term ϕ^6 requires a coupling constant with mass dimensions -2, the term ϕ^8 requires a coupling constant with mass dimensions -4, and so forth. By contrast, the term ϕ^4 has a dimensionless coupling constant, while the term ϕ^2 requires a coupling constant with mass dimensions 2.

Looking at table 1.1 and recalling that we are considering a spacetime dimension of 4, we can classify field operators as follows: operators with dimensions greater than 4 are termed *irrelevant*, those with dimension exactly 4 are *marginal*, and those with dimensions less than 4 are *relevant*. These adjectives refers to the behaviour of the corresponding operators at low energy.

In 4 space-time dimensions, the Weinberg theorem implies that including in the lagrangian operators with dimension greater than 4 leads to a non-renormalizable theory, which means that the divergences arising in radiative (loop) corrections cannot be reabsorbed in a finite number of parameters. This, however, does not represent an issue in the framework of EFT, where the theory is only meant to be valid below a certain energy *scale* and is expected to break down at higher energies. Within the validity of the EFT, and given a finite experimental precision, the power counting ensures that only a finite number of EFT operators lead to observable effects. Moreover, it can be shown that renormalization does not spoil this behavior [6, 13].

1.2.2 The mass–scale distinction

We now repeat the reasoning of the previous Section, but with a slight change in the choice of units. As before, we set the speed of light to unity, $c = 1$, but this time we do not set $\hbar$ to one [14]. This new analysis will make explicit the distinction between a scale and a mass - a difference that is not apparent in natural units, where both share the same dimension, 1.

The starting point is, as before, the uncertainty principle, that this time reads as:

$$\Delta E\Delta t \geq \frac{\hbar}{2}, \quad \Delta p\Delta x \geq \frac{\hbar}{2}. \tag{1.16}$$

Again, lengths and times share the same dimensions but they are no longer expressed as the inverse

of an energy. Moreover, with this choice of units, the mass term in the lagrangian carries a factor of $\hbar$. It is therefore convenient to express masses as inverse of lengths: $\tilde{m} = m/\hbar$, so that there are no explicit factor of $\hbar$ in the classical lagrangian in position space.

Dimensional analysis now gives:

$$\begin{aligned} &[\hbar] = EL\,, && [\mathcal{L}] = EL^{-3}\,, && [\phi] = [A_\mu] = E^{\frac{1}{2}}L^{-\frac{1}{2}}, && [\psi] = E^{\frac{1}{2}}L^{-1}\,, \\ &[\partial_\mu] = L^{-1}\,, && [g] = [y] = E^{-\frac{1}{2}}L^{-\frac{1}{2}}\,, && [\lambda] = E^{-1}L^{-1}\,, \end{aligned} \tag{1.17}$$

where g, y and λ are the gauge coupling, the Yukawa coupling, and the scalar quartic coupling, respectively. According to perturbative theory, $[g^2] = [y^2] = [\lambda]$, and it can be demonstrated that performing loops calculations, each of them carry a factor of $\hbar$. In particular these factors are dimensionless in L and E units and look like $\hbar g^2/(4\pi)^2$, $\hbar y^2/(4\pi)^2$ and $\hbar\lambda/(4\pi)^2$.

With this new choice of units also couplings, and not only masses, have dimensions. We can then introduce another units system which is based on mass and coupling:

$$\tilde{M} \equiv L^{-1}\,, \qquad\qquad C \equiv E^{-\frac{1}{2}}L^{-\frac{1}{2}}\,. \tag{1.18}$$

Let us now consider the generic operator in D dimensions:

$$\frac{1}{\Lambda^{D-4}}\partial^{n_\partial}\Phi^{n_B}\psi^{n_F}\,, \tag{1.19}$$

where n_∂ is the number of derivatives, n_B is the number of boson fields ($\Phi = \phi$, A_μ), and n_F is the number of fermion fields, with the constraint that $n_\partial + n_B + \frac{3}{2}n_F = D$. The dimensionful quantity Λ defines the strength of the interaction and we will call it a *scale*. The dimensions of Λ are:

$$[\Lambda] = \frac{\tilde{M}}{C^{\frac{n_B+n_F-2}{D-4}}}\,. \tag{1.20}$$

Noting that the exponent of C in eq. (1.20) is strictly positive, it is straightforward to understand that the dimension of a scale and the one of a mass are different. Namely, these two quantities are incommensurable, and in particular a scale is the ratio between a mass and a coupling.

This difference underlines also that the two quantities are associated to different physical meanings: a mass, $E_m = \tilde{m}\hbar$ is linked to the energy at which a new degree of freedom appears. A scale, $E_\Lambda = \Lambda\hbar^{\frac{2D-n_B-n_F-6}{2(D-4)}}$ is instead associated to the energy at which a theory becomes strongly coupled if no new degrees of freedom appear.

1.3 The Standard Model and its limitations

In the following, we briefly recap the basics of the Standard Model (SM) and outline its limitations. These motivate why it can be expanded within the the Effective Field Theory framework, leading to the so-called Standard Model Effective Field Theory (SMEFT). A vast literature of textbooks [15–

20] and review articles [21–23] discusses the SM in detail, therefore the reader is referred to them for a complete and formal treatment.

1.3.1 The Standard Model

The Standard Model (SM) is the quantum field theory that describes the known fundamental particles and their interactions. The strong and electroweak forces are taken into account within this theory, while gravity is, at this day, not included. The particle content can be summarized as follows:

- Matter is made up of fermions, organized in families. Fermions within the same family share the same quantum numbers but have different masses and usually are reported from the lightest to the heaviest:
  - 3 families of up-type quarks, which are sensitive to strong, electromagnetic and weak interactions: u, c, t.
  - 3 families of down-type quarks, which are sensitive to strong, electromagnetic and weak interactions: d, s, b.
  - 3 families of leptons with no electric charge, the neutrinos, which are sensitive to just the weak interactions: ν_e, ν_μ, ν_τ.
  - 3 families of electrically charged leptons, which are sensitive to the electromagnetic and weak force: e, μ, τ.
- The force mediators are vector bosons, called also gauge bosons, and are responsible for the interactions among particles:
  - the photon γ mediates the electromagnetic interaction, which involves the electrically charged particles.
  - the $W^\pm$ and the Z bosons are responsible for the weak interactions. They are respectively involved in charged and neutral processes. Weak interactions are possible for all fermions and also involve the gauge bosons themselves.
  - the gluon g is the vector boson associated to the strong force. Strong interactions happen only between particles which carry a quantum number called "colour" (the quarks and the gluon itself).
- The Higgs boson is a scalar particle, necessary for the electroweak symmetry breaking, which is the process that explains how particles get masses

In Figure 1.1 a schematic representation of the SM particle content is provided. Formally, the particle content and interactions are encoded in the SM lagrangian, as will be detailed in the following.

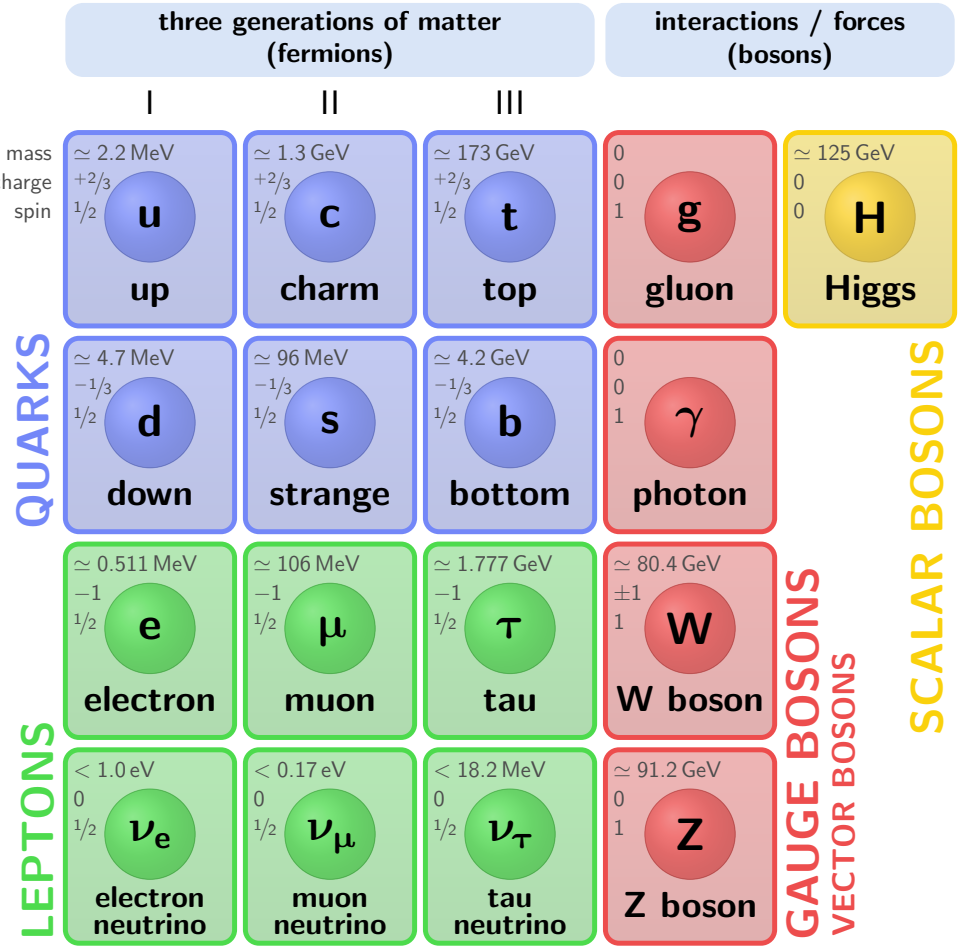


Figure 1.1: The Standard Model particle content. In the figure the particles which constitute matter and those responsible for mediating interactions are highlighted. The mass, electric charge and spin of the particles are also reported.

The SM lagrangian exhibits a space-time Poincaré symmetry and a local $SU(3)_C \times SU(2)_L \times U(1)_Y$ gauge invariance. The suffixes have the following meaning: C refers to colour, which is the quantum number associated to the strong interaction, while L and Y refer to the left-handed chirality component and to the hypercharge, respectively, which are the relevant attributes within electroweak interactions. The complete lagrangian can be written schematically as:

$$\mathcal{L}_{\rm SM} = \mathcal{L}^{\text{Yang-Mills}} + \mathcal{L}^{\rm ferm} + \mathcal{L}^{\phi} + \mathcal{L}^{\rm Yukawa} + \mathcal{L}^{\rm ghost}\,. \tag{1.21}$$

We now provide a brief explanation of each part:

$$\mathcal{L}^{\text{Yang-Mills}} = -\frac{1}{4}B_{\mu\nu}B^{\mu\nu} - \frac{1}{2}\mathrm{tr}(W_{\mu\nu}W^{\mu\nu}) - \frac{1}{2}\mathrm{tr}(G_{\mu\nu}G^{\mu\nu}) \tag{1.22}$$

is the Yang-Mills lagrangian, which contains the kinetic terms and the mutual interactions of the vector bosons. The $B_{\mu\nu}$ and $W_{\mu\nu}$ field strengths are associated with the electroweak interaction, while the $G_{\mu\nu}$ with the strong one.

$$\begin{aligned}\mathcal{L}^{\rm ferm} = {} & \bar{Q}_L^{a\alpha}\, i\gamma^\mu D_\mu^{ab,\alpha\beta} Q_L^{b\beta} + \bar{U}_R^a\, i\gamma^\mu D_\mu^{ab} U_R^b + \bar{D}_R^a\, i\gamma^\mu D_\mu^{ab} D_R^b \\ & + \bar{L}_L^\alpha\, i\gamma^\mu D_\mu^{\alpha\beta} L_L^\beta + \bar{E}_R\, i\gamma^\mu D_\mu E_R\end{aligned} \tag{1.23}$$

is the fermion lagrangian, which contains the kinetic terms of the fermions and the interactions between the fermions and the gauge bosons. The capital letters are used as a compact notation to indicate each generation of fermion, so that, for example:

$$\bar{L}_L^\alpha\, i\gamma^\mu D_\mu^{\alpha\beta} L_L^\beta = \begin{bmatrix}\ell_{e,L}^\alpha & \ell_{\mu,L}^\alpha & \ell_{\tau,L}^\alpha\end{bmatrix} \begin{bmatrix} \not{D}^{\alpha\beta} & 0 & 0 \\ 0 & \not{D}^{\alpha\beta} & 0 \\ 0 & 0 & \not{D}^{\alpha\beta}\end{bmatrix} \begin{bmatrix}\ell_{e,L}^\beta \\ \ell_{\mu,L}^\beta \\ \ell_{\tau,L}^\beta\end{bmatrix}\,.$$

The indices L and R refers to the chirality. The left-handed fermions are organized in $SU(2)$ doublets, for instance:

$$\ell_{e,L} = \begin{bmatrix}\nu_{e,L} \\ e_L\end{bmatrix}\,, \tag{1.24}$$

while the right-handed ones are $SU(2)$ singlets. The latin indices a, b run over the $SU(3)$ colour charge, while the greek indices α, β are the indices of the $SU(2)$ representation. $\gamma^\mu D_\mu$ or, equivalently, $\not{D}$ denotes the covariant derivative. D_μ is defined as:

$$D_\mu^{ab,\alpha\beta} = \partial_\mu\, \delta_a^b\, \delta_\alpha^\beta - i g_s\, G_\mu^A (T^A)_a^b\, \delta_\alpha^\beta - i g\, W_\mu^I\, \delta_a^b \left(\frac{\tau^I}{2}\right)_\alpha^\beta - i g' Y\, B_\mu\, \delta_a^b\, \delta_\alpha^\beta\,, \tag{1.25}$$

where G_μ^A, W_μ and B_μ represent the bosons acting as force mediators. For completeness, we report in Table 1.2 the quantum numbers associated to the five fermion fields q_L, ℓ_L, u_R, d_R, e_R of the first generation (the ones of the other two generations are the same).

Field	$SU(3)_C \times SU(2)_L \times U(1)_Y$	Y	Q
q_L	$(\mathbf{3}, \mathbf{2}, 1/6)$	$1/6$	$\begin{cases} u_L : & 2/3 \\ d_L : & -1/3 \end{cases}$
u_R	$(\mathbf{3}, \mathbf{1}, 2/3)$	$2/3$	$2/3$
d_R	$(\mathbf{3}, \mathbf{1}, -1/3)$	$-1/3$	$-1/3$
ℓ_L	$(\mathbf{1}, \mathbf{2}, -1/2)$	$-1/2$	$\begin{cases} \nu_L : & 0 \\ e_L : & -1 \end{cases}$
e_R	$(\mathbf{1}, \mathbf{1}, -1)$	-1	-1

Table 1.2: Quantum number of the fermion fields in the Standard Model (one generation).

$$\mathcal{L}^{\phi} = |D_\mu \phi|^2 + \mu^2 \phi^\dagger \phi - \lambda (\phi^\dagger \phi)^2 \tag{1.26}$$

is the Higgs lagrangian, which contains the kinetic term for the Higgs boson, its coupling with the gauge bosons and its self-coupling. After the electroweak symmetry breaking this term gives masses to the $W^\pm$ and Z bosons and to the Higgs boson itself.

$$\mathcal{L}^{\text{Yukawa}} = (Y_U)(\overline{Q}_L)\tilde{\phi}(U_R) + (Y_D)(\overline{D}_L)\phi(D_R) + (Y_E)(\overline{L}_L)\tilde{\phi}(E_R)\,, \tag{1.27}$$

is the Yukawa lagrangian. Y_U, Y_D and Y_E are 3×3 matrices containing the Yukawa couplings, which, after the electroweak symmetry breaking, determine the fermion masses. $\tilde{\phi}$ is defined as: $\tilde{\phi} = i\sigma_2 \phi$. Finally,

$$\mathcal{L}^{\text{ghost}} = \partial_\mu \overline{c}^a \partial^\mu c^a + g f^{abc} (\partial^\mu \overline{c}^a) A^b_\mu c^c\,, \tag{1.28}$$

is the ghost lagrangian, needed to maintain the consistency of the path integral formulation and necessary to preserve unitarity and gauge invariance in non-abelian gauge theories. A^b_μ refers generically to the gauge bosons of non-abelian gauge theories.

Starting from this Lagrangian, one can compute predictions for experimentally measurable quantities, known as observables, which allow testing the validity of the theory by comparison with data. Among the most relevant observables are cross-sections, considered either as total cross-sections or as differential distributions with respect to specific kinematic variables, such as, for example, the invariant mass, the transverse momentum, scattering angles, ... Throughout decades of experiments, the SM has been extensively and rigorously tested, yielding remarkable successes, most notably the theoretical prediction of the existence of the Higgs boson in 1964 [24, 25], almost 50 years before its experimental discovery in 2012 by the ATLAS [26] and CMS [27] experiments at the LHC. Nonetheless, it is now clear that the SM cannot be considered the ultimate theory of Nature, as numerous experimental observations remain unexplained. Over the years, several theories extending the SM have been proposed to address these open issues. This line of research, which aims to expand the SM in order to explain such observations, is generally referred to as Beyond Standard Model (BSM) physics.

1.3.2 The Standard Model limitations

In this subsection we will briefly list some experimental evidence that are not accounted for by the SM, along with some of its conceptual and phenomenological limitations. Our purpose is neither to examine each case in detail nor to provide an exhaustive list, but rather to support the statement made in the previous Section that "the SM cannot be considered the ultimate theory of Nature".

- **Naturalness problem** [28, 29]: this issue is strictly related with the gigantic separation of scales between the fundamental interactions described by the SM (strong and electro-weak) and gravity.[1] The existence of a large UV scale such as $m_{\rm P}$ poses a serious theoretical problem for the Higgs sector of the SM. Indeed, the Higgs mass, or, equivalently, its vacuum expectation value (VEV), which is the only dimensionful parameter in the SM, is expected to receive quantum corrections proportional to any UV scale in the theory. Naive dimensional analysis would then suggest the Higgs mass to be as large as the scale of gravity, or any other existing UV scale. However, the observed value $m_H \sim 125$ GeV, is extremely smaller than the scale of gravity, suggesting that, in absence of any particular mechanism, this would require gigantic cancellations in the quantum corrections to such parameter. This large cancellation is usually called *fine-tuning*. Given that the typical level of fine-tuning observed in all other physical systems, from atomic phenomena, to the sun to moon size ratio that allows for total solar eclipse, never exceeds the % level, a fine-tuning at the level of $m_H^2/m_{\rm P}^2 \sim 10^{-34}$ is considered theoretically unacceptable, or, at least, it is believed to require an explanation. The main idea of Beyond the SM (BSM) theories addressing the Naturalness problem is to try to screen the quantum corrections to the Higgs boson mass from large UV scales through new physics at a scale $\Lambda_{\rm NP}$ not far above the EW scale. In this way, the large fine-tuning mentioned above, is replaced by something of the order $m_H^2/\Lambda_{\rm UV}^2$. It is not hard to see that if $\Lambda_{\rm NP}$ is in the TeV region, then the fine-tuning is in the percent region, which would represent an "acceptable" numerical accident.

  Examples of such BSM theories are Supersymmetry, where the aforementioned screening of the Higgs mass from large quantum corrections is realized through an enhanced symmetry in the UV, or Composite Higgs, where the Higgs boson is a composite particle, similar to the pion in QCD, arising in the infrared (IR), and simply does not exist as a fundamental particle in the UV.

- **Neutrino masses** [30]: it is verified through experimental measurements that neutrinos have a non-zero mass, which is not accounted for by the SM. The absolute value of these masses is unknown (only an upper bound of around 0.1 eV exists [31]), and only mass differences among different neutrino species have been measured until now. There are ways to add neutrino masses to the SM, both for Dirac and Majorana neutrinos, and this constitutes one interesting branch of investigation of BSM physics.

[1] The QCD scale is about $\Lambda_{\rm QCD} \sim 200\,{\rm MeV}$, the EW scale is set by the Fermi constant, and is around 100 GeV, while the scale of gravity, as set by the Newton constant (or, equivalently, the Planck mass), is around $m_{\rm P} \sim 10^{19}$ GeV.

- **Gravitational interactions** [32]: a renormalizable theory of quantum gravity has not been proposed yet and the SM cannot account for gravitational phenomena. The known theory of quantum gravity is only an effective field theory and no description is known which provides a full unification of all four known forces. The hope is that string theory could provide such candidate theory, but its link to phenomenology and to the SM is still under investigation.

- **Matter Anti-Matter asymmetry** [33]: astronomical observations highlight that there is far more matter than anti-matter in the observable Universe. This is something that the SM cannot explain because the theory does not distinguish between matter and anti-matter and it is perfectly symmetric under their interchange.

- **Strong CP problem** [34]: when we introduced the SM, we only mentioned continuous symmetries. Nonetheless, there are other three discrete symmetries C (charge conjugation), P (parity) and T (time-reversal), which are all broken in the SM.[2] In particular, weak interactions break both P and CP (or T) by an order-one amount, parametrized by a phase in the Cabibbo–Kobayashi–Maskawa (CKM) matrix.. Also strong interactions could in principle break CP. However, from the current measured upper bound to the neutron electric dipole moment ($d_n < 0.0 \pm 1.1 \times 10^{-26} \mathrm{e} \cdot \mathrm{cm}$), it turns out that this breaking, if it is there, is unexpectedly small. To explain such a small parameter, theorists have considered axion models, where such parameter is promoted to a dynamical field. Axion physics is another interesting and active branch of BSM research.

- **Dark matter** [35]: astrophysics observations have made clear that an important part of Universe ($\sim 26\%$) is made up by a kind of matter which is different from the ordinary one and does not emit electromagnetic radiation. The existence of this so called dark matter is necessary, for instance, to explain the rotation speed of stars around the center of the galaxy. The SM particle content does not include any particle that could act as such dark matter candidate.

All these unanswered questions are signals that the SM does not provide a complete description of Nature. A possible solution is to think that the theory could be consistently extended to account for phenomena it cannot explain in its present formulation. Such extensions include new particles and phenomena that are generically referred to as New Physics (NP).

Signals of NP have been searched for in experiments for decades. Currently, our best chance to observe them is at colliders. In particular, the Large Hadron Collider (LHC) reaches the highest energies ever achieved in a laboratory. This is promising for NP searches, since particles not yet observed might exist at energies beyond our current reach.

[2] Notice that the CPT theorem implies that the combination of all these three symmetries is always an exact symmetry.

1.4 The SMEFT as a tool for NP search

In this last part of the Introduction, we describe how the EFT framework can be used to extend the Standard Model (SM), giving rise to the so-called Standard Model Effective Field Theory (SMEFT), and employed as a tool for searching New Physics (NP). Moreover, we highlight why this framework is promising in the context of the Large Hadron Collider (LHC). Finally, we introduce the channel that will be the focus of the first part of this thesis: the Drell-Yan process.

1.4.1 SMEFT

The search for NP can be pursued through direct or indirect searches. As the terms suggest, direct searches aim to directly observe in experiments new particles that could account for unexplained phenomena. The most famous example is the discovery of the Higgs boson in 2012 at the LHC [26, 27], which confirmed the prediction of the Standard Model in explaining how particles get mass after electroweak symmetry breaking. The strength of direct searches is their conclusive power when successful, but their drawback is evident: if no new particle is observed, nothing can be said about the existence of NP. Indirect searches, on the other hand, aims to detected signal of NP by looking for small deviations from the SM expectations in measurable quantities. Given the current experimental status, with no new discoveries since the Higgs boson, indirect searches appear, at present, to be the most promising approach to look for NP.

Another distinction is typically made between model-dependent and model-independent searches. The former focus on specific models, where new, well-defined particles with specified properties (such as mass, spin, charge, couplings, ...), are introduced to address unexplained phenomena. Some examples of these models are: supersymmetry [36], which can solve the hierarchy problem and provide candidates for dark matter, axion physics [37], to solve the strong CP problem and again giving a candidate for dark matter, seesaw models [38] for neutrino masses, introducing right-handed neutrino to give mass to neutrinos and explain why they are so small, ... The latter, instead, aim to capture the broadest possible underlying NP that could account for a given deviation from SM predictions, without committing to any specific NP model. The most general approach to carry out an indirect, model-independent analysis is to combine the concepts of EFT with the SM framework, giving rise to what is known as the SMEFT [39, 40]. In the SMEFT framework, the SM is considered as the low energy manifestation of some more fundamental and still unknown UV theory. Schematically, this is represented in Figure 1.2. Following the concepts outlined in the previous Sections, the SM lagrangian can be extended by adding all the operators, made up by SM fields, which respect the symmetries of the theory (namely Poincaré invariance and invariance under the gauge group $SU(3)_C \times SU(2)_L \times U(1)_Y$), regardless of their mass dimensions. These operators have mass dimensions equal or higher than 5 and must be accompanied by the appropriate inverse power of an energy scale, conventionally denoted by Λ, which represents the scale at which the EFT description is expected to break down. Operators with odd mass dimension, typically violate baryon and/or lepton number (e.g. the Weinberg operator for neutrino masses, which is the only

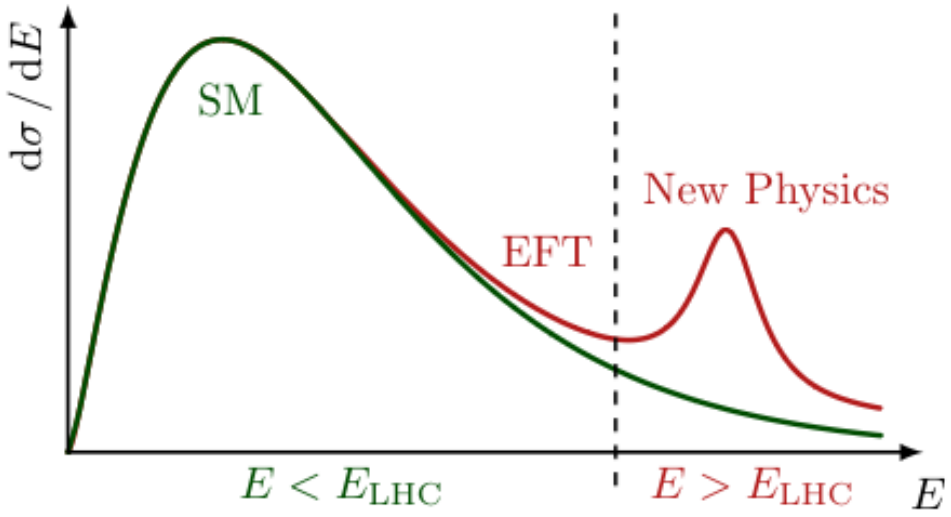


Figure 1.2: A schematic representation of the SMEFT. The SM appears as the low energy manifestation of a UV theory, and NP is present at energy higher than the LHC reach.

dimension 5 operator). For this reason, given that B and L violation are very tightly constrained experimentally, many SMEFT analyses do not include them. Considering only even mass dimension operators, *power counting* implies that the leading contributions beyond the SM come from dimension-6 operators, suppressed by $1/\Lambda^2$, followed by those of dimension 8, suppressed by $1/\Lambda^4$, and so on. Therefore, at the energies currently accessible in experiments, the dimension-6 sector is expected to dominate, except for specific observables in which the dimension-6 contributions are accidentally suppressed, typically due to selection rules.

In Reference [41] a standard basis for dimension-6 operators, called the Warsaw basis, has been defined and is typically used as a reference to classify different kind of operators. We report in Tables 1.3 and 1.4 the classes of operators and their definitions as listed in Ref. [41]. In total, there are 59 dimension-6 independent operators which do not violate the baryon number conservation in the flavor-universal case. When the complete three-generation flavor structure is included, allowing for generic flavor couplings, this number rises to 2499. When considering higher dimensions operators, this number rises drastically. For example, dimension-8 operators are more than ten thousand.

The procedure to conduct a SMEFT analysis is well established. The first step is to select a process that is sensitive to potential NP. This can occur when a SM measurement is so precise that even small NP contributions could be detected. Then, the relevant SMEFT operators contributing to the chosen process are added to the SM lagrangian. This is done considering both the hierarchy among operators (with dimension-6 operators typically being more important than dimension-8 ones, and so on) and by the characteristics of the measurement itself. For instance, cancellations may enhance operators that are usually suppressed, or in certain regions of phase space such operators

X^3		φ^6 and $\varphi^4 D^2$		$\psi^2\varphi^3$	
Q_G	$f^{ABC} G_\mu^{A\nu} G_\nu^{B\rho} G_\rho^{C\mu}$	Q_φ	$(\varphi^\dagger\varphi)^3$	$Q_{e\varphi}$	$(\varphi^\dagger\varphi)(\bar{l}_p e_r \varphi)$
$Q_{\widetilde{G}}$	$f^{ABC} \widetilde{G}_\mu^{A\nu} G_\nu^{B\rho} G_\rho^{C\mu}$	$Q_{\varphi\Box}$	$(\varphi^\dagger\varphi)\Box(\varphi^\dagger\varphi)$	$Q_{u\varphi}$	$(\varphi^\dagger\varphi)(\bar{q}_p u_r \widetilde{\varphi})$
Q_W	$\varepsilon^{IJK} W_\mu^{I\nu} W_\nu^{J\rho} W_\rho^{K\mu}$	$Q_{\varphi D}$	$(\varphi^\dagger D^\mu \varphi)^\star (\varphi^\dagger D_\mu \varphi)$	$Q_{d\varphi}$	$(\varphi^\dagger\varphi)(\bar{q}_p d_r \varphi)$
$Q_{\widetilde{W}}$	$\varepsilon^{IJK} \widetilde{W}_\mu^{I\nu} W_\nu^{J\rho} W_\rho^{K\mu}$				
$X^2\varphi^2$		$\psi^2 X\varphi$		$\psi^2\varphi^2 D$	
$Q_{\varphi G}$	$\varphi^\dagger\varphi\, G_{\mu\nu}^A G^{A\mu\nu}$	Q_{eW}	$(\bar{l}_p\sigma^{\mu\nu} e_r)\tau^I \varphi W_{\mu\nu}^I$	$Q_{\varphi l}^{(1)}$	$(\varphi^\dagger i\overleftrightarrow{D}_\mu \varphi)(\bar{l}_p\gamma^\mu l_r)$
$Q_{\varphi\widetilde{G}}$	$\varphi^\dagger\varphi\, \widetilde{G}_{\mu\nu}^A G^{A\mu\nu}$	Q_{eB}	$(\bar{l}_p\sigma^{\mu\nu} e_r)\varphi B_{\mu\nu}$	$Q_{\varphi l}^{(3)}$	$(\varphi^\dagger i\overleftrightarrow{D}^I_\mu \varphi)(\bar{l}_p\tau^I\gamma^\mu l_r)$
$Q_{\varphi W}$	$\varphi^\dagger\varphi\, W_{\mu\nu}^I W^{I\mu\nu}$	Q_{uG}	$(\bar{q}_p\sigma^{\mu\nu} T^A u_r)\widetilde{\varphi} G_{\mu\nu}^A$	$Q_{\varphi e}$	$(\varphi^\dagger i\overleftrightarrow{D}_\mu \varphi)(\bar{e}_p\gamma^\mu e_r)$
$Q_{\varphi\widetilde{W}}$	$\varphi^\dagger\varphi\, \widetilde{W}_{\mu\nu}^I W^{I\mu\nu}$	Q_{uW}	$(\bar{q}_p\sigma^{\mu\nu} u_r)\tau^I \widetilde{\varphi} W_{\mu\nu}^I$	$Q_{\varphi q}^{(1)}$	$(\varphi^\dagger i\overleftrightarrow{D}_\mu \varphi)(\bar{q}_p\gamma^\mu q_r)$
$Q_{\varphi B}$	$\varphi^\dagger\varphi\, B_{\mu\nu} B^{\mu\nu}$	Q_{uB}	$(\bar{q}_p\sigma^{\mu\nu} u_r)\widetilde{\varphi} B_{\mu\nu}$	$Q_{\varphi q}^{(3)}$	$(\varphi^\dagger i\overleftrightarrow{D}^I_\mu \varphi)(\bar{q}_p\tau^I\gamma^\mu q_r)$
$Q_{\varphi\widetilde{B}}$	$\varphi^\dagger\varphi\, \widetilde{B}_{\mu\nu} B^{\mu\nu}$	Q_{dG}	$(\bar{q}_p\sigma^{\mu\nu} T^A d_r)\varphi G_{\mu\nu}^A$	$Q_{\varphi u}$	$(\varphi^\dagger i\overleftrightarrow{D}_\mu \varphi)(\bar{u}_p\gamma^\mu u_r)$
$Q_{\varphi WB}$	$\varphi^\dagger\tau^I\varphi\, W_{\mu\nu}^I B^{\mu\nu}$	Q_{dW}	$(\bar{q}_p\sigma^{\mu\nu} d_r)\tau^I \varphi W_{\mu\nu}^I$	$Q_{\varphi d}$	$(\varphi^\dagger i\overleftrightarrow{D}_\mu \varphi)(\bar{d}_p\gamma^\mu d_r)$
$Q_{\varphi\widetilde{W}B}$	$\varphi^\dagger\tau^I\varphi\, \widetilde{W}_{\mu\nu}^I B^{\mu\nu}$	Q_{dB}	$(\bar{q}_p\sigma^{\mu\nu} d_r)\varphi B_{\mu\nu}$	$Q_{\varphi q}$	$i(\varphi^\dagger \overleftrightarrow{D}_\mu \varphi)(\bar{u}_p\gamma^\mu d_r)$

Table 1.3: Dimension-six operators other than the four-fermion ones.

can become particularly important. Once the operators are specified, the SMEFT predictions for the observable of interest are calculated and compared with experimental data (when available under the same assumptions) or with pseudo-data generated under the SM-only hypothesis. Typically, these comparisons expoit statistical tools such as χ^2 minimization and likelihood ratio test, to derive Confidence Intervals on the Wilson coefficients of the included SMEFT operators. These regions for the Wilson coefficients can then be mapped onto possible new physics scenarios, constraining both the masses and the coupling of specific UV models. This final step is usually referred to as the "bottom-up approach".

In the following we explore in more details how SMEFT analyses are carried out at the LHC.

1.4.2 The SMEFT in the LHC high-p_T precision program

As discussed previously, the LHC is the experiment which reaches the highest energies ever observed, and therefore represents our best chance to discover NP. A natural question that arises is whether it is possible to apply a SMEFT analysis in the LHC context and the answer is not straightforward.

The main point is that the LHC is a hadron collider, where protons collide with each other to produce new particles. To appreciate the implications of this feature, it is useful to compare it with a lepton collider. In the past, the Large Electron–Positron (LEP) collider was the experiment exploited to study particle physics through electron–positron collisions. Each type of collider comes

$(\overline{L}L)(\overline{L}L)$		$(\overline{R}R)(\overline{R}R)$		$(\overline{L}L)(\overline{R}R)$	
Q_{ll}	$(\bar{l}_p\gamma_\mu l_r)(\bar{l}_s\gamma^\mu l_t)$	Q_{ee}	$(\bar{e}_p\gamma_\mu e_r)(\bar{e}_s\gamma^\mu e_t)$	Q_{le}	$(\bar{l}_p\gamma_\mu l_r)(\bar{e}_s\gamma^\mu e_t)$
$Q_{qq}^{(1)}$	$(\bar{q}_p\gamma_\mu q_r)(\bar{q}_s\gamma^\mu q_t)$	Q_{uu}	$(\bar{u}_p\gamma_\mu u_r)(\bar{u}_s\gamma^\mu u_t)$	Q_{lu}	$(\bar{l}_p\gamma_\mu l_r)(\bar{u}_s\gamma^\mu u_t)$
$Q_{qq}^{(3)}$	$(\bar{q}_p\gamma_\mu\tau^I q_r)(\bar{q}_s\gamma^\mu\tau^I q_t)$	Q_{dd}	$(\bar{d}_p\gamma_\mu d_r)(\bar{d}_s\gamma^\mu d_t)$	Q_{ld}	$(\bar{l}_p\gamma_\mu l_r)(\bar{d}_s\gamma^\mu d_t)$
$Q_{lq}^{(1)}$	$(\bar{l}_p\gamma_\mu l_r)(\bar{q}_s\gamma^\mu q_t)$	Q_{eu}	$(\bar{e}_p\gamma_\mu e_r)(\bar{u}_s\gamma^\mu u_t)$	Q_{qe}	$(\bar{q}_p\gamma_\mu q_r)(\bar{e}_s\gamma^\mu e_t)$
$Q_{lq}^{(3)}$	$(\bar{l}_p\gamma_\mu\tau^I l_r)(\bar{q}_s\gamma^\mu\tau^I q_t)$	Q_{ed}	$(\bar{e}_p\gamma_\mu e_r)(\bar{d}_s\gamma^\mu d_t)$	$Q_{qu}^{(1)}$	$(\bar{q}_p\gamma_\mu q_r)(\bar{u}_s\gamma^\mu u_t)$
		$Q_{ud}^{(1)}$	$(\bar{u}_p\gamma_\mu u_r)(\bar{d}_s\gamma^\mu d_t)$	$Q_{qu}^{(8)}$	$(\bar{q}_p\gamma_\mu T^A q_r)(\bar{u}_s\gamma^\mu T^A u_t)$
		$Q_{ud}^{(8)}$	$(\bar{u}_p\gamma_\mu T^A u_r)(\bar{d}_s\gamma^\mu T^A d_t)$	$Q_{qd}^{(1)}$	$(\bar{q}_p\gamma_\mu q_r)(\bar{d}_s\gamma^\mu d_t)$
				$Q_{qd}^{(8)}$	$(\bar{q}_p\gamma_\mu T^A q_r)(\bar{d}_s\gamma^\mu T^A d_t)$
$(\overline{L}R)(\overline{R}L)$ and $(\overline{L}R)(\overline{L}R)$		B-violating			
Q_{ledq}	$(\bar{l}_p^j e_r)(\bar{d}_s q_t^j)$	Q_{duq}	$\varepsilon^{\alpha\beta\gamma}\varepsilon_{jk}\,[(d_p^\alpha)^T C u_r^\beta][(q_s^j)^T C l_t^k]$		
$Q_{quqd}^{(1)}$	$(\bar{q}_p^j u_r)\varepsilon_{jk}(\bar{q}_s^k d_t)$	Q_{qqu}	$\varepsilon^{\alpha\beta\gamma}\varepsilon_{jk}\,[(q_p^{\alpha j})^T C q_r^{\beta k}][(u_s^\gamma)^T C e_t]$		
$Q_{quqd}^{(8)}$	$(\bar{q}_p^j T^A u_r)\varepsilon_{jk}(\bar{q}_s^k T^A d_t)$	Q_{qqq}	$\varepsilon^{\alpha\beta\gamma}\varepsilon_{jn}\varepsilon_{km}\,[(q_p^{\alpha j})^T C q_r^{\beta k}][(q_s^{\gamma m})^T C l_t^n]$		
$Q_{lequ}^{(1)}$	$(\bar{l}_p^j e_r)\varepsilon_{jk}(\bar{q}_s^k u_t)$	Q_{duu}	$\varepsilon^{\alpha\beta\gamma}\,[(d_p^\alpha)^T C u_r^\beta][(u_s^\gamma)^T C e_t]$		
$Q_{lequ}^{(3)}$	$(\bar{l}_p^j\sigma_{\mu\nu} e_r)\varepsilon_{jk}(\bar{q}_s^k\sigma^{\mu\nu} u_t)$				

Table 1.4: Four-fermion operators.

with its own advantages and disadvantages. The main strength of a hadron collider like the LHC is its ability to reach much higher energies. On the other hand, a lepton collider such as LEP enables far more precise measurements, given the cleaner nature of lepton collisions compared to hadron ones. Below we provide a schematic comparison between LEP and the LHC, highlighting the aspects just discussed:

	on Z pole		off Z pole	
	LEP	LHC	LEP	LHC
Energy	100 GeV	1 TeV	100 GeV	1 TeV
Accuracy	$\sim$‰	$\sim 10\%$	$\sim 1\%$	$\sim 10\%$

In the previous Section, we highlighted how, in the SMEFT context, the contributions from higher dimensional operators, which represent NP, manifest as small deviations from SM expectations. For this reason, high precision is essential to distinguish genuine NP signals from fluctuations due to uncertainties. At first sight, this makes LEP appear as a more suitable environment for a SMEFT analysis, thanks to its superior precision. Nonetheless, this consideration does not hold for any class of SMEFT operators. For example, dimension-6 operators can contribute to the 2 in 2

amplitude with different scalings, namely:

$$\frac{v^2}{\Lambda^2}, \quad \frac{v\sqrt{s}}{\Lambda^2}, \quad \frac{s}{\Lambda^2}, \tag{1.29}$$

where v denotes the vacuum expectation value (VEV) and $\sqrt{s}$ the center of mass (c.o.m.) energy of the process. The contributions proportional to v arise from operators containing Higgs fields which, after electroweak symmetry breaking, are replaced by the VEV. In contrast, the energy-growing contributions proportional to s originate from operators containing derivatives, which translate into factors of the external momenta in the scattering amplitude.

In general, SMEFT operators can contribute through appropriate combinations of v^2 and s. If we target operators whose contributions grows with energy, such as the dimension-6 ones that contribute with s/Λ^2 to the amplitude, their effect becomes significantly enhanced at the LHC compared to LEP, thanks to the higher energy reached by the hadron collider. Approximately, the enhancement factor is given by:

$$\frac{E^2_{\text{LHC}}}{E^2_{\text{LEP}}} \sim 100, \tag{1.30}$$

which is large enough to compensate for the reduced precision at the LHC.

To summarize, SMEFT operators can be broadly divided into two categories: those whose contributions to observables depend only on the Higgs VEV, and those whose contributions scales with the energy of the process. A lepton collider, such as LEP, is ideal to probe the former class, thanks to the intrinsic high precision it achieves. On the other hand, a hadron collider, such as the LHC, is optimal to study the latter class, whose effects are strongly enhanced by the higher energy reached. This observation motivated the development of a dedicated branch of study, referred to as the "LHC high-p_T precision program" [42, 43].

1.5 A benchmark channel: the Drell-Yan process

Insights into New Physics (NP) can be pursued through different channels, motivated by different reasons. Within the LHC high-p_T precision program, some remarkable examples are: di-jets [44–47], di-bosons [48–55], di-quarks [45], and di-tops [56–59], and even four-tops [60]. Among these, the production of a lepton and an anti-lepton in the final state, referred to as Drell-Yan (DY) process, stands out for its simplicity. The final state can be either electrically neutral or charged, namely:

$$\text{Neutral DY:} \quad pp \longrightarrow \ell^+\ell^- + X, \tag{1.31}$$

$$\text{Positively charged DY:} \quad pp \longrightarrow \ell^+\nu_\ell - + X, \tag{1.32}$$

$$\text{Negatively charged DY:} \quad pp \longrightarrow \ell^-\overline{\nu}_\ell - + X, \tag{1.33}$$

where ℓ denotes a generic lepton ($\ell = e, \mu, \tau$), ν_ℓ the corresponding neutrino and X the additional particles produced by the process which are not the targeted leptons.

The DY process, both in its neutral and charged channels, is arguably the easiest channel to

1. $\varphi^4 D^2$		**2.** $\psi^2\varphi^3$	
$Q_{\varphi\Box}$	$(\varphi^\dagger\varphi)\Box(\varphi^\dagger\varphi)$	$Q_{e\varphi}$	$(\varphi^\dagger\varphi)(\bar{\ell}_p e_r\varphi)$
$Q_{\varphi D}$	$(\varphi^\dagger D^\mu\varphi)^*(\varphi^\dagger D_\mu\varphi)$	$Q_{u\varphi}$	$(\varphi^\dagger\varphi)(\bar{q}_p u_r\tilde{\varphi})$
		$Q_{d\varphi}$	$(\varphi^\dagger\varphi)(\bar{q}_p d_r\varphi)$
3. $\psi^2 X\varphi$		**4.** $\psi^2\varphi^2 D$	
Q_{eW}	$(\bar{\ell}_p\sigma^{\mu\nu}e_r)\tau^I\varphi W^I_{\mu\nu}$	$Q^{(1)}_{\varphi l}$	$(\varphi^\dagger i\overleftrightarrow{D}_\mu\varphi)(\bar{\ell}_p\gamma^\mu\ell_r)$
Q_{eB}	$(\bar{\ell}_p\sigma^{\mu\nu}e_r)\varphi B_{\mu\nu}$	$Q^{(3)}_{\varphi l}$	$(\varphi^\dagger i\overleftrightarrow{D}^I_\mu\varphi)(\bar{\ell}_p\tau^I\gamma^\mu\ell_r)$
Q_{uW}	$(\bar{q}_p\sigma^{\mu\nu}u_r)\,\tau^I\tilde{\varphi}\,W^I_{\mu\nu}$	$Q_{\varphi e}$	$(\varphi^\dagger i\overleftrightarrow{D}_\mu\varphi)(\bar{e}_p\gamma^\mu e_r)$
Q_{uB}	$(\bar{q}_p\sigma^{\mu\nu}u_r)\,\tilde{\varphi}\,B_{\mu\nu}$	$Q^{(1)}_{\varphi q}$	$(\varphi^\dagger i\overleftrightarrow{D}_\mu\varphi)(\bar{q}_p\gamma^\mu q_r)$
Q_{dW}	$(\bar{q}_p\sigma^{\mu\nu}d_r)\,\tau^I\varphi\,W^I_{\mu\nu}$	$Q^{(3)}_{\varphi q}$	$(\varphi^\dagger i\overleftrightarrow{D}^I_\mu\varphi)(\bar{q}_p\tau^I\gamma^\mu q_r)$
Q_{dB}	$(\bar{q}_p\sigma^{\mu\nu}d_r)\,\varphi\,B_{\mu\nu}$	$Q_{\varphi u}$	$(\varphi^\dagger i\overleftrightarrow{D}_\mu\varphi)(\bar{u}_p\gamma^\mu u_r)$
		$Q_{\varphi d}$	$(\varphi^\dagger i\overleftrightarrow{D}_\mu\varphi)(\bar{d}_p\gamma^\mu d_r)$
		$Q_{\varphi ud}$	$i(\tilde{\varphi}^\dagger D_\mu\varphi)(\bar{u}_p\gamma^\mu d_r)$
5. $(\bar{L}L)(\bar{L}L)$		**6.** $(\bar{R}R)(\bar{R}R)$	
$Q^{(1)}_{lq}$	$(\bar{\ell}_p\gamma_\mu\ell_r)(\bar{q}_s\gamma^\mu q_t)$	Q_{eu}	$(\bar{e}_p\gamma_\mu e_r)(\bar{u}_s\gamma^\mu u_t)$
$Q^{(3)}_{lq}$	$(\bar{\ell}_p\gamma_\mu\tau^I\ell_r)(\bar{q}_s\gamma^\mu\tau^I q_t)$	Q_{ed}	$(\bar{e}_p\gamma_\mu e_r)(\bar{d}_s\gamma^\mu d_t)$
7. $(\bar{L}L)(\bar{R}R)$		**8.** $(\bar{L}R)(\bar{R}L)$ **and** $(\bar{L}R)(\bar{L}R)$	
Q_{lu}	$(\bar{\ell}_p\gamma_\mu\ell_r)(\bar{u}_s\gamma^\mu u_t)$	Q_{ledq}	$(\bar{\ell}^j_p e_r)(\bar{d}_s q^j_t)$
Q_{ld}	$(\bar{\ell}_p\gamma_\mu\ell_r)(\bar{d}_s\gamma^\mu d_t)$	$Q^{(1)}_{lequ}$	$(\bar{\ell}^j_p e_r)\epsilon_{jk}(\bar{q}^k_s u_t)$
Q_{qe}	$(\bar{q}_p\gamma_\mu q_r)(\bar{e}_s\gamma^\mu e_t)$	$Q^{(3)}_{lequ}$	$(\bar{\ell}^j_p\sigma_{\mu\nu}e_r)\epsilon_{jk}(\bar{q}^k_s\sigma^{\mu\nu}u_t)$

Table 1.5: Dimension-6 operators contributing to the Drell-Yan process in the Warsaw basis. The notation is the same as in Ref. [41]. The operators are grouped according to their field content and the first column indicates the operator's name.

study at the LHC, from both a theoretical and experimental point of view. On the theoretical side, the simple structure of the interaction has allowed calculations at high perturbative orders, while, experimentally, charged leptons are easy to detect, they are not strongly affected by QCD backgrounds, and the cross-section is large enough to collect significant statistics. As a result, the DY process has been widely explored at the LHC. Measurements are available in different binning schemes, both single and multi differential, and at various center of mass (c.o.m.) energies, ranking among the most precise performed.

For all the reasons listed above, the DY process represent an important channel to search for NP using the SMEFT tools introduced in the previous Sections. During the first part of this thesis, we will focus on the leading contributions from the SMEFT, namely the dimension-6 operators, relevant to the DY.

To provide a complete overview, we report in Table 1.5 all the dimension-6 operators contributing to this process, written in the Warsaw basis notation. Below, we detail their behaviour focusing on two main features. The first is whether they contribute at the linear or at the quadratic level

(i.e. whether they interfere with the SM amplitude or not), the second is their behavior with respect to the energy of the process, i.e. whether they lead to amplitudes that grow with energy or not.

- **1.** $\varphi^4 D^2$
  The operators in this class modify the Higgs and gauge boson propagators on-shell, generating corrections to their kinetic terms suppressed by v^2/Λ^2. The Drell-Yan amplitude involving these operators are proportional to v^2/Λ^2, leading to an interference with the SM amplitude proportional to v^2/Λ^2, and to a contribution to the amplitude squared proportional to v^4/Λ^4. These contributions do not grow with energy, and are therefore better constrained in very precise experiments at the threshold for on-shell production of SM bosons (such as LEP or future lepton colliders).

- **2.** $\psi^2\varphi^3$
  These operators are v^2/Λ^2 corrections to the Yukawa couplings of the SM. Obviously, their flavor structure may determine stronger or weaker bounds from flavor physics. However, their contribution to the DY amplitude is generally suppressed by the fermion mass and the additional factor v^2/Λ^2. For the light fermions, which give the largest contribution to the DY process, the contribution of such operators is therefore negligible.

- **3.** $\psi^2 X\varphi$
  This class of operators contain dipole interactions between fermions and gauge bosons. Some combinations of these operators, such as those leading to electron dipole moments, are very strongly constrained by low-energy experiments. However, other combinations can be effectively constrained from DY measurements. Indeed, the contribution of these chirality-breaking operators to the DY amplitude is proportional to $v\sqrt{s}/\Lambda^2$. Due to the chirality structure of the fermionic tensor, the DY amplitude generated by these operators does not interfere with the SM amplitude. Therefore, their contribution to the amplitude squared scales as $v^2 s/\Lambda^4$, growing linearly with s.

- **4.** $\psi^2\varphi^2 D$
  These operators have the form of product of a fermion and a Higgs current. They lead to a v^2/Λ^2 corrections to the interactions of fermions with gauge bosons and the Higgs. Their contribution to the DY amplitude does interfere with the SM ones, but does not grow with energy, generating terms proportional to v^2/Λ^2 and v^4/Λ^4 in the amplitude squared. As for the operators of class 1, these contributions are better constrained in precise experiments at the threshold for on-shell production of SM bosons.

- **5.** $(\bar{L}L)(\bar{L}L)$**, 6.** $(\bar{R}R)(\bar{R}R)$**, 7.** $(\bar{L}L)(\bar{R}R)$ The seven operators in these classes correspond to the product of two fermion currents. They are the only operators that lead to contributions to the DY amplitude that are proportional to s/Λ^2, and, at the same time, interfere with the SM amplitudes. Therefore, they generate contributions to the squared amplitude proportional

to s/Λ^2 and s^2/Λ^4. These operators have been extensively studied in the context of the DY process at the LHC [1, 43, 61, 62].

- **8.** $(\bar{L}R)(\bar{R}L)$ **and** $(\bar{L}R)(\bar{L}R)$
  The last class of operators contains chirality-breaking interactions of two scalar or two tensor fermion bilinears. We will refer to them as "scalar" and "tensor" four-fermion operators, respectively. Their contribution to the DY amplitude has the same energy behavior of the current-current four-fermion operators.[3] However, due to the chirality structure of the fermionic tensors, they do not interfere with the SM amplitude, leading to contributions to the squared amplitude that are proportional to s^2/Λ^4.

The first part of this thesis is devoted to the study of the Drell-Yan process within SMEFT framework, with the aim of understanding its implications for LHC phenomenology. In particular, in Chapter 2, we focus on the operators that can be effectively probed at the LHC (i.e. those whose effects grow with energy) and provide the larger contribution (i.e. those who interfere with the SM). Only one class satisfies both criteria: the 4-fermions operators described in classes 5-7. In Chapter 3, we turn to an observable in which the contribution of these operators vanishes, allowing us to study the remaining ones whose effects grow with energy, namely those in point 3 and 8.

[3]The growth with energy of the four-fermion interactions is entirely determined by the spinor polarization of the external fermions, which carry a power of $\sqrt{E}$, and not by the Lorentz structure of the operator, so that scalar, vector, and tensor four-fermion operators all lead to amplitudes that grow with the square of the energy s.

Chapter 2

Current–Current SMEFT operators contributions to the Drell–Yan cross section

In this Chapter, we study the Drell-Yan (DY) process within the Standard Model Effective Field Theory (SMEFT) framework, considering the dimension-6 two-quarks-two-leptons four-fermion operators as written in the Warsaw basis [41], listed in Table 2.1. The notation we adopt is slightly different: we denote by l the left-handed lepton doublet and by q the left-handed quark doublet. When turning to right-handed particles, we indicate with e the generic lepton and with u and d the generic up-family quark and down-family quark, respectively. Chirality is explicitly specified as L and R for for left- and right-handed states. As specified in the previous Chapter, those operators are the only class providing growing-with-energy effects and interfering with the Standard Model (SM) in the DY process [62].

Dimension-6 current-current operators
$O_{lq}^{(3)} = (\bar{l}_L \sigma_I \gamma^\mu l_L)(\bar{q}_L \sigma_I \gamma_\mu q_L)$
$O_{lq}^{(1)} = (\bar{l}_L \gamma^\mu l_L)(\bar{q}_L \gamma_\mu q_L)$
$O_{qe} = (\bar{q}_L \gamma^\mu q_L)(\bar{e}_R \gamma_\mu e_R)$
$O_{lu} = (\bar{l}_L \gamma^\mu l_L)(\bar{u}_R \gamma_\mu u_R)$
$O_{ld} = (\bar{l}_L \gamma^\mu l_L)(\bar{d}_R \gamma_\mu d_R)$
$O_{eu} = (\bar{e}_R \gamma^\mu e_R)(\bar{u}_R \gamma_\mu u_R)$
$O_{ed} = (\bar{e}_R \gamma^\mu e_R)(\bar{d}_R \gamma_\mu d_R)$

Table 2.1: The seven dimension-6 contact operators contributing to Drell-Yan processes written in the notation of ref. [41].

In our analysis we consider both the neutral DY, with a di-lepton final state, and the charged DY, with a lepton and a neutrino final state, resulting experimentally in a single lepton plus missing energy final state. Our aim is to generalize previous constraints on the Wilson coefficients of SMEFT

operators looking at the cross section, including at the same time all the relevant operators [62] and all the relevant uncertainties [61]. Moreover, we investigate the impact of different binning specifications and single vs multi-differential measurements to assess the best strategy of presenting measurements to maximize the sensitivity to SMEFT operators. We exploit both a multi-differential cross-section, triple differential for the neutral channel, and double differential for the charged one, and the single differential cross-section, in the di-lepton invariant mass $m_{\ell\ell}$ (neutral channel) and lepton transverse momentum p_T (charged channel).

At the time of this reaserch, no new experimental measurements of differential DY at high energy have been appeared since the analysis of Ref. [62]. Therefore, we rely, in our analysis, on simulated data, allowing us to explore the impact of different binning strategies. We compare the sensitivity of analyses based on the fully-differential cross-section, on the single differential cross-section obtained by integrating the fully-differential one over the angular variables, and the "enhanced" single differential cross-section obtained with a finer binning in the dimensionful kinematic variables, that are the invariant mass of the two leptons $m_{\ell\ell}$, in the neutral DY, and the lepton transverse momentum, p_T in the charged DY.

We find that, for parameters that are particularly sensitive to the shape of the distribution in the dimensionful kinematic variables, the single differential cross-section in which the full statistics is used to optimize the binning compatibly with the statistical uncertainty, can be as sensitive as the fully differential one. This suggests that, in forthcoming experimental analyses, it would be optimal to have available both fully-differential and single differential (with optimized binning) cross-section information.

The Chapter is organized as follows. in Section 2.1, we describe how we obtained the simulated values for the cross section and the parametrization used to perform the analysis. In Section 2.2 we list all the relevant sources of uncertainty and detail how each was treated to obtain the bounds. In Section 2.3 we provide the technical details about the statistical procedure adopted to extract the bounds. In Section 2.4 we report the results obtained in the various configurations. Finally, in Section 2.5 we summarize the conclusions of the analysis.

2.1 Cross-sections parametrization

The SM+EFT cross-section is obtained using the reweighting strategy outlined in Ref. [61]. This is based on the fact that in DY the new physics contributions factorize not only with respect to the tree-level cross-section, but also with respect to QCD radiative corrections. Such factorization holds separately for each chirality channel, and allows one to generate events for the full SM+EFT process, for any value of the Wilson coefficients, and at NLO QCD accuracy, by just reweighting events for the SM process at NLO QCD accuracy. The reweighting coefficients depend on the chirality of the quarks and leptons and on kinematic quantities such as the partonic center-of-mass energy squared s. The parton shower at NLO QCD accuracy is not modified by the EFT operators, therefore there is no need to account for their contribution. The SM cross-section and the parton shower at NLO

QCD accuracy have been obtained using respectively POWHEG [63, 64] and PYTHIA 8 [65]. To obtain results whose precision matches the experimental measurements, also the NNLO QCD corrections have to be taken into account for the Standard Model contributions (using FEWZ [66]). These contributions are completely negligible for the new physics. Finally, the reweighting coefficients can be modified in order to include also the EW next-to-Leading-Log (NLL) corrections, which become important in the high energy regime we are interested in [61].

We write the cross section in each bin, for both the neutral and the charged channel, exploiting the so-called Cholesky decomposition. This mathematical tool is useful because it guarantees the positivity of the cross section for any value of the Wilson coefficient and it applies here since the cross section can be expressed as a quadratic polynomial in the Wilson coefficients. This method leads to weights parameterizing the SM contribution, the interference between SM and EFT operators, and the quadratic EFT contributions. In particular, we obtain a total of 36 weights for the neutral channel (1 for the SM, 7 for the interference, and 28 for quadratic terms) and 3 weights (1 SM, 1 interference, and 1 quadratic) for the charged one (see Appendix 2.A for details). All relevant uncertainties have been taken into account following the prescription of Ref. [61] and are parametrized as fluctuations of these weights, which we refer to as Cholesky coefficients.

2.2 Uncertainties

Uncertainties represent a crucial aspect of this analysis, as they must be accounted for to ensure a fair comparison with precise experimental data and to obtain reliable constraints. They are included in our framework using nuisance parameters as in Ref. [61].[1] We assume that the nuisance parameters deriving from theoretical uncertainties modify the Cholesky coefficients (in each bin), while the ones linked to the experimental uncertainties have a direct impact on the number of expected events in each bin. We discuss the implementation of both classes of uncertainties in the following.

We indicate with $\bar{c}_{i,I}$ the i-th Cholesky coefficient in the I-th bin, calculated using the Standard Model Central values of α_S, Parton Density Functions (PDFs), and factorization and renormalization scale (μ_F and μ_R)[2]. The SM Central Values input are specified below. On the other hand, we denote with $c_{i,I}$ the corresponding Cholesky coefficient as function of the nuisance parameters. For simplicity, we do not explicitly distinguish between the neutral and charged channels; however, it should be understood that all quantities differ between channels.

- **Theory uncertainties**
  - *Monte Carlo statistic*
    The uncertainty deriving from Monte Carlo statistics is negligible if the simulations provide accurate enough predictions for the SM terms, well below 1%. This is guaranteed by the fact that the new physics contributions are accounted for using reweighting, so

[1] Details on our treatment of nuisance parameters are given in Appendix 2.B.

[2] Notice that, from the cross-section definition in eq. (2.9) and (2.12), this convention implies that $\bar{c}_0 = 1$ for every bin

that their accuracy aligns with that of the SM terms. Within this analysis, we simulated a number of events high enough to ensure that in each bin this requirement is satisfied.

– *Strong coupling constant*

The uncertainty associated to the value of α_S is accounted for through a single nuisance parameter θ^{α_S}, which is the same across all channels and bins. The effect of θ^{α_S} is estimated using POWHEG SM DY [63] Monte Carlo samples reweighted for upper ($\alpha_S^u = 0.1195$), lower ($\alpha_S^l = 0.1165$), and central value ($\alpha_S = 0.1180$) of α_S at the scale of the Z mass. Since this uncertainty is not the leading one in the SM, we can ignore its effect on the new physics Cholesky coefficients and only retain the SM part, parameterized by the coefficient c_0:

$$\begin{aligned} c_{0,I}(\theta^{\alpha_S}) &= \overline{c}_{0,I} e^{k_I^{\alpha_S}\theta^{\alpha_S}} = e^{k_I^{\alpha_S}\theta^{\alpha_S}} , \\ k_I^{\alpha_S} &= \max\left(|\, c_{0,I}^u - \overline{c}_{0,I}\,|, |\, c_{0,I}^l - \overline{c}_{0,I}|\right) . \end{aligned} \tag{2.1}$$

with $c_{0,I}^l = c_{0,I}(\alpha_S^l)$ and $c_{0,I}^u = c_{0,I}(\alpha_S^u)$.

– *Missing higher orders (QCD and EW)*

The uncertainty deriving from the truncation of the perturbative QCD series are accounted for by the introduction of a nuisance parameter θ_I^{TU} for each bin. We consider different values of the factorization and renormalization scales, μ_F and μ_R, respectively: their central values are set to $\overline{\mu}_R = \overline{\mu}_F = \sqrt{\hat{s}}$ and we let them vary independently by multiplicative factors $2^{\pm 1}$, $2^{\pm\frac{1}{2}}$, and 1, with the latter value corresponding to the central value. This gives a grid with 25 values.

Again, the missing higher order uncertainty is not leading in the SM. In particular, while the contribution deriving from the truncation of the NLO EW perturbative series could be completely neglected, the one linked to the truncation of the QCD NNLO perturbative series is relevant only for the c_0 SM coefficient. Such contribution is parameterized as:

$$\begin{aligned} c_{0,I}(\theta_I^{TU}) &= \overline{c}_{0,I} e^{k_I^{\theta^{TU}}\theta_I^{TU}} = e^{k_I^{\theta^{TU}}\theta_I^{TU}} , \\ k_I^{\theta^{TU}} &= \max\left(\frac{|\, c_{0,I}^{max} - \overline{c}_{0,I}\,|}{10}, \frac{|\, c_{0,I}^{min} - \overline{c}_{0,I}\,|}{10}\right) , \end{aligned} \tag{2.2}$$

where $c_{0,I}^{max}$ and $c_{0,I}^{min}$ are the maximum and minimum value of $c_{0,I}$ within the 25 different replicas specified above. The factor 1/10 is due to the fact that the missing higher orders uncertainty on the Cholesky coefficients is obtained from a NLO QCD simulation, while it is known that the NNLO contribution is a factor ~ 10 smaller (see Refs. [61, 66]).

– *Parton Distribution Functions*

PDF uncertainty is the most important theoretical uncertainty in the SM DY process [67–69]. Therefore, we account for it in the Cholesky coefficients of both the SM and the new physics contributions. The PDF uncertainties are parametrized by a vector of nuisance parameter θ_i^{PDF}, corresponding to the eigenvalues of the PDFs within the

Hessian representation, for each bin. As before, we use POWHEG to get the weights of the different Hessian components in the SM calculation. In our case it is enough to consider the 30 components in the PDF set PDF4LHC15_NLO_30_PDFAS (code 90400 in the LHAPDF database [70]). The advantage of the Hessian set is that it automatically provides a definition of the relevant nuisance parameters that can be used across different processes, simplifying the combination of different channels. The parametrization of the Cholesky coefficients as function of the PDFs nuisance parameters is:

$$\begin{aligned} & c_{k,I}(\theta_i^{\rm PDF}) = \overline{c}_{k,I} \exp\left[\sum_{i=1}^{30} \frac{c_{k,I}^{(i)} - \overline{c}_{k,I}}{\overline{c}_{k,I}} \theta_i^{\rm PDF}\right] , \\ & \text{for } k = 0, 1, ..., 35 \qquad \text{(Neutral channel)}, \\ & \text{for } k = 0, 1, 2 \qquad\quad \text{(Charged channels)}. \end{aligned} \tag{2.3}$$

- **Experimental uncertainties**

  Uncertainties associated to the experimental setup could only be accounted for within an analysis or with available information from the experiments. Despite our code gives us full flexibility in accounting for experimental systematic uncertainties with full correlation information, possibly different in the electron and muon channels, in this work we considered a relatively simple pattern of uncertainties, consistent with past analyses. We assume a 2% uncorrelated uncertainty, parametrized by a single nuisance parameters $\theta^{\rm L}$, across all bins and channels, from the determination of the integrated luminosity. For all the other experimental systematic uncertainties, we consider 2% and 5% uncertainties in each bin, uncorrelated among bins, for the neutral and charged channels, respectively. These are parametrized by a set of nuisance parameters $\theta_I^{\rm exp}$ for each bin and each channel.

  We assume that the experimental uncertainties do not affect directly the Cholesky coefficients, and only modify the number of expected events in each bin, referred to as μ_I, from its theoretical prediction $\mu_I^{\rm th}$ as follows:

$$\mu_I = \mu_I^{\rm th} \exp\left(\sum_J \left[\sqrt{\Sigma^{\rm exp}}\right]_I^J \theta_J^{\rm exp} + 0.02\theta^{\rm L}\right) . \tag{2.4}$$

  Here $\Sigma^{\rm exp}$ is the covariance matrix of the experimental systematic uncertainties in the space of bins, that we take proportional to the identity matrix, and $\mu_I^{\rm th} = L \cdot \sigma_I^{\rm th}$, where L is the integrated luminosity, and $\sigma_I^{\rm th}$ is the cross-section in each bin including all sources of theoretical uncertainties specified before.

2.3 The likelihood

The constraints on the Wilson coefficients are obtained using the profiled likelihood ratio test. In general, the test-statistic, referred to as $t_{\boldsymbol{\psi}}$, for this tested is defined as:

$$t_{\boldsymbol{\psi}} = -2\Big(\sup_{\boldsymbol{\theta}} \log \mathscr{L}(\boldsymbol{\psi}, \boldsymbol{\theta}) - \sup_{(\boldsymbol{\psi},\boldsymbol{\theta})} \log \mathscr{L}(\boldsymbol{\psi}, \boldsymbol{\theta})\Big) \,. \tag{2.5}$$

Here $\mathscr{L}$ is the likelihood as function of the parameters, $\boldsymbol{\psi}$ is the vector of parameters of interest, that in our case are the seven Wilson coefficients corresponding to the EFT operators in Table 2.1, and $\boldsymbol{\theta}$ is the vector of nuisance parameters.

The Wilks' theorem ensures that $t_{\boldsymbol{\psi}}$ follows asymptotically a χ^2 distribution with a number of degrees of freedom equal to the number of parameters of interest [71]. Assuming we are in the asymptotic regime, we can construct confidence level boundaries on the Wilson coefficients.

In our analysis, bin by bin, the likelihood is a Poisson distribution of the number of observed events n_I with mean μ_I, multiplied by the likelihood that constrains each nuisance parameter (with auxiliary data). This, can be parametrized as a standard Normal distribution, since the relevant scales have already been taken into account in defining the dependence of the Cholesky coefficients on the nuisance parameters. The combined likelihood is then simply written as:

$$\mathscr{L}_{\rm comb}(\boldsymbol{G}, \boldsymbol{\theta}) = \mathscr{L}_{\rm n}(\boldsymbol{G}, \boldsymbol{\theta}_{\rm n}) \times \mathscr{L}_{\rm c,+}(\boldsymbol{G}, \boldsymbol{\theta}_{\rm c}) \times \mathscr{L}_{\rm c,-}(\boldsymbol{G}, \boldsymbol{\theta}_{\rm c}) \times \mathscr{L}_{\rm aux}(\boldsymbol{\theta}) \,, \tag{2.6}$$

where $\boldsymbol{G}$ is a vector containing the Wilson coefficients, that are our parameter of interest, $\boldsymbol{\theta}$ is a vector containing all the nuisance parameters while $\boldsymbol{\theta}_{\rm n/c}$ is a vector containing the nuisance parameters only relevant for the neutral/charged channel. The explicit definition of each term is given in Appendix 2.C.

2.4 Projections and results

In this Section, we present the results obtained in our analysis.

We report the 95% confidence level (CL) projected bounds, at the LHC at 13 TeV, on each of the Wilson coefficients of the operators in Table 2.1. We assume that the SM central value for the number of observed events in each bin is given by $n_I = L \cdot \overline{\sigma}_I$, where L is the luminosity and $\overline{\sigma}_I$ is the cross-section in each bin calculated assuming the central values for $\alpha_{\rm S}$, PDFs, and factorization and renormalization scale, and setting to zero all the experimental uncertainties and the EFT effects.

We consider the benchmark luminosity values of 100, 300, and 3000 fb^{-1} as in Ref. [61]. These correspond, respectively, to the luminosity available at the time of the analysis, at the end of the LHC run, and at the end of the High-Luminosity LHC (HL-LHC) run.

As outlined in the previous Sections, we consider different binning schemes, looking both at multi- and single-differential distributions. The binning specification for the multi-differential analysis is analogous to the one of Ref. [62]. In particular, given the dilepton invariant mass $m_{\ell\ell}$, the

scattering angle $c_* = \cos\theta_*$, and the absolute value of the ratio between the dilepton rapidity and its maximum value $|y/y_{\max}|$, the binning scheme for the neutral channel is given by

$$\begin{cases} m_{\ell\ell} & : \{300, 360, 450, 600, 800, 1100, \\ & \quad 1500, 2000, 2600, 13000\}\,\text{GeV}\,, \\ c_* & : \{-1, -0.6, -0.2, 0.2, 0.6, 1\}\,, \\ \left|\dfrac{y}{y_{\max}}\right| & : \{0, 1/3, 2/3, 1\}\,. \end{cases} \tag{2.7}$$

Analogously, for the charged channel, given the lepton transverse momentum p_T and the absolute value of the ratio between the lepton pseudo-rapidity and its maximum value $|\eta/\eta_{\max}|$, the binning is given by

$$\begin{cases} p_T & : \{150, 180, 275, 300, 400, 550, \\ & \quad 750, 1000, 1300, 7500\}\,\text{GeV}\,, \\ \left|\dfrac{\eta}{\eta_{\max}}\right| & : \{0, 1/3, 2/3, 1\}\,. \end{cases} \tag{2.8}$$

Turning to the single differential analysis, we considered two different scenarios. In the first one, the single differential cross-sections are obtained integrating the fully differential ones over angular and rapidity variables (neutral channel) and over pseudo-rapidity variable (charged channel) keeping the aforementioned binning for the dimensionful kinematic variables. In the second one, the large number of expected events is exploited to consider a finer binning in the $m_{\ell\ell}/p_T$ (neutral/charged) variables. Such binning is determined by requiring, in each bin, the minimum number of events that keeps the statistical uncertainty of our Monte Carlo simulations negligible (below 0.2%). This procedure results in 113 bins in $m_{\ell\ell}$ for the neutral channel and 26 bins in p_T for the charged one (the full binning specification is reported in Appendix 2.D).

The constraints we obtained in the multi-differential analysis and in the single differential one after the integration over the rapidity and the angular variables are compatible with the results reported in Ref. [62]. Our bounds are slightly more conservative because we also included the uncertainties deriving from the value of the strong coupling constant α_S, and the truncation of the perturbative series. Furthermore, we considered a 2% experimental uncertainty in the neutral channel and a 5% in the charged one, while in Ref. [62] a 2% uncertainty was considered for both channels. The single differential analysis with a finer binning was not considered in previous works.

Table 2.2 reports the 1-dimensional 95% CL projected bounds on the Wilson coefficients in the three scenarios discussed above: considering a fully-differential cross-section, a single differential cross-section obtained integrating the fully-differential one over angular and rapidity variables, and a single differential cross-section with a finer binning. The three aforementioned integrated luminosity values are considered. The numerical values in the table are obtained considering only one Wilson coefficient at a time, with all the others set to zero, building the test-statistic $t_{\boldsymbol{\psi}}$ as function of a single parameter, and assuming it is asymptotically distributed as a χ^2 with one degree of freedom. The same results are also shown in Figure 2.1, which highlights the differences between the three

95% CL	$\mathcal{L} = 100\,\text{fb}^{-1}$			$\mathcal{L} = 300\,\text{fb}^{-1}$			$\mathcal{L} = 3000\,\text{fb}^{-1}$		
$[10^{-9}\text{GeV}^{-2}]$	Fully-Dif	Single integrated	Single fine bins	Fully-Dif	Single integrated	Single fine bins	Fully-Dif	Single integrated	Single fine bins
$G_{lq}^{(3)}$	[−2.04, 2.10]	[−2.23, 2.31]	[−2.07, 2.14]	[−1.43, 1.47]	[−1.65, 1.71]	[−1.50, 1.56]	[−0.73, 0.75]	[−1.02, 1.08]	[−0.91, 0.96]
$G_{lq}^{(1)}$	[−8.50, 14.7]	[−9.48, 20.2]	[−8.85, 18.0]	[−5.75, 9.01]	[−6.72, 15.5]	[−6.07, 12.1]	[−2.66, 3.36]	[−3.68, 9.22]	[−3.04, 4.54]
G_{qe}	[−8.72, 15.2]	[−11.1, 18.0]	[−10.4, 16.8]	[−5.98, 10.8]	[−8.10, 13.9]	[−7.38, 12.5]	[−2.88, 5.26]	[−4.68, 9.10]	[−3.95, 7.19]
G_{lu}	[−8.21, 13.5]	[−11.3, 21.1]	[−10.5, 18.5]	[−5.40, 8.02]	[−8.03, 16.0]	[−7.22, 12.7]	[−2.41, 3.03]	[−4.43, 10.0]	[−3.65, 5.93]
G_{ld}	[−27.1, 18.1]	[−30.5, 21.6]	[−29.3, 20.5]	[−20.3, 12.8]	[−23.6, 16.1]	[−22.3, 14.9]	[−11.7, 6.70]	[−15.2, 9.72]	[−13.6, 8.41]
G_{eu}	[−6.29, 7.37]	[−6.91, 8.54]	[−6.27, 7.42]	[−4.16, 4.64]	[−4.77, 5.57]	[−4.16, 4.68]	[−1.88, 1.99]	[−2.52, 2.75]	[−2.00, 2.11]
G_{ed}	[−27.7, 15.5]	[−29.9, 16.7]	[−27.1, 15.4]	[−20.5, 10.9]	[−23.2, 12.2]	[−19.8, 10.9]	[−11.4, 5.56]	[−15.4, 7.09]	[−11.1, 5.84]

Table 2.2: One dimensional single parameter 95% confidence intervals for the seven EFT Wilson coefficients in units of $10^{-9}\,\text{GeV}^{-2}$, for integrated luminosity values of $100\,\text{fb}^{-1}$, $300\,\text{fb}^{-1}$ and $3000\,\text{fb}^{-1}$. The results are obtained considering a fully-differential cross-section, a single differential cross-section obtained from the fully-differential one integrating over angular and rapidity variables and a single differential cross-section with a fine binning.

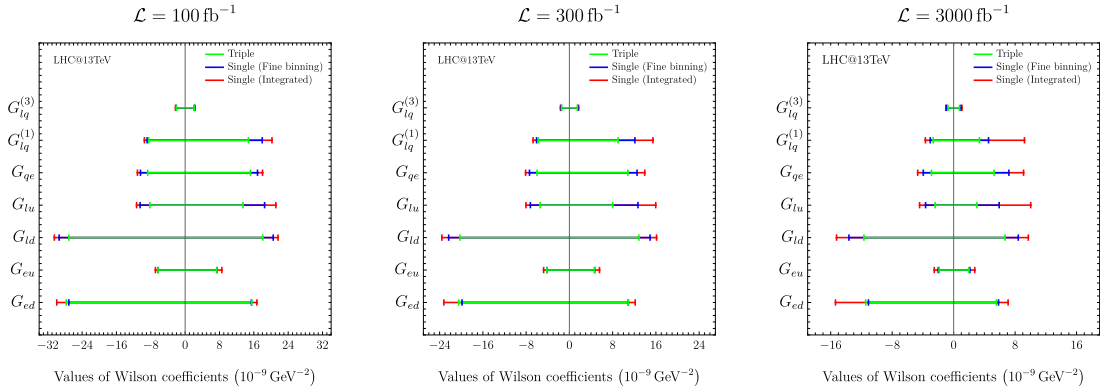


Figure 2.1: Comparison between the one dimensional single parameter (switch on one coefficient at a time and set the other six to 0) 95% confidence intervals for the seven EFT Wilson coefficients. These results are obtained considering a multi-differential cross-section (green), a single differential cross-section obtained from the multi-differential one integrating over angular and rapidity variables (red) and a single differential cross-section with a fine binning (blue). Collider energy is set to 13 TeV and integrated luminosity of $100\,\text{fb}^{-1}$, $300\,\text{fb}^{-1}$ and $3000\,\text{fb}^{-1}$ are considered.

binning strategies.

From the figures and the table we can draw some important conclusions.

- As expected, the fully-differential analysis always provides the tightest constraints on the Wilson coefficients.

- The single-differential analysis with fine bins often reaches the same level of sensitivity as the fully-differential one. This is particularly true for the Wilson coefficients $G_{lq}^{(3)}$, G_{eu}, and G_{ed}, which are therefore much less sensitive to angular distributions. The sensitivity is still weaker than the fully-differential analysis for the other Wilson coefficients, but the difference is, especially in the negative end of the bound, generally less than 10%, which can be considered within the precision of our determination, based on asymptotic formulae for the likelihood-ratio test-statistic distribution.

95% CL	$\mathcal{L} = 100\,\text{fb}^{-1}$			$\mathcal{L} = 300\,\text{fb}^{-1}$			$\mathcal{L} = 3000\,\text{fb}^{-1}$		
$[10^{-9}\text{GeV}^{-2}]$	Fully-Dif	Single integrated	Single fine bins	Fully-Dif	Single integrated	Single fine bins	Fully-Dif	Single integrated	Single fine bins
$G_{lq}^{(3)}$	[−2.17, 2.17]	[−2.35, 2.39]	[−2.16, 2.18]	[−1.53, 1.54]	[−1.73, 1.77]	[−1.57, 1.59]	[−0.82, 0.82]	[−1.05, 1.10]	[−0.93, 0.98]
$G_{lq}^{(1)}$	[−14.1, 17.2]	[−15.7, 20.8]	[−14.9, 19.5]	[−10.5, 12.5]	[−12.1, 16.3]	[−11.2, 14.8]	[−6.11, 6.78]	[−7.77, 11.2]	[−6.52, 9.22]
G_{qe}	[−12.3, 15.6]	[−16.9, 18.3]	[−16.1, 17.1]	[−9.01, 11.5]	[−13.2, 14.2]	[−12.2, 12.8]	[−5.12, 6.50]	[−8.77, 9.36]	[−7.42, 7.52]
G_{lu}	[−10.9, 17.4]	[−18.3, 22.1]	[−17.4, 20.5]	[−7.56, 12.4]	[−14.1, 17.0]	[−13.0, 15.2]	[−3.84, 6.55]	[−9.11, 11.2]	[−7.59, 8.81]
G_{ld}	[−27.9, 24.6]	[−32.7, 32.0]	[−30.9, 30.8]	[−20.9, 18.4]	[−25.7, 25.5]	[−23.6, 24.1]	[−12.2, 10.7]	[−17.4, 17.8]	[−14.6, 15.8]
G_{eu}	[−10.4, 17.3]	[−11.8, 19.2]	[−11.3, 17.6]	[−7.47, 12.9]	[−8.85, 14.9]	[−8.30, 13.1]	[−4.05, 7.75]	[−5.45, 9.88]	[−4.74, 7.65]
G_{ed}	[−28.7, 25.7]	[−30.2, 28.7]	[−28.1, 27.7]	[−21.7, 19.5]	[−23.5, 22.9]	[−21.1, 21.7]	[−13.1, 11.5]	[−15.6, 16.0]	[−12.5, 14.3]

Table 2.3: One dimensional profiled (constraining one by one each coefficient while treating the other six as nuisance parameters) 95% confidence intervals for the seven EFT Wilson coefficients in units of $10^{-9}\,\text{GeV}^{-2}$, for integrated luminosity values of $100\,\text{fb}^{-1}$, $300\,\text{fb}^{-1}$ and $3000\,\text{fb}^{-1}$. The results are obtained considering a fully-differential cross-section, a single differential cross-section obtained from the fully-differential one integrating over angular and rapidity variables and a single differential cross-section with a fine binning.

- The single differential analysis obtained by integrating the fully-differential cross-section over the angular and rapidity variables is generally less sensitive than the fully-differential one. This is particularly true for the operators $G_{lq}^{(1)}$, G_{qe}, G_{lu}, and G_{ld}, whose upper bounds reach differences from the fully-differential analysis of more than a factor of two. This implies that the bound on these parameters is particularly sensitive to the angular and rapidity distributions.

Similar results, this time obtained marginalizing the likelihood over the other Wilson coefficients, instead of setting them to zero, are reported in Table 2.3 and shown in Figure 2.2. The conclusions are unchanged.

Finally, for completeness, we analyse also the 2-dimensional 95% confidence level intervals for the twenty-one possible pairs of Wilson coefficients. We do it considering only the multi-differential distribution, which has proven to be generally more powerful and we adopt the same specifications described above. Additionally, we compare these results with those obtained when Taylor expanding the likelihood to the second order in both the Wilson coefficients and the nuisance parameters. Figure 2.3 and 2.4 show the comparison between the two approaches for two representative operator pairs, $G_{lq}^{(3)}$ with G_{qe} and $G_{lq}^{(1)}$ with G_{eu}.

In this thesis, we present for simplicity just these two examples for the 2-dimensional bounds, although all possible operators pairs have been analysed. The results obtained are, as before, comparable with those of Ref. [62], even if we included more sources of uncertainty. It is straightforward to observe that the linear approximation, while capturing the dominant NP contribution, fails to account for correlations among different operators (see Fig 2.4) and is therefore not reliable. The same conclusions hold even when switching to the operators basis proposed again in [62], where the impact of angular dependencies are less relevant. In figure 2.5 and 2.6 we report, as before, two significant examples to make this manifest.

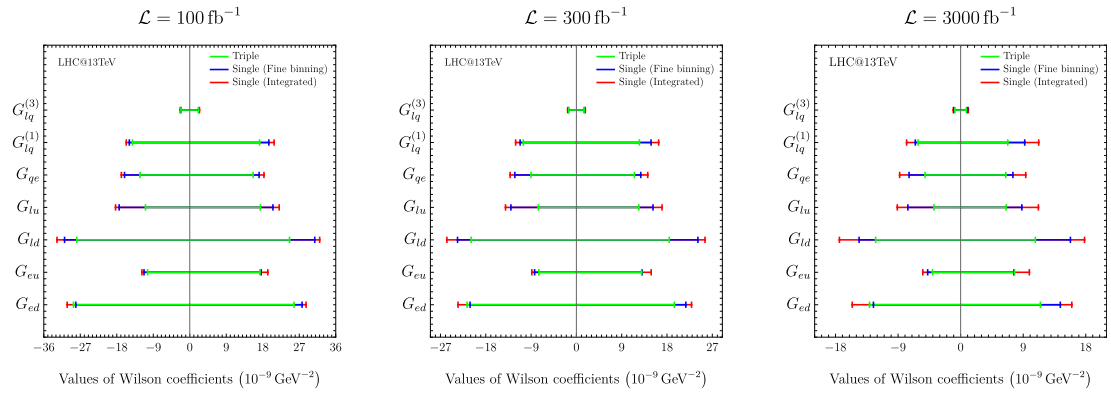


Figure 2.2: Comparison between the one dimensional profiled 95% confidence intervals for the seven EFT Wilson coefficients. These results are obtained considering a multi-differential cross-section (green), a single differential cross-section obtained from the multi-differential one integrating over angular and rapidity variables (red) and a single differential cross-section with a fine binning (blue). Collider energy is set to 13 TeV and integrated luminosity of $100\,\mathrm{fb}^{-1}$, $300\,\mathrm{fb}^{-1}$ and $3000\,\mathrm{fb}^{-1}$ are considered.

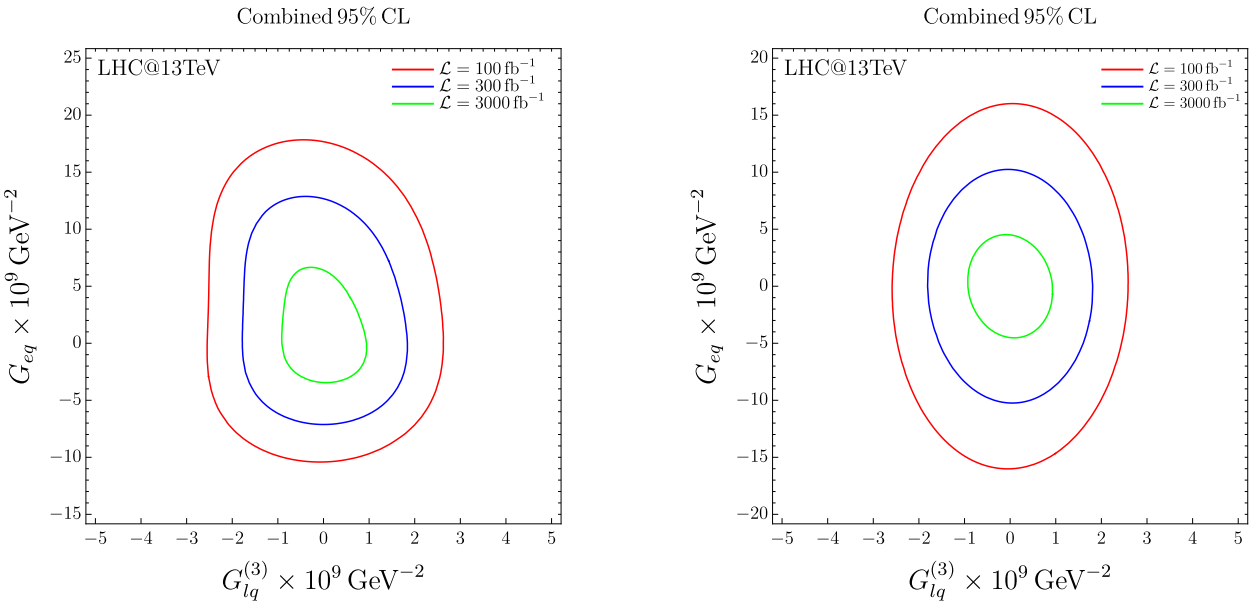


Figure 2.3: 95% confidence level interval for the $G_{lq}^{(3)}$, G_{qe} pair of Wilson coefficient setting the others to zero. Collider energy is set to 13 TeV and integrated luminosity of $100\,\mathrm{fb}^{-1}$, $300\,\mathrm{fb}^{-1}$ and $3000\,\mathrm{fb}^{-1}$ are considered. Left: the bound is obtained using the exact formulas described within the Chapter. Right: the bound is obtained Taylor expanding the likelihood at the second order in both the Wilson coefficients and nuisance parameters.

2.5 Conclusions

In this Chapter, we reported the work where we generalized previous analysis on the determination of the bounds on the Wilson coefficients of the EFT operators entering in the DY processes at the

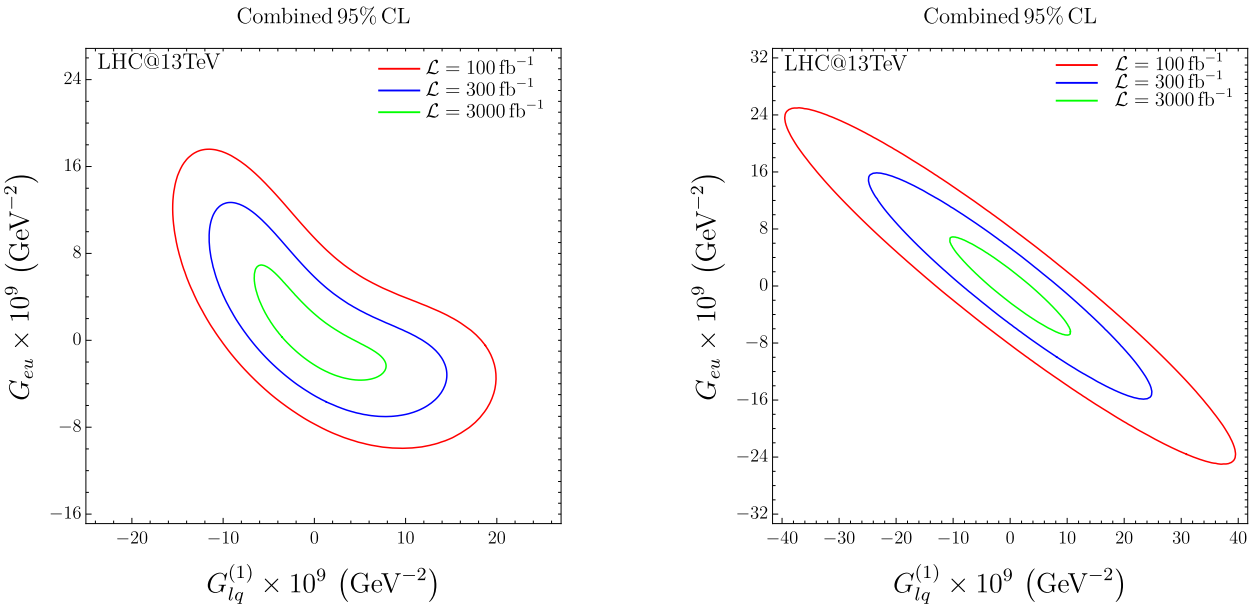


Figure 2.4: 95% confidence level interval for the $G_{lq}^{(1)}$, G_{eu} pair of Wilson coefficient setting the others to zero. Collider energy is set to 13 TeV and integrated luminosity of $100\,\mathrm{fb}^{-1}$, $300\,\mathrm{fb}^{-1}$ and $3000\,\mathrm{fb}^{-1}$ are considered. Left: the bound is obtained using the exact formulas described within the Chapter. Right: the bound is obtained Taylor expanding the likelihood at the second order in both the Wilson coefficients and nuisance parameters.

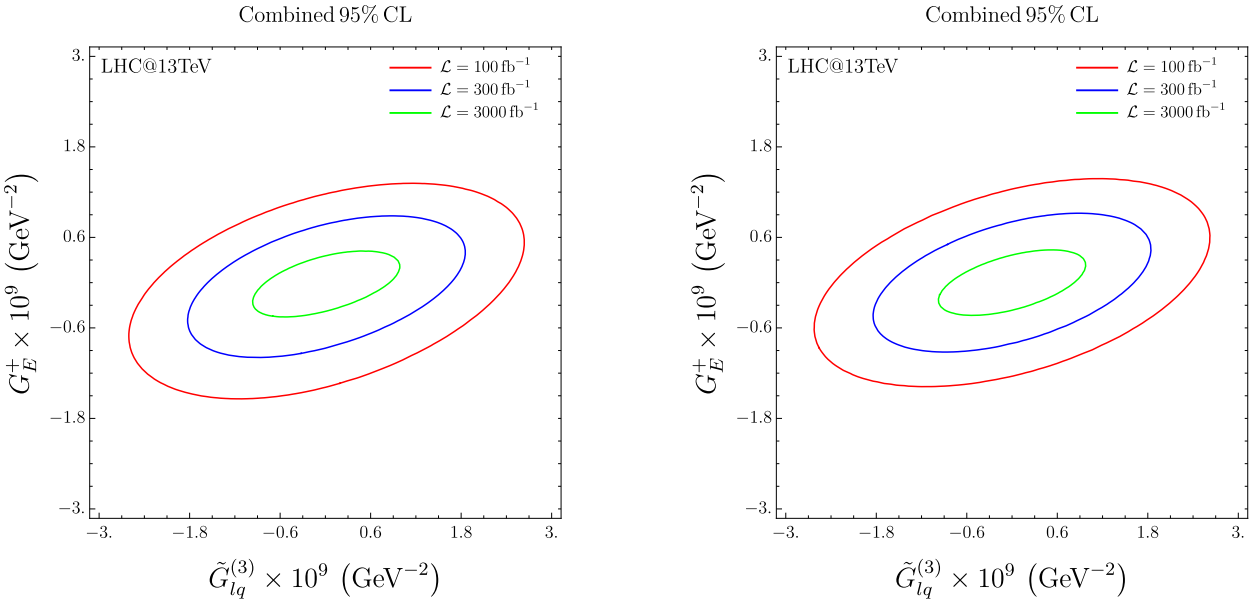


Figure 2.5: 95% confidence level interval for the $\tilde{G}_{lq}^{(3)}$, G_E^+ (defined in **??**) pair of Wilson coefficient setting the others to zero. Collider energy is set to 13 TeV and integrated luminosity of $100\,\mathrm{fb}^{-1}$, $300\,\mathrm{fb}^{-1}$ and $3000\,\mathrm{fb}^{-1}$ are considered. Left: the bound is obtained using the exact formulas described within the Chapter. Right: the bound is obtained Taylor expanding the likelihood at the second order in both the Wilson coefficients and nuisance parameters.

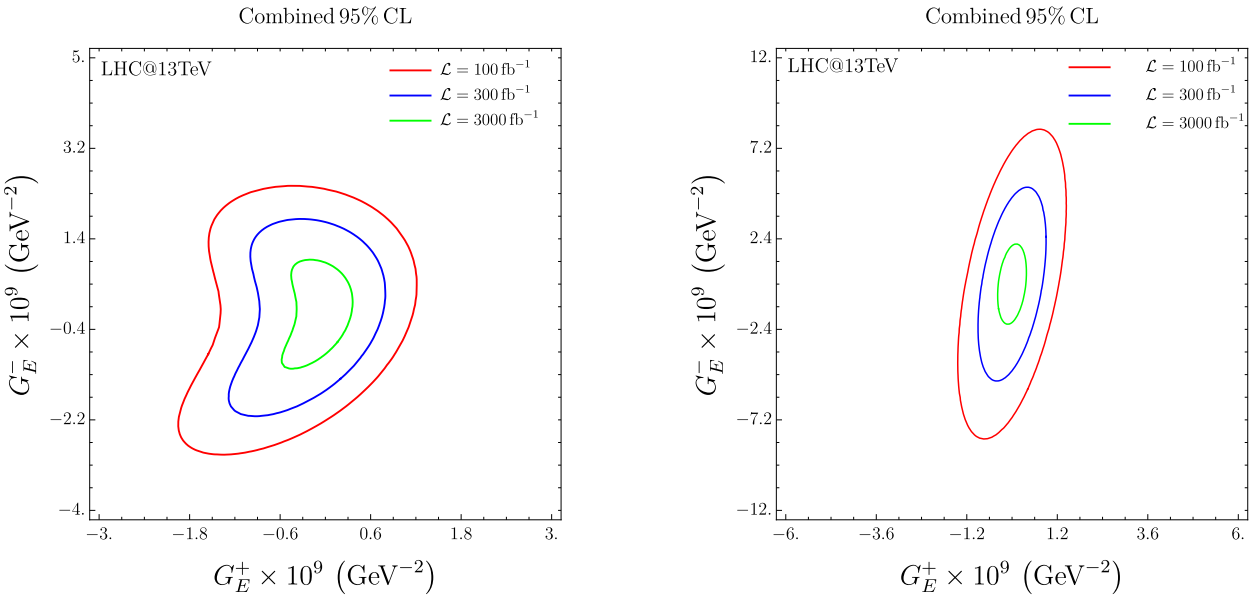


Figure 2.6: 95% confidence level interval for the G_E^+, G_E^- (defined in **??**) pair of Wilson coefficient setting the others to zero. Collider energy is set to 13 TeV and integrated luminosity of $100\,\mathrm{fb}^{-1}$, $300\,\mathrm{fb}^{-1}$ and $3000\,\mathrm{fb}^{-1}$ are considered. Left: the bound is obtained using the exact formulas described within the Chapter. Right: the bound is obtained Taylor expanding the likelihood at the second order in both the Wilson coefficients and nuisance parameters.

LHC. In particular, we considered the operators providing the leading contributions to the cross-section (i.e. the dimension-6 ones) and whose effects are enhanced at the LHC (i.e. grow with energy). Referring to the Warsaw basis [41], these operators are the 4-fermions that share the same chirality structure as the SM.

We have put together the “all-operators" approach of Ref. [62] with the “all-uncertainties" approach of Ref. [61]. We found consistent results with the previous works in the multi-differential and integrated single differential analysis [62] and we added a single differential “fine binning" analysis.

We found that, in general, a multi-differential approach provides tighter constraints. However, also the single-differential analysis with an optimized binning yielded meaningful results, providing, for a subset of the bounds, similar performances. In this particular case we tried to constrain seven operators looking at a single process, therefore is not surprising that a multi-differential analysis, containing more information, results in better results . Nonetheless, if an analysis targets some operators that are particularly sensitive to a single kinematic variable, while simultaneously considering different processes, our results suggest that a finely optimized binning in that variable could even outperform the multi-differential approach. For these reasons, we emphasize the importance, in future experimental analysis, of presenting results in multiple complementary ways, providing the community with the maximum amount of information extracted from the analysis.

Appendix

2.A Cross-section parametrization

The cross-section as a function of the Wilson coefficients is a non-negative quadratic polynomial in $G_{lq}^{(3)}$, $G_{lq}^{(1)}$, G_{qe}, G_{lu}, G_{ld}, G_{eu}, G_{ed}. The cross-section in each bin I can therefore be parametrized using the Cholesky decomposition. In the neutral channel, this can be written as:

$$\sigma_I^{\text{th,n}} = \overline{\sigma}_I^{\text{SM,n}} c_{0,I}^2 \left\| \sum_{j=1}^{8} \mathcal{C}_{Ij} G_j \right\|^2 , \tag{2.9}$$

with the Cholesky matrix

$$\mathcal{C}_{Ij} = \begin{pmatrix} 1 & c_{1,I} & c_{2,I} & c_{3,I} & c_{4,I} & c_{5,I} & c_{6,I} & c_{7,I} \\ 0 & c_{8,I} & c_{15,I} & c_{16,I} & c_{17,I} & c_{18,I} & c_{19,I} & c_{20,I} \\ 0 & 0 & c_{9,I} & c_{21,I} & c_{22,I} & c_{23,I} & c_{24,I} & c_{25,I} \\ 0 & 0 & 0 & c_{10,I} & c_{26,I} & c_{27,I} & c_{28,I} & c_{29,I} \\ 0 & 0 & 0 & 0 & c_{11,I} & c_{30,I} & c_{31,I} & c_{32,I} \\ 0 & 0 & 0 & 0 & 0 & c_{12,I} & c_{33,I} & c_{34,I} \\ 0 & 0 & 0 & 0 & 0 & 0 & c_{13,I} & c_{35,I} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & c_{14,I} \end{pmatrix} \tag{2.10}$$

and the vector of Wilson coefficients G_j with $j = 1, ..., 8$

$$\mathbf{G} = \left(1, G_{lq}^{(1)}, G_{lq}^{(3)}, G_{qe}, G_{lu}, G_{ld}, G_{eu}, G_{ed}\right) , \tag{2.11}$$

where the 1 accounts for the SM contribution, and the Cholesky coefficients $c_{k,I}$ are functions of the nuisance parameters θ^{α_S}, θ_i^{PDF}, and θ_I^{TU}, as described in the main text. The same parametrization holds for the charged channel, where it is more simply written as

$$\sigma_I^{\text{th,c}} = \overline{\sigma}_I^{\text{SM,c}} c_{0,I}^2 \left\| \begin{pmatrix} 1 & c_{1,I} \\ 0 & c_{2,I} \end{pmatrix} \begin{pmatrix} 1 \\ G_{lq}^{(3)} \end{pmatrix} \right\|^2 . \tag{2.12}$$

2.B Nuisance parameters

Uncertainties in the theoretical predictions are parametrized by nuisance parameters with the following procedure. The theoretical predictions are expressed as functions of the Cholesky coefficients, which are, in turn, functions of the theory input parameters. Schematically, we could write:

$$c_{k,I} \equiv c_{k,I}(f_i(x,q^2), \alpha_S, \mu_R, \mu_F, \ldots)\,, \tag{2.13}$$

where the index k runs over the Cholesky coefficients, the index I runs over the bins, and where $f_i(q^2,x)$ are the PDFs (for parton i, with Bjorken scaling x, and squared momentum transfer q^2), α_S is the strong coupling constant, μ_R and μ_F are the renormalization and factorization scales, respectively, and the ellipsis stands for any other relevant input parameter. We assume that these parameters are not exactly known, but their estimators are determined, with an uncertainty, by auxiliary measurements, and have a central value $\overline{x}$ and a standard deviation σ_x. We can therefore write, for each parameter, a relation of the form:

$$\hat{x} \sim \overline{x} + \sigma_x \theta^x\,, \tag{2.14}$$

where $\hat{x}$ represents the estimator of the parameter x, which is a random variable, expressed as a function of θ^x, a standard normal random variable parametrizing uncertainty in the determination of $\hat{x}$. We can then write the Cholesky coefficients as functions of the nuisance parameter as:

$$\begin{aligned} c_{k,I} &\simeq c_{k,I}(\overline{x} + \sigma_x\theta^x) \simeq c_{k,I}(\overline{x}) + \sigma_x\theta^x \\ &\equiv \overline{c}_{k,I} + \sigma_x\theta^x = \overline{c}_{k,I}\left(1 + \frac{\sigma_x}{\overline{c}_{k,I}}\theta^x\right) \\ &\simeq \overline{c}_{k,I}\exp\left(\frac{\sigma_x}{\overline{c}_{k,I}}\theta^x\right)\,, \end{aligned} \tag{2.15}$$

where $\overline{c}_{k,I}$ represents a central value for the Cholesky coefficient, which we assume to correspond to the central SM prediction, and θ^x is a standard normal random variable. Notice that, even if we dropped the hat over $c_{k,I}$, this last expression has to be interpreted as the estimator of the Cholesky coefficient and, as such, as a random variable. The scope of the last step in eq. (2.15), in which we express the parenthesis as an exponential, is to guarantee that the sign of the Cholesky coefficient is preserved, and is needed to guarantee a positive cross-section for any value of the random variable θ^x. Obviously, assuming that uncertainties are small compared to the central values, which is certainly the case in our analysis, the exponential parametrization gives results comparable to the linear one. Effectively, the determination of the σ_x parameter is performed by running the Monte Carlo simulations with varying values of the input parameters and computing the distribution of the (estimators of the) Cholesky coefficients. The way in which σ_x is determined for different parameters is described in the main text. The procedure can also be simply generalized to correlated uncertainties, such as the PDF uncertainty, by considering the covariance matrix for

the nuisance parameters.

2.C Likelihood

Each factor of the combined likelihood in equation eq. (2.6) is defined in the following. The neutral/charged channel likelihoods are:

$$\begin{aligned} \mathscr{L}_{\rm n}(\boldsymbol{G},\boldsymbol{\theta}_{\rm n}) &= \prod_{I_{\rm n}=1}^{N_{\rm n}} \text{Poisson}\left[n_{I_{\rm n}} \mid \mu_{I_{\rm n}}(\mathbf{G},\boldsymbol{\theta}_{\rm n})\right] , \\ \mathscr{L}_{{\rm c},\pm}(\boldsymbol{G},\boldsymbol{\theta}_{\rm c}) &= \prod_{I_{\rm c}=1}^{N_{\rm c}} \text{Poisson}\left[n^{\pm}_{I_{\rm c}} \mid \mu^{\pm}_{I_{\rm c}}(\mathbf{G},\boldsymbol{\theta}_{\rm c})\right] , \end{aligned} \tag{2.16}$$

where $\mathbf{G}$ is the vector defined in eq. (2.11), and

$$\begin{aligned} \boldsymbol{\theta}_{\rm n} &= \left(\theta^{\alpha_{\rm S}}, \theta_i^{\rm PDF}, \theta_{I_{\rm n}}^{\rm TU}, \theta_{I_{\rm n}}^{\rm exp}, \theta^{\rm L}\right) , \\ \boldsymbol{\theta}_{\rm c} &= \left(\theta^{\alpha_{\rm S}}, \theta_i^{\rm PDF}, \theta_{I_{\rm c}}^{\rm TU}, \theta_{I_{\rm c}}^{\rm exp}, \theta^{\rm L}\right) . \end{aligned} \tag{2.17}$$

The auxiliary likelihood of the nuisance parameters is:

$$\begin{aligned} \mathscr{L}_{\rm aux}(\boldsymbol{\theta}) &= f_{\alpha_{\rm S}}(\theta^{\alpha_{\rm S}}) f_{\rm L}(\theta^{\rm L}) \left(\prod_{i=1}^{30} f_{\rm PDF}(\theta_i^{\rm PDF})\right) \\ &\quad \times \prod_{I_{\rm n}=1}^{N_{\rm n}} f_{\rm TU}(\theta_{I_{\rm n}}^{\rm TU}) f_{\rm exp}(\theta_{I_{\rm n}}^{\rm exp}) \\ &\quad \times \prod_{I_{\rm c}=1}^{N_{\rm c}} f_{\rm TU}(\theta_{I_{\rm c}}^{\rm TU}) f_{\rm exp}(\theta_{I_{\rm c}}^{\rm exp}) . \end{aligned} \tag{2.18}$$

Where the function $f(\theta)$ represents the Likelihood of the nuisance parameter θ as determined by auxiliary measurements, scaled, as explained in the previous appendix, to zero mean $\mu_\theta = 0$ and unit standard deviation $\sigma_\theta = 1$:

$$f(\theta) = \frac{1}{\sqrt{2\pi}} e^{-\frac{\theta^2}{2}} . \tag{2.19}$$

2.D Fine Binning

The "fine binning" have been obtained imposing to have a negligible MC statistical error in each bin (below 1%). We achieved this imposing to have $\sim 2 \cdot 10^5$ MC events for each bin in our simulation, that corresponds to a statistical error of $\sim 0.22\%$. The resulting binning, for the neutral and charged channels, is:

- Neutral channel:
  $m_{\ell\ell}$: {300, 305, 309, 315, 320, 326, 332, 337, 342, 348, 355, 362, 368, 375, 381, 389, 397, 405,

413, 420, 427, 435, 444, 453, 462, 470, 478, 487, 497, 507, 518, 527, 536, 546, 556, 567, 580, 592, 602, 614, 626, 639, 653, 669, 684, 696, 709, 723, 739, 756, 774, 792, 805, 819, 834, 850, 868, 888, 908, 924, 941, 959, 979, 1001, 1026, 1053, 1077, 1095, 1115, 1137, 1161, 1188, 1218, 1251, 1275, 1297, 1322, 1349, 1379, 1413, 1451, 1494, 1521, 1548, 1577, 1610, 1647, 1689, 1737, 1792, 1823, 1854, 1888, 1926, 1969, 2018, 2076, 2142, 2184, 2220, 2259, 2305, 2357, 2417, 2489, 2575, 2680, 2910, 3105, 3365, 3752, 4126, 4802, 13000} GeV.

- Charged channel:
  p_T : {150, 163, 177, 191, 207, 225, 244, 264, 288, 313, 342, 373, 407, 445, 488, 537, 591, 652, 723, 802, 896, 1003, 1130, 1292, 1493, 1770, 6500} GeV.

Chapter 3

Isolating chirality-breaking SMEFT operators with Drell-Yan angular analysis

In this Chapter, we investigate another way to probe New Physics (NP) using the Drell–Yan (DY) process. We focus on the so-called Lam–Tung relation [72], whose violation can be exploited to search for NP effects that would otherwise be suppressed. In particular, looking at the classification outlined in Sec 1.5, these effects are those deriving from operators in classes 3 and 8. These operators lead to contributions to the DY amplitude that grow with energy, but do not interfere with the SM amplitude, generating contributions to the squared amplitude that are proportional to $v^2 s/\Lambda^4$ (for class 3) and s^2/Λ^4 (for class 8), respectively. Considering only the first two generations, which are those relevant for our work, we rewrite them, as:

$$\begin{aligned}
\mathcal{L}_{\psi^2 X\varphi} = \frac{1}{\Lambda^2} \sum_{p=e,\mu} \Big[& (\bar{\ell}_p \sigma^{\mu\nu} e_p)\big(c_{e_p B} B_{\mu\nu} + c_{e_p W} \tau^I W^I_{\mu\nu}\big)\,\varphi + (\bar{q}_p \sigma^{\mu\nu} d_p)\big(c_{d_p B} B_{\mu\nu} + c_{d_p W} \tau^I W^I_{\mu\nu}\big)\,\varphi \\
& + (\bar{q}_p \sigma^{\mu\nu} u_p)\big(c_{u_p B} B_{\mu\nu} + c_{u_p W} \tau^I W^I_{\mu\nu}\big)\,\tilde{\varphi} + \text{h.c.}\Big], \\
\mathcal{L}_{\psi^4} = \frac{1}{\Lambda^2} \sum_{\substack{p=e,\mu \\ r=u,d,s,c}} \Big[& c_{\ell_p e_p dq}\,(\bar{\ell}^j_p e_p)(\bar{d}_r q^j_r) + c^{(1)}_{\ell_p e_p qu} (\bar{\ell}^j_p e_p)\epsilon_{jk} (\bar{q}^k_r u_r) \\
& + c^{(3)}_{\ell_p e_p qu} (\bar{\ell}^j_p \sigma_{\mu\nu} e_p)\epsilon_{jk} (\bar{q}^k_r \sigma^{\mu\nu} u_r) + \text{h.c.}\Big],
\end{aligned} \tag{3.1}$$

where we have introduced the dimensionless Wilson coefficients c_i and we have made explicit our assumption of a diagonal flavor structure in the lepton sector and of flavor universality in the quark sector (the Wilson coefficients of the four-fermion operators do not depend on the quark flavor index r). The UV origin of these operators can be diverse and a detailed discussion of explicit models is beyond the phenomenological scope of this Chapter. However, we present in Appendix 3.F a brief

overview of possible UV completions that can generate some of these operators, possibly with a sizable coefficient.

The main idea that we exploit in this Chapter is on the one hand to profit of the energy growth of the chirality-breaking operators discussed above, and on the other hand to isolate their contribution from the SM and the SMEFT operators that interfere with the SM by considering observables where the contribution of the latter is suppressed. To do so, we focus on the angular distribution of the lepton pair in the DY process, which is sensitive to the chirality structure of the operators, and on the violation of the Lam-Tung relation [72, 73], which guarantees a suppression of the SM contribution, and of the NP contribution with the same chirality structure of the SM, to the relevant observables. On top of considering observables that are suppressed for the SM due to the Lam-Tung relation, we also exploit the growing-with-energy behavior of the chirality-breaking operators to enhance their contribution with respect to the SM.

The Chapter is organized as follows. In Section 3.1 we describe the angular decomposition of the Drell-Yan differential cross-section, the related angular coefficients, and the Lam-Tung relation, which is a key ingredient in our analysis. In Section 3.2 we review the experimental status of the measurements of the angular coefficients in DY, while in Section 3.3 we present our precise estimates of the cross-section and of the angular observables in the SM, both in the transverse momentum and invariant mass distributions of the lepton pair, for the LHC and HL-LHC. In Section 3.4 we discuss the Lam-Tung relation and compute the contribution of the chirality-breaking SMEFT operators to the relevant observables. In Section 3.5 we present our analysis of the LHC and HL-LHC sensitivity on the relevant SMEFT operators, based on a likelihood fit to the aforementioned observables. In Section 3.6 we summarize the results, while in Section 3.7 we draw our conclusions. The Chapter is complemented by a number of appendices reporting our choices on the definition of the observables and some analytic results relevant for our analysis.

3.1 Angular coefficients and the $A_0 - A_2$ observable

The fully-differential inclusive cross-section of the neutral DY process $pp \to \ell^+\ell^-$, where ℓ is a lepton[1], can be expressed in terms of the invariant mass $m_{\ell\ell}$, the transverse momentum $p_T^{\ell\ell}$, the rapidity $y_{\ell\ell}$ of the lepton pair, and the polar and azimuthal angles θ and ϕ of the negative charged

[1]Throughout this Chapter, we use ℓ to denote either an electron or a muon. Although electrons and muons differ from an experimental point of view, they have been treated equivalently in our theoretical discussion.

lepton in the Collins-Soper (CS) frame [74] (see Appendix 3.A for details), as

$$\begin{aligned}\frac{d\sigma}{dm_{\ell\ell}^2 dp_T^{\ell\ell} dy_{\ell\ell} d\cos\theta d\phi} &= \frac{3}{16\pi}\frac{d\sigma}{dm_{\ell\ell}^2 dp_T^{\ell\ell} dy_{\ell\ell}} \\ &\Big\{(1+\cos^2\theta) + \frac{1}{2}A_0(1-3\cos^2\theta) + A_1\sin 2\theta\cos\phi \\ &+\frac{1}{2}A_2\sin^2\theta\cos 2\phi + A_3\sin\theta\cos\phi + A_4\cos\theta \\ &+A_5\sin^2\theta\sin 2\phi + A_6\sin 2\theta\sin\phi + A_7\sin\theta\sin\phi\Big\}\ .\end{aligned} \tag{3.2}$$

The coefficients A_l are dimensionless and can be extracted from experimental data by making a fit to the angular distribution of the lepton pair. This expression is completely general and also holds in the presence of QCD and electroweak corrections.

A crucial observable for this process is $A_0 - A_2$, which, according to the so-called Lam-Tung relation [73], is expected to vanish in the Standard Model (SM) up to corrections of order $\mathcal{O}(\alpha_S)$. As a result, the SM contribution to this observable is suppressed by QCD effects at order $\mathcal{O}(\alpha_S^2)$, making it an excellent probe for potential new physics (NP) effects. Moreover, as noted in Ref. [75], the only dimension-six operators that can violate the Lam-Tung relation are the chirality-breaking ones, specifically those belonging to classes 3 and 8 discussed in the previous Section. For these reasons, $A_0 - A_2$ stands out as a distinctive observable where the contribution of such operators can be effectively isolated. This makes it a particularly suitable candidate for constraining them, especially when compared to the differential cross-section, where their effect is subleading relative to operators that do interfere with the SM. Accordingly, we propose a systematic study of the sensitivity of the LHC and HL-LHC to chirality-breaking SMEFT operators via the $A_0 - A_2$ observable.

The A_l coefficients are defined through expectation values of suitable combinations of spherical harmonics in the θ and ϕ angles, normalized to the cross-section integrated over such angles.[2] In particular, for each bin of $m_{\ell\ell}$, $p_{T,\ell\ell}$, and $y_{\ell\ell}$, one can define

$$S_l = \langle P_l(\cos\theta,\phi)\rangle = \frac{\displaystyle\int d\sigma(\cos\theta,\phi,m_{\ell\ell},p_T^{\ell\ell},y_{\ell\ell}) d\cos\theta\, d\phi\, P_l(\cos\theta,\phi)}{\displaystyle\int d\sigma(\cos\theta,\phi,m_{\ell\ell},p_T^{\ell\ell},y_{\ell\ell}) d\cos\theta\, d\phi}\ . \tag{3.3}$$

It can be verified that the $P_l(\cos\theta,\phi)$ choice that projects the amplitude onto the A_l coefficients is

[2] see the latest discussion in Refs. [76–80].

given by

$$\begin{aligned}
S_0 &= \langle P_0(\cos\theta,\phi)\rangle = \langle \frac{1}{2}\left(1-3\cos^2\theta\right)\rangle = \frac{3}{20}\left(A_0-\frac{2}{3}\right),\\
S_1 &= \langle P_1(\cos\theta,\phi)\rangle = \langle \sin 2\theta\cos\phi\rangle = \frac{1}{5}A_1,\\
S_2 &= \langle P_2(\cos\theta,\phi)\rangle = \langle \sin^2\theta\cos 2\phi\rangle = \frac{1}{10}A_2,\\
S_3 &= \langle P_3(\cos\theta,\phi)\rangle = \langle \sin\theta\cos\phi\rangle = \frac{1}{4}A_3,\\
S_4 &= \langle P_4(\cos\theta,\phi)\rangle = \langle \cos\theta\rangle = \frac{1}{4}A_4,\\
S_5 &= \langle P_5(\cos\theta,\phi)\rangle = \langle \sin^2\theta\sin 2\phi\rangle = \frac{1}{5}A_5,\\
S_6 &= \langle P_6(\cos\theta,\phi)\rangle = \langle \sin 2\theta\sin\phi\rangle = \frac{1}{5}A_6,\\
S_7 &= \langle P_7(\cos\theta,\phi)\rangle = \langle \sin\theta\sin\phi\rangle = \frac{1}{4}A_7,
\end{aligned} \tag{3.4}$$

so that the A_l coefficients can be extracted from the S_l ones as

$$\begin{aligned}
A_0 &= \frac{20}{3}S_0+\frac{2}{3}, & A_1 &= 5S_1,\\
A_2 &= 10S_2, & A_3 &= 4S_3,\\
A_4 &= 4S_4, & A_5 &= 5S_5,\\
A_6 &= 5S_6, & A_7 &= 4S_7.
\end{aligned} \tag{3.5}$$

The exact relation between the $P_l(\cos\theta,\phi)$ functions and the ordinary spherical harmonics is given in Appendix 3.B. The SM contribution to the A_l observables can be explitly written as a function of

$$S_l^{\rm SM} = \frac{\displaystyle\int d\sigma^{\rm SM}(\cos\theta,\phi,m_{\ell\ell},p_T^{\ell\ell},y_{\ell\ell})d\cos\theta\, d\phi\, P_l(\cos\theta,\phi)}{\displaystyle\int d\sigma^{\rm SM}(\cos\theta,\phi,m_{\ell\ell},p_T^{\ell\ell},y_{\ell\ell})d\cos\theta\, d\phi}\,. \tag{3.6}$$

As already mentioned, in this Chapter we do not make use of the lepton system rapidity $y_{\ell\ell}$ distribution, and we only consider the dependence of the A_0-A_2 observables either on the transverse momentum $p_T^{\ell\ell}$ or on the invariant mass $m_{\ell\ell}$ distribution of the lepton pair.

When any of the dipole, scalar, or tensor four-fermion operators, corresponding to the Wilson coefficient $c_{\rm NP}$ is switched on, we can rewrite the differential cross-section, integrated over $y_{\ell\ell}$, as

$$d\sigma(\cos\theta,\phi,m_{\ell\ell},p_T^{\ell\ell}) = d\sigma^{\rm SM}(\cos\theta,\phi,m_{\ell\ell},p_T^{\ell\ell}) + \left(\frac{c_{\rm NP}}{\Lambda^2}\right)^2 d\tilde{\sigma}^{\rm NP}(\cos\theta,\phi,m_{\ell\ell},p_T^{\ell\ell})\,, \tag{3.7}$$

where $d\tilde{\sigma}^{\rm NP}$ is, up to the NP coefficient $\left(c_{\rm NP}/\Lambda^2\right)^2$, the differential cross-section generated by the NP operator, which does not interfere with the SM. Substituting Eq. (3.7) into Eq. (3.3), we obtain

the S_l observable in the presence of NP contributions:

$$
\begin{aligned}
S_l \;=\; & \frac{\displaystyle\int d\sigma^{\rm SM}(\cos\theta,\phi,m_{\ell\ell},p_T^{\ell\ell})d\cos\theta\, d\phi\, P_l(\cos\theta,\phi)}{d\sigma^{\rm SM}(m_{\ell\ell},p_T^{\ell\ell})+\left(\frac{c_{\rm NP}}{\Lambda^2}\right)^2 d\tilde{\sigma}^{\rm NP}(m_{\ell\ell},p_T^{\ell\ell})} \\
& + \left(\frac{c_{\rm NP}}{\Lambda^2}\right)^2 \frac{\displaystyle\int d\tilde{\sigma}^{\rm NP}(\cos\theta,\phi,m_{\ell\ell},p_T^{\ell\ell})d\cos\theta\, d\phi\, P_l(\cos\theta,\phi)}{d\sigma^{\rm SM}(m_{\ell\ell},p_T^{\ell\ell})+\left(\frac{c_{\rm NP}}{\Lambda^2}\right)^2 d\tilde{\sigma}^{\rm NP}(m_{\ell\ell},p_T^{\ell\ell})},
\end{aligned}
\tag{3.8}
$$

where we made explicit the integration of the denominator over $\cos\theta$ and ϕ by omitting the corresponding variables in the argument of the differential cross-sections. The A_l observables in the presence of NP contributions can then be obtained from Eq. (3.8) by using Eq. (3.5).

In the limit $(c_{\rm NP}/\Lambda^2) \to 0$, this expression reduces to the SM one in Eq. (3.6). Moreover, we know that, in this limit, $A_0 - A_2$ is non-zero only starting from $\mathcal{O}(\alpha_{\rm S}^2)$, due to the Lam-Tung relation. This means that, to correctly account for the SM contribution to the $A_0 - A_2$ observable, we need to evaluate the SM differential cross-section at order $\mathcal{O}(\alpha_{\rm S}^2)$. Equation (3.8) also shows that generally, even though the SM and the SMEFT contributions do not interfere, their contributions to the S_l observables, and therefore also to the A_l and $A_0 - A_2$ combinations, can not be simply disentangled, since the denominator contains both the SM and the NP contributions. In other words, the behavior with respect to the Wilson coefficients $c_{\rm NP}$ of the A_l observables can be approximated as quadratic only in the limit where the NP contribution to the differential cross-section is subdominant with respect to the SM one. This may not be the case in the regions of the phase space where the SMEFT contribution is enhanced, for example at large invariant mass or transverse momentum of the lepton pair, which turn out to be the most sensitive regions to constrain the chirality-breaking operators. Therefore, in our analysis we present results obtained both with the exact relation of Eq. (3.8), which would of course be subject to corrections from operators of dimension higher than six that we do not consider, and with its quadratic approximation in $c_{\rm NP}/\Lambda^2$, which corresponds to a fixed order truncation in the SMEFT expansion. This allows one to assess the possible impact of higher-order terms in the SMEFT expansion on the bounds that we derive.

In order to compute the $p_T^{\ell\ell}$ and $m_{\ell\ell}$ dependence of the A_l observables we proceeded as follows. We estimated the SM differential cross-section for the process $pp \to \ell^+\ell^-$ at $\mathcal{O}(\alpha_{\rm S}^2)$ by generating events with the `MiNNLOPS` [81, 82] Monte Carlo generator, implemented in the `POWHEG` framework [64, 83, 84]. We have checked that the inclusion of showering effects, available in the `MiNNLOPS` framework through a matching with `PYTHIA8` [65, 85, 86] showering, does not affect the results of our analysis, and we therefore proceeded computing parton level events with a lepton–anti-lepton pair and up to two jets in the final state. Notice that, since we consider the differential $p_T^{\ell\ell}$ distribution for a finite $p_T^{\ell\ell} \geq 10$ GeV cut, we could have in principle computed the NLO QCD correction to the process $pp \to \ell^+\ell^- + j$. However, we decided to use the `MiNNLOPS` framework which allowed us to directly generate weighted events in Les Houches (lhe) format [87] and to properly check the effect of the parton showering.

The generated SM events were used to define the differential SM cross-sections $d\sigma^{\rm SM}/dp_T^{\ell\ell}$ within a window 80 GeV $< m_{\ell\ell} <$ 100 GeV and $d\sigma^{\rm SM}/dm_{\ell\ell}^2$ in the region $p_T^{\ell\ell} > 10$ GeV, by binning the events in the relevant kinematic variable.

In order for the differential cross-section to be defined at finite $p_T^{\ell\ell}$, the SMEFT contribution was computed analytically by considering the process $pp \to \ell^+\ell^- + j$ at the leading order in QCD. The partonic cross-sections for all sub-processes have been calculated using the `FeynCalc` [88] and `FeynArts` [89] packages in `Mathematica` [90], after truncating the squared amplitude at $\mathcal{O}(1/\Lambda^4)$, which is equivalent to considering a single NP insertion. The hadronic cross-section was then derived by convoluting these results with the PDFs. For this purpose, we employed the `ManeParse` package [91] for `Mathematica`, which provides all the necessary tools for PDF integration. We adopted the PDF set `NNPDF31_nnlo_as_0118` (ID 303600 in the LHAPDF [70] set), neglecting contributions from the b and t quarks. Details about the analytical expression of the integral can be found in appendix 3.C. This procedure allowed us to compute the differential cross-sections $d\tilde{\sigma}^{\rm NP}/dp_T^{\ell\ell}$ and $d\tilde{\sigma}^{\rm NP}/dm_{\ell\ell}^2$ in the same kinematic region as the SM ones, and to use them in Eq. (3.8) to compute the S_l and A_l observables in the presence of NP contributions. The evaluation of Eq. (3.8) also required the calculation of the projections of the cross-section onto the corresponding angular polynomials $P_l(\cos\theta, \phi)$. This was done by Monte Carlo integration for the SM, using the generated events, and by numerical integration of the analytic expression for the NP contributions.

3.1.1 Monte Carlo simulation and uncertainty estimation

The angular observables defined in the previous Section can be predicted theoretically by Monte Carlo estimation of the integrals. Given a sample of weighted events i in bin I with weights w_i, the cross-section in that bin is given by

$$\sigma_I = \sum_{i \in I} w_i \,. \tag{3.9}$$

The angular observables in bin I are then defined from the projections of the cross-section in that bin on the corresponding polynomial functions $P_l(\cos\theta, \phi)$, which appear in the numerator of Eq. (3.3). We can compute this by multiplying the MC weights w_i by the value of the relevant polynomial functions $P_l^i(\cos\theta, \phi)$, given in Eq. (3.5), computed for the i-th event kinematic, and by defining a new set of angular weights

$$w_i^{(l)} = w_i P_l^i(\cos\theta, \phi) \,. \tag{3.10}$$

Then, the differential cross-section projected over the polynomial functions $P_l(\cos\theta, \phi)$, that we simply denote by $\sigma_I^{(l)}$, is given by

$$\sigma_I^{(l)} = \sum_{i \in I} w_i^{(l)} \,. \tag{3.11}$$

This projection can be used to estimate the value of the $S_{l,I}$ observables of Eq. (3.4) in each bin I

$$S_{l,I} = \frac{\sigma_I^{(l)}}{\sigma_I} \tag{3.12}$$

and, in turn, to compute the $A_{l,I}$ observables through Eq. (3.5).

Extracting a meaningful bound on the new physics contribution from the differential angular observables A_l, and in particular from the $A_0 - A_2$ combination, requires a careful estimate of the expected experimentally measured values and uncertainties of the observables, especially for what concerns the SM contribution. Given the very large statistics available at the LHC for the DY process, at relatively low $p_T^{\ell\ell}$ and for di-lepton invariant masses $m_{\ell\ell}$ not far above the Z boson mass, the uncertainty is dominated by experimental systematics. Since no recent analysis of the multi-differential DY cross-section is available, and based on the existing experimental results, we decided to make the simplifying assumption of a flat, uncorrelated, 3% uncertainty on the differential cross-section in all bins.

Concerning the statistical uncertainty, which becomes dominant in corner regions of the phase space, such as very large $p_T^{\ell\ell}$ and/or $m_{\ell\ell}$ much above the Z boson mass, we generated enough statistics to cover the expected experimental statistics at the LHC with $\mathcal{L} = 300$ fb^{-1},[3] which is the target integrated luminosity of the LHC and extrapolated with a factor of square root of ten to the HL-LHC with $\mathcal{L} = 3$ ab^{-1}. The central values were kept unchanged in this extrapolation.

In order to estimate the central values and the statistical uncertainties in each bin of the two distributions, we have proceeded as follows. We denoted with $n_I^{\rm MC}$ the available number of MC-generated events in each bin I and with

$$\sigma_I^{\rm MC} = \sum_{i \in I} w_i \tag{3.13}$$

the MC-estimated cross-section in that bin. We then fixed an initial value for the number of events corresponding to the experimental observation with a given integrated luminosity $\mathcal{L}$

$$n_I^{\rm exp} = \sigma_I^{\rm MC} \times \mathcal{L}\,. \tag{3.14}$$

To estimate the expected value and the standard deviation of the differential cross-section and of the angular observables in each bin (of the relevant kinematic distribution), we simulated pseudo-experiments, each corresponding to the statistics expected at the LHC with $\mathcal{L} = 300$ fb^{-1}.[4] Pseudo experiments were drawn from the available MC sample with replacement.[5] In order to fix the correct value of the cross-section, for each pseudo-experiment we rescaled the value of the corresponding weights. In order to also introduce a fluctuation in the normalization, and not only in the shape, we also considered fluctuations in the normalization, computed from the original MC sample. Ob-

[3] This required the generation of about a billion events for the $p_T^{\ell\ell}$ distribution, which is at the edge of what we could afford with our computing resources. The number of events generated for the $m_{\ell\ell}$ distribution was substantially lower, thanks to the ability of the MiNNLOPS generator to cut on the di-lepton invariant mass.

[4] The result for $\mathcal{L} = 3$ ab^{-1} has been obtained assuming that the variance scales with the ratio of the number of events and therefore scaling the error with a square root of ten.

[5] When the available MC sample is smaller or roughly equal in size than the expected statistics, such as in low $p_T^{\ell\ell}$ bins, our procedure may underestimate the standard deviation. On the other hand these are the bins where the expected statistical uncertainty is much smaller than the expected systematic uncertainty, so that the effect of such underestimation becomes negligible.

viously, fluctuations on the normalization only affect the cross-section estimate and not the angular coefficients, which are built from ratios. The detailed procedure is outlined in the following:

- For each pseudo-experiment p and bin I, we consider a Poisson fluctuation of the number of MC events
$$n_I^{\mathrm{MC},(p)} = \mathrm{Poisson}(\mu = n_I^{\mathrm{MC}})\,, \tag{3.15}$$
and draw, with replacement, $n_I^{\mathrm{MC},(p)}$ events from the available MC events. We use this sample to estimate a fluctuation of the cross-section in that bin, that we denote by $\sigma_I^{\mathrm{MC},(p)}$, and that we use as fluctuating normalization of the cross-section in that bin.

- For each pseudo-experiment p and bin I, we consider a Poisson fluctuation of the number of experimental events
$$n_I^{\mathrm{exp},(p)} = \mathrm{Poisson}(\mu = n_I^{\mathrm{exp}})\,, \tag{3.16}$$
and draw, with replacement, $n_I^{\mathrm{exp},(p)}$ from the available MC events.

- For each pseudo-experiment p and bin I, we use the corresponding $n_I^{\mathrm{exp},(p)}$ events to define a new set of normalized weights
$$w_i^{\mathrm{exp},(p)} = \kappa_I^{(p)} w_i^{(p)} \qquad \forall i \in I\,, \tag{3.17}$$
with normalization
$$\kappa_I^{(p)} = \frac{\sigma_I^{\mathrm{MC},(p)}}{\sum_{i\in I} w_i^{(p)}}\,. \tag{3.18}$$
Notice that this definition does automatically set the corresponding cross-section to its correct fluctuating normalization:
$$\sigma_I^{\mathrm{exp},(p)} \equiv \sum_{i\in I} w_i^{\mathrm{exp},(p)} = \kappa_I^{(p)} \sum_{i\in I} w_i^{(p)} = \sigma_I^{\mathrm{MC},(p)}\,. \tag{3.19}$$

- For each pseudo-experiment p and bin I, we use the corresponding $n_I^{\mathrm{exp},(p)}$ events to define a new set of normalized angular weights, analog to those in Eq. (3.10),
$$w_i^{\mathrm{exp},(l),(p)} = \kappa_I^{(p)} w_i^{(p)} P_l^{i,(p)}(\cos\theta, \phi) \qquad \forall i \in I\,. \tag{3.20}$$

- For each pseudo-experiment p and bin I, we use the new set of normalized angular weights to estimate the cross-section angular projections as
$$\sigma_I^{\mathrm{exp},(l),(p)} = \sum_{i\in I} w_i^{\mathrm{exp},(l),(p)} = \kappa_I^{(p)} \sum_{i\in I} w_i^{(p)} P_l^{i,(p)}(\cos\theta, \phi)\,. \tag{3.21}$$

- For each pseudo-experiment p and bin I, we use Eqs. (3.19) and (3.20) to compute the S_l's

obsevables as

$$S_{l,I}^{\mathrm{exp},(p)} = \frac{\sigma_I^{\mathrm{exp},(l),(p)}}{\sigma_I^{\mathrm{exp},(p)}} = \frac{\sum_{i\in I} w_i^{(p)} P_l^{i,(p)}(\cos\theta,\phi)}{\sum_{i\in I} w_i^{(p)}}\,, \tag{3.22}$$

which is, as expected, independent on the cross-section normalization $\kappa_I^{(p)}$.

- We estimate the central values and (one dimensional) covariance matrix of the differential cross-section as

$$\mu_{\sigma_I} = \mathrm{Mean}_{p\in\mathcal{P}}\left(\sigma_I^{\mathrm{MC},(p)}\right)\,, \qquad (\Sigma_{\sigma_I,\sigma_I})^{\mathrm{stat}} = \mathrm{Var}_{p\in\mathcal{P}}\left(\sigma_I^{\mathrm{MC},(p)}\right)\,, \tag{3.23}$$

  where we denoted with $\mathrm{Mean}_{p\in\mathcal{P}}$ and $\mathrm{Var}_{p\in\mathcal{P}}$ the mean and variance computed over a set of 100 pseudo-experiments.

- We estimate the central values and covariance matrix of the S_l's observables as

$$\mu_{S_{l,I}} = \mathrm{Mean}_{p\in\mathcal{P}}\left(S_{l,I}^{\mathrm{exp},(p)}\right)\,, \qquad (\Sigma_{S_{l,I},S_{m,I}})^{\mathrm{stat}} = \mathrm{Cov}_{p\in\mathcal{P}}\left(S_{l,I}^{\mathrm{exp},(p)}, S_{m,I}^{\mathrm{exp},(p)}\right)\,, \tag{3.24}$$

  where we denoted with $\mathrm{Mean}_{p\in\mathcal{P}}$ and $\mathrm{Cov}_{p\in\mathcal{P}}$ the mean and covariance matrix computed over a set of 100 pseudo-experiments.

- Analogously, we estimate the central values and covariance matrix of the A_l observables through Eq. (3.5).

Once the central values $\mu_{\mathcal{O}}$ and covariance matrix $(\Sigma_{\mathcal{O},\mathcal{O}'})^{\mathrm{stat}}$ of the observables $\mathcal{O}$ are computed, we combine the uncertainties with the corresponding quantities parametrizing the systematic uncertainty $(\Sigma_{\mathcal{O},\mathcal{O}'})^{\mathrm{syst}}$ and get the final estimate of the observables as

$$\mu_{\mathcal{O}} \pm \sqrt{\Sigma_{\mathcal{O},\mathcal{O}}} \qquad \text{with} \qquad \Sigma_{\mathcal{O},\mathcal{O}'} = \Sigma^{\mathrm{stat}}_{\mathcal{O},\mathcal{O}'} + \Sigma^{\mathrm{syst}}_{\mathcal{O},\mathcal{O}'}\,. \tag{3.25}$$

3.2 Experimental status

The measurements relevant for the present Chapter fall in the realm of precision measurements for new physics searches.[6] Such precision measurements usually take a long term experimental effort and this is why not many such measurements are already available from LHC Run 2 and Run 3. Here we briefly summarize the most relevant existing measurements in the Drell-Yan channel, and use the information we can gather from them to motivate our assumptions on the uncertainties discussed in the previous Section. Table 3.1 lists the measurements of differential cross-sections and angular coefficients in di-lepton final states performed over the past decade.

As it can immediately be seen from the table, only a single measurement of the angular coefficients from each experiment ATLAS, CMS, and LHCb is available. The first two measurements

[6] We omitted LHCb measurements of the inclusive cross-section since they generally have less statistics. Obviously, in the case in which the forward region becomes more relevant the LHCb measurements become important.

Collab.	Year	Energy	Luminosity	Observable	Ref.
ATLAS	2024	5.02, 13 TeV	$255, 338$ pb^{-1}	$d\sigma_{\ell\ell}/dp_{T,ll}$	[92]
ATLAS	2024	8 TeV	20.2 fb^{-1}	$d\sigma_{\ell\ell}/dp_{T,ll}dy_{\ell\ell}$	[93]
CMS	2023	13 TeV	36.3 fb^{-1}	$d\sigma_{\ell\ell}/dp_{T,ll}dm^2_{\ell\ell}$, $d\sigma_{\ell\ell}/d\phi^*_\eta dm^2_{\ell\ell}$	[94]
LHCb	2022	13 TeV	5.1 fb^{-1}	$A_0, \ldots, A_7$ as functions of $p_{T,ll}$	[95]
ATLAS	2020	13 TeV	36.1 fb^{-1}	$d\sigma_{\ell\ell}/dp_{T,ll}$, $d\sigma_{\ell\ell}/d\phi^*_\eta$	[96]
CMS	2019	13 TeV	35.9 fb^{-1}	$d\sigma_{\ell\ell}/dp_{T,ll}$, $d\sigma_{\ell\ell}/d|y_{\ell\ell}|$, $d\sigma_{\ell\ell}/d\phi^*_\eta$	[97]
CMS	2019	13 TeV	$2.8^{(\mu)}/2.3^{(e)}$ fb^{-1}	$d\sigma_{\ell\ell}/dm^2_{\ell\ell}$	[98]
CMS	2018	8 TeV	19.7 fb^{-1}	$d\sigma_{\ell\ell}/d\phi^*_\eta$, $d\sigma_{\ell\ell}/d\phi^*_\eta d|y_{\ell\ell}|$	[99]
ATLAS	2017	8 TeV	20.2 fb^{-1}	$d\sigma_{\ell\ell}/dm^2_{\ell\ell}d|y_{\ell\ell}|d\cos\theta^*$	[100]
CMS	2017	8 TeV	18.4 fb^{-1}	$d\sigma_{\ell\ell}/dp_{T,ll}$	[101]
ATLAS	2016	8 TeV	20.3 fb^{-1}	$d\sigma_{\ell\ell}/dm^2_{\ell\ell}d|y_{\ell\ell}|$, $d\sigma_{\ell\ell}/dm^2_{\ell\ell}d|\Delta\eta_{\ell\ell}|$	[102]
ATLAS	2016	8 TeV	20.3 fb^{-1}	$A_0, \ldots, A_7$ as functions of $p_{T,ll}$	[103]
ATLAS	2016	8 TeV	20.3 fb^{-1}	$d\sigma_{\ell\ell}/dp_{T,ll}$, $d\sigma_{\ell\ell}/d\phi^*_\eta$	[104]
CMS	2015	8 TeV	19.7 fb^{-1}	$A_0, \ldots, A_7$ as functions of $p_{T,ll}$	[105]
CMS	2015	8 TeV	19.7 fb^{-1}	$d\sigma_{\ell\ell}/dp_{T,ll}$, $d\sigma_{\ell\ell}/d|y_{\ell\ell}|$	[106]

Table 3.1: Summary of the possibly relevant experimental measurements of Drell-Yan observable over the last ten years. For ease of reading, hovering over the info icon shows the bibliography item.

are at 8 TeV and an integrated luminosity of around 20 fb^{-1}, while the LHCb measurement in the forward region is the only measurement of the angular coefficient to date at an energy of 13 TeV, with an integrated luminosity of 5.1 fb^{-1}. No updated measurements of the angular coefficients have yet been performed at 13 TeV, where the LHC has collected a much larger integrated luminosity. The measurements of the differential cross-section are more abundant, with different differential distributions measured at different energies and integrated luminosities. However, even the most recent measurements at 13 TeV are performed with a limited integrated luminosity of approximately 36 fb^{-1}, which is only a small fraction of the total integrated luminosity collected at 13 TeV during Run 2 (around 160 fb^{-1}).

Both the CMS [105] and ATLAS [103] measurements of the angular coefficients at 8 TeV are performed in the Z-boson mass peak region, and, to take into account the effect of the finite acceptance on the leptons, which affects the angular distributions in Eq. (3.2), are performed through matching with Monte Carlo templates of the different angular observables in the θ-ϕ plane. Entering in the details of the experimental measuremets is beyond the scope of this study, and we refer the reader to the original experimental papers for more details. Here, we are only interested in understanding the expectation about the measurement uncertainties, which we need to consider for a realistic projection of the sensitivity to new physics in the angular coefficients.

From Tables 11 to 14 of Ref. [103], we can see that the statistical and systematic uncertainties on the $p_T^{\ell\ell}$ distribution of the A_0, A_2, and $A_0 - A_2$ observables are comparable over the whole spectrum, but in the last few bins, where, obviously, the statistical uncertainty dominates. Even though most of the systematic uncertainties, such as Monte Carlo statistics, have large margins for improvement, it is reasonable to expect that, with the full LHC and HL-LHC integrated luminosity, the systematic uncertainties will be dominant in the whole spectrum, but the last few bins, where

the statistical uncertainty will still be dominant. We have already discussed in the previous Section how we model the statistical uncertainty through pseudo-experiments simulation. Concerning the systematic uncertainty, assuming a projected uncertainty directly on the A_l coefficients from the aforementioned angular analyses is not possible, since we can not guess how this will improve in the future. However, we can take the expected systematic uncertainties on the differential cross-section measurements from recent 13 TeV analyses as a guideline. We consider in particular the ATLAS analysis of Ref. [96] as reference analysis of the $p_T^{\ell\ell}$ spectrum and the CMS analysis of Ref. [98] as reference analysis of the $m_{\ell\ell}$ spectrum. We adopt the same binning for the two distributions and use the experience of those analyses to motivate our assumption on the systematic uncertainties on the differential cross-sections. Starting from such uncertainties, and assuming the same uncertainties for the cross-sections projections on the different angular polynomials, we propagate the uncertainty to the A_l coefficients. The result is consistent with a systematic uncertainty that dominates in the low $p_T^{\ell\ell}$ and $m_{\ell\ell}$ region (above the Z boson mass) and becomes subleading in the tails of the distributions.

In the next Section, we show projections for the expected measurements of the $p_T^{\ell\ell}$ and $m_{\ell\ell}$ distributions, as well as the angular coefficients.

3.3 Standard Model predictions

In this Section, we present SM predictions for the $p_T^{\ell\ell}$ and $m_{\ell\ell}$ distributions of the cross-section and angular observables in the Drell-Yan process. Such predictions will be combined with predictions of the effect of new physics and used to derive projected limits on the chirality-breaking operator of Eq. (3.1) in the next Section.

3.3.1 Transverse momentum distribution

The left panel of Figure 3.1 shows the SM prediction for the di-lepton $p_T^{\ell\ell}$ spectrum in the DY process at the LHC at 13 TeV, assuming a flat, uncorrelated systematic uncertainty of 3%, corresponding, for instance, to a 2% systematic from the integrated luminosity measurement and about 2% additional systematic uncertainty, including missing higher orders (scale variation), PDFs, Monte Carlo, and experimental uncertainties. These numbers for the systematic uncertainties are consistent with those reported in the aforementioned ATLAS analysis of Ref. [96], at least for the low and intermediate $p_T^{\ell\ell}$ region. The high $p_T^{\ell\ell}$ region has much larger systematic uncertainties, which are clearly statistically dominated. We do not have a clear prescription to project this uncertainty to the future measurements, and therefore we assume the flat 3% uncertainty also for the high $p_T^{\ell\ell}$ region. Statistical and systematic uncertainties are added in quadrature, and the statistical uncertainty is derived through pseudo-experiments as explained in the previous Section. Uncertainties are too small to be visible in the plot, so the result is visually the same for both 300 fb^{-1} and 3 ab^{-1} of integrated luminosity. The exact numerical values (both central values and uncertainties) corresponding to 300 fb^{-1} and 3 ab^{-1} of integrated luminosity are given in Table 3.7 of Appendix 3.E.

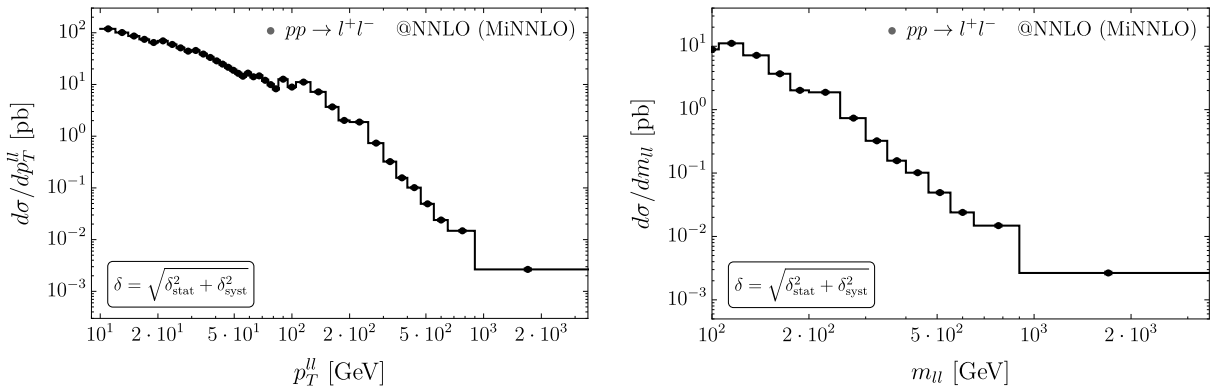


Figure 3.1: Standard Model prediction for the $p_T^{\ell\ell}$ (left panel) and $m_{\ell\ell}$ (right panel) spectra in the DY process at the LHC at 13 TeV with 300 fb^{-1}. Uncertainties are too small to be visible, so the result is visually the same for both 300 fb^{-1} and 3 ab^{-1} of integrated luminosity. The exact numerical values (both central values and uncertainties) corresponding to 300 fb^{-1} and 3 ab^{-1} of integrated luminosity are given in Tables 3.7 and 3.11 of Appendix 3.E.

Figure 3.2 shows the SM prediction for the angular coefficients A_0, A_2, and $A_0 - A_2$ as functions of $p_T^{\ell\ell}$ in the DY process at the LHC at 13 TeV with two assumptions for the statistical uncertainty: the larger error bars represent the combined statistical and systematic uncertainty with 300 fb^{-1} of integrated luminosity, while the smaller error bars represent the combined statistical and systematic uncertainty with 3 ab^{-1} of integrated luminosity. In both cases, we assume a flat, uncorrelated systematic uncertainty of 3% on the measurement of the cross-section and of its angular projections (denoted as $\sigma_I^{(l)}$ in Section 3.1) and propagate it to the angular coefficients as explained in the previous Section. As before, statistical and systematic uncertainties are added in quadrature, and the statistical uncertainty is derived through pseudo-experiments as explained in the previous Section.

Notice that the systematic uncertainty on the A_0 observable is much smaller than that on the A_2 observable, so that, even including correlation among them, the combined systematic uncertainty on $A_0 - A_2$ is not largely affected, and remains dominated by the uncertainty on A_2. For this reason we only show the result under the assumption of no correlation among the systematic uncertainties of the different angular coefficients.

All the numbers of the cross-sections and angular coefficients predictions shown in Figures 3.1 (left) and 3.2 are summarized in Tables 3.7, 3.8, 3.9, and 3.10 of Appendix 3.E.

3.3.2 Invariant mass distribution

The right panel of Figure 3.1 shows the SM prediction for the di-lepton $m_{\ell\ell}$ spectrum in the DY process at the LHC at 13 TeV, assuming, as in the case of the $p_T^{\ell\ell}$ distribution, a flat, uncorrelated systematic uncertainty of 3% (see above). These numbers for the systematic uncertainties are consistent with those reported in the aforementioned CMS analysis of Ref. [98], at least for systematic uncertainties that are not statistically dominated. In the high $m_{\ell\ell}$ region, we do not have a clear procedure to project the systematic uncertainty to the future measurements, and therefore we as-

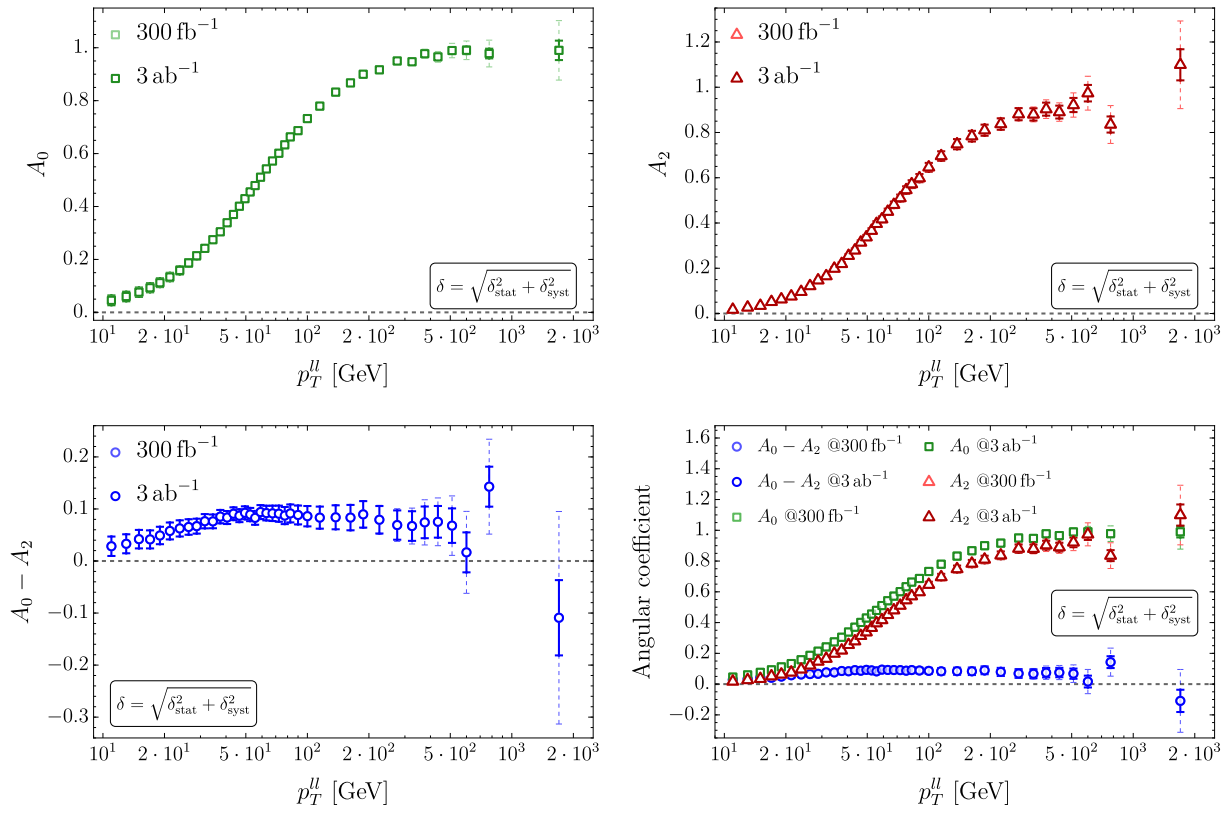


Figure 3.2: Standard Model prediction for the $p_T^{\ell\ell}$ dependence of the angular observables A_0 (upper left), A_2 (upper right), and $A_0 - A_2$ (lower left) in the DY process at the LHC at 13 TeV with 300 fb^{-1} and 3 ab^{-1} both inclusive in rapidity. The lower right panel gives a combined view of the result. The numerical values of the observables appearing in the plots are given in Tables 3.8, 3.9, and 3.10 of Appendix 3.E.

sume the flat 3% uncertainty also in that region. As before, statistical and systematic uncertainties are added in quadrature, the statistical uncertainty is derived through pseudo-experiments as explained in the previous Section, and uncertainties are too small to be visible in the plot, so the result is visually the same for both 300 fb^{-1} and 3 ab^{-1} of integrated luminosity. The exact numerical values (both central values and uncertainties) corresponding to 300 fb^{-1} and 3 ab^{-1} of integrated luminosity are given in Table 3.11 of Appendix 3.E.

Figure 3.3 shows the SM prediction for the angular coefficients A_0, A_2, and $A_0 - A_2$ as functions of $m_{\ell\ell}$ in the DY process at the LHC at 13 TeV with two assumptions for the statistical uncertainty: the larger error bars represent the combined statistical and systematic uncertainty with 300 fb^{-1}, while the smaller error bars represent the combined statistical and systematic uncertainty with 3 ab^{-1}. In both cases, we assume a flat, uncorrelated systematic uncertainty of 3% on the measurement of the cross-section and of its angular projections (denotes as $\sigma_I^{(l)}$ in Section 3.1) and propagate it to the angular coefficients as explained in the previous Section. As before, statistical and systematic uncertainties are added in quadrature, and the statistical uncertainty is derived through pseudo-

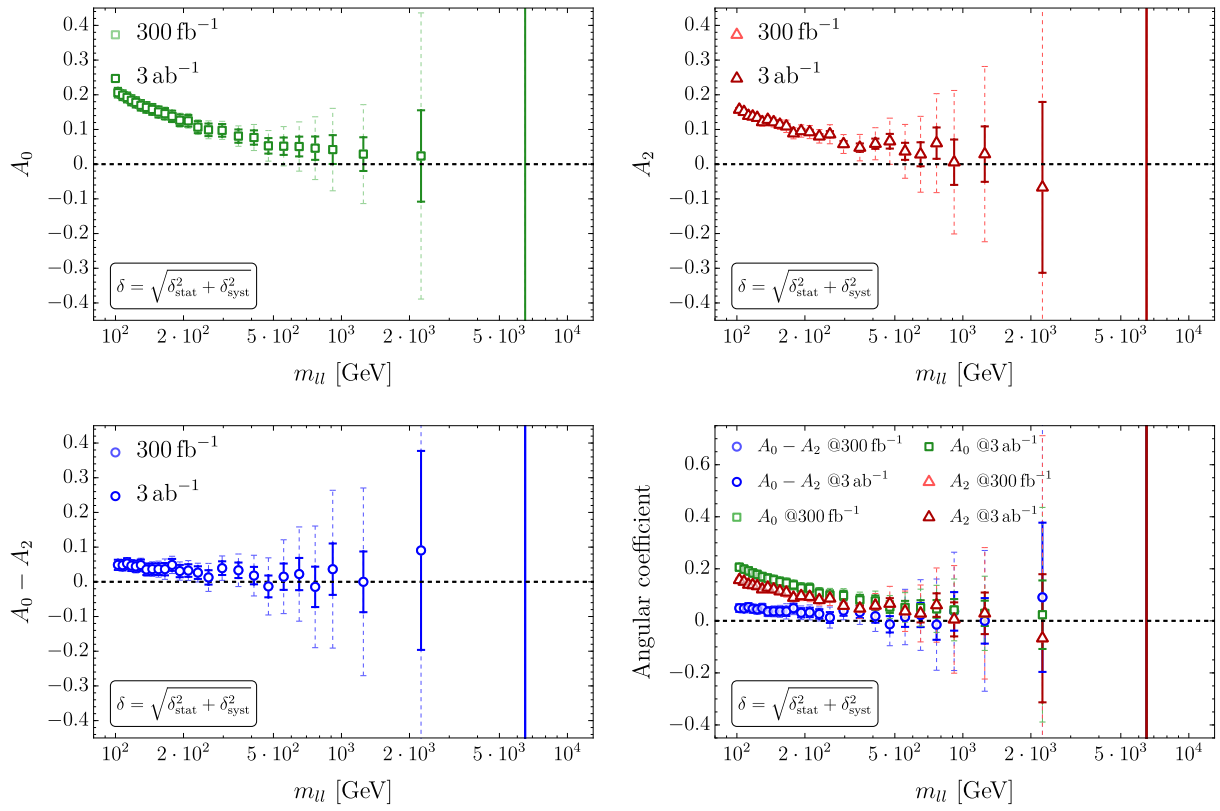


Figure 3.3: Standard Model prediction for the $m_{\ell\ell}$ dependence of the angular observables A_0 (upper left), A_2 (upper right), and $A_0 - A_2$ (lower left) in the DY process at the LHC at 13 TeV with 300 fb^{-1} and 3 ab^{-1} both inclusive in rapidity. The lower right panel gives a combined view of the result. The numerical values of the observables appearing in the plots are given in Tables 3.12, 3.13, and 3.14 of Appendix 3.E.

experiments as explained in the previous Section.

Notice that, also in this case, the systematic uncertainty on the A_0 observable is much smaller than that on the A_2 observable, so that, even including correlation among them, the combined systematic uncertainty on $A_0 - A_2$ is not largely affected, and remains dominated by the uncertainty on A_2. For this reason we only show the result under the assumption of no correlation among the systematic uncertainties of the different angular coefficients.

All the numbers of the cross-sections and angular coefficients projections shown in Figures 3.1 (right) and 3.3 are summarized in Tables 3.11, 3.12, 3.13, and 3.14 of Appendix 3.E.

3.4 SMEFT predictions

In this Section we analyze the effects of the dimension-6 operators introduced in Section **??** on the $A_0 - A_2$ observable. In particular, we argue that the dipoles and scalar/tensor four-fermion operators are the only dimension-6 operators that can break the Lam-Tung relation at order $\mathcal{O}(\alpha_S^0)$,

that is when QCD corrections are ignored. This remains true also at order $\mathcal{O}(\alpha_{\rm S})$, motivating our focus on just these two classes of operators. From now on, we omit the flavor index in the Wilson coefficients, and it is understood that all predictions and constraints hold separately, and identically, for electron and muons.

3.4.1 Lam-Tung relation breaking at $\mathcal{O}(\alpha_{\rm S}^0)$

At the zero-th order in $\alpha_{\rm S}$, and under the assumption of negligible intrinsic transverse momentum of the partons and of negligible effects from the possible QED radiation, the transverse momentum of the lepton pair vanishes. As a consequence, in the di-lepton rest frame the parton beams are collinear, and, in turn, the $\hat{z}$ axis of the CS frame (see Appendix 3.A for its definition) lies in the direction of the beams and can be identified with the beam axis, denoted by z. Therefore, the process features azimuthal symmetry, the CS frame can be identified with the center of mass frame (the ϕ angle is arbitrary), and the angular distribution of the leptons in such frame only depends on the angle θ, that is the angle between the negatively charged lepton and the z axis.

The fully differential cross-section of $pp \to \ell^+\ell^- + X$ in the CS frame in the SM is given by Eq. (3.2). This five-differential cross-section reduces, at the zero-th order in $\alpha_{\rm S}$, to the triple differential cross-section for the process $pp \to \ell^+\ell^-$. Integrating over the arbitrary angle ϕ we can write

$$\int_0^{2\pi} d\phi \lim_{\alpha_S \to 0} \left(\frac{d\sigma}{dm_{\ell\ell}^2 dp_T^{\ell\ell} dy_{\ell\ell} d\cos\theta d\phi} \right) = \left(\frac{d\sigma}{dm_{\ell\ell}^2 dy_{\ell\ell} d\cos\theta} \right)_{\rm LO} . \tag{3.26}$$

Since, at $\mathcal{O}(\alpha_S^0)$, Eq. (3.2) only depends on θ and not on ϕ, that means that it should hold for arbitrary values of ϕ, then all the terms proportional to functions of ϕ must vanish, implying that only the coefficients A_0 and A_4 can be non-zero in this limit. Therefore, we can formally write the limit in Eq. (3.26) as[7]

$$\begin{aligned} \int_0^{2\pi} d\phi \lim_{\alpha_S \to 0} \left(\frac{d\sigma}{dm_{\ell\ell}^2 dp_T^{\ell\ell} dy_{\ell\ell} d\cos\theta d\phi} \right) \;=\; & \frac{3}{8} \left(\frac{d\sigma}{dm_{\ell\ell}^2 dy_{\ell\ell}} \right)_{\rm LO} \Big[(1+\cos^2\theta) \\ & + \frac{1}{2} A_0 (1 - 3\cos^2\theta) + A_4 \cos\theta \Big] . \end{aligned} \tag{3.27}$$

It is important to notice that Eq. (3.27) does not imply that A_0 and A_4 are both non-vanishing in the leading order limit.

Table 3.2 summarizes the angular dependence of the squared amplitude at leading order in α_S arising from the SM and the different dimension-6 operators that we consider. In particular, we immediately see that the contributions proportional to a linear combination of $(1-\cos\theta)^2$ and $(1+\cos\theta)^2$ can only arise from Eq. (3.27) if $A_0 = 0$. This is the Lam-Tung relation at order $\mathcal{O}(\alpha_S^0)$:

$$\lim_{\alpha_S \to 0} (A_0 - A_2) = A_0 = A_2 = 0. \tag{3.28}$$

[7] The additional factor $1/(2\pi)$ appearing in Eq. (3.2) has been absorbed by the integral over ϕ.

	SM	$\mathcal{O}_{(3)}$	$\mathcal{O}_{(1,2,4)}$	$\mathcal{O}_{(5,6,7)}$	$\mathcal{O}_{(8)}$
SM	$(1 \pm \cos\theta)^2$				
$\mathcal{O}_{(3)}$	0	$1 - \cos^2\theta$			
$\mathcal{O}_{(1,2,4)}$	$(1 \pm \cos\theta)^2$	0	$(1 \pm \cos\theta)^2$		
$\mathcal{O}_{(5,6,7)}$	$(1 \pm \cos\theta)^2$	0	0	$(1 \pm \cos\theta)^2$	
$\mathcal{O}_{(8)}$	0	0	0	0	1

Table 3.2: Dependence on the $\cos\theta$ angular variable of the squared amplitude of the process $pp \to \ell^+\ell^-$ (at leading order in α_S) arising from the interference of the SM and the SMEFT operators considered in our analysis. The notation $(1 \pm \cos\theta)^2$ is used to indicate that the angular dependence is a linear combination of both $(1 + \cos\theta)^2$ and $(1 - \cos\theta)^2$ terms. The operators are labeled as in Section **??**: $\mathcal{O}_{(1,2)}$ are the operators $\phi^4 D^2$ and $\psi^2\phi^3$, $\mathcal{O}_{(3)}$ are the dipole operator $\psi^2 X\varphi$, $\mathcal{O}_{(4)}$ are the operators $\psi^2\varphi^2 D$, $\mathcal{O}_{(5,6,7)}$ are the current-current four-fermion operators, and $\mathcal{O}_{(8)}$ are the scalar/tensor four-fermion operators. . The table is symmetric and we only filled the lower part. The two shaded entries highlight the cases where the Lam-Tung relation is violated.

Table 3.2 also shows that the only operators that can break this relation at order $\mathcal{O}(\alpha_S^0)$ are the dipole operators $\mathcal{O}_{(3)}$ and the chirality-breaking scalar/tensor four-fermion operators $\mathcal{O}_{(8)}$ with angular dependences respectively proportional to $1 - \cos^2\theta$ and to a constant term.

3.4.2 Lam-Tung relation breaking at $\mathcal{O}(\alpha_S)$

At $\mathcal{O}(\alpha_S)$ the SM does not violate the Lam-Tung relation as a consequence of the fact that gluons couple to the vector quark current, as explained in details in Ref. [107]. This implies that four-fermion operators involving the vector quark current (namely $\mathcal{O}_{(5,6,7)}$, following the notation in Table 3.2) can not violate the Lam-Tung relation. As already mentioned in Section **??**, the effect of the operators $\mathcal{O}_{(1,2,4)}$ does not grow with energy and can safely be neglected.[8] Thus, even at order $\mathcal{O}(\alpha_S)$, the only operators expected to give an observable Lam-Tung breaking effect are the dipole and the scalar/tensor four-fermion operators. The $\mathcal{O}(\alpha_S)$ contributions of these operators to the $A_0 - A_2$ observable is compared to the first non-vanishing SM contribution (corresponding to $\mathcal{O}(\alpha_S^2)$) in Figure 3.4. The SM curves are the same as those shown in Figures 3.2 and 3.3, while the SMEFT contributions are computed at order $\mathcal{O}(\alpha_S)$, for Wilson coefficients fixed to the 3000 fb^{-1} (positive) exclusion bound reported in Tables 3.3 and 3.4 (see Section 3.5 for details on how these bounds are derived).

3.5 Analysis and projected limits

In this Section, we detail the procedure used to extract the projected constraints on the Wilson coefficient appearing in Eq. (3.1), and we specify the hypotheses underlying our analysis.

We considered, at the LHC, a center-of-mass energy of 13 TeV and two benchmark integrated luminosities: 300 fb^{-1}, corresponding approximately to the dataset expected by the end of Run 3,

[8] One could expect that contributions to the Drell–Yan process from the s-channel Higgs boson exchance could break Lam-Tung already at $\mathcal{O}(\alpha_S^0)$. However, even if they did, these contributions are too small to be observed.

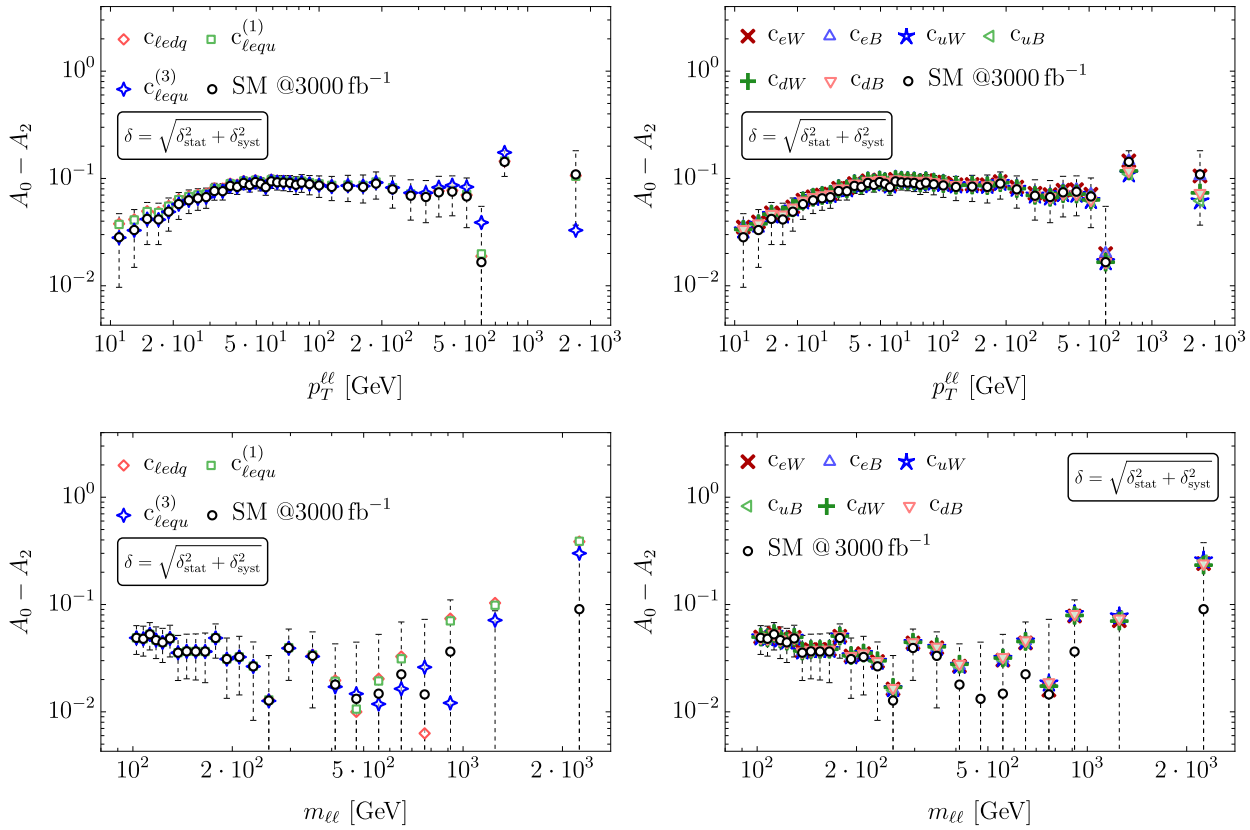


Figure 3.4: Comparison between the contribution of the different SMEFT operators to the transverse momentum (first row) and invariant mass distribution (second row) of the $A_0 - A_2$ observable. The left column shows scalar/tensor four-fermion operators, the right one shows the dipole operators. For each distribution, the Wilson coefficients are fixed equal to the largest positive value consistent with the bounds from the corresponding kinematic distribution at 3000 fb^{-1} (Table 3.3 for $p_T^{\ell\ell}$ and Table 3.4 for $m_{\ell\ell}$).

and 3000 fb^{-1}, representative of the High-Luminosity LHC (HL-LHC) scenario. Since measurements of $A_0 - A_2$ at 13 TeV are not yet available, we rely on simulations to obtain pseudo-data, as specified in Section 3.1. These simulations are performed under the SM-only hypothesis for the process $pp \to \ell^+\ell^- + X$ at next-to-next-to-leading order (NNLO) in QCD, *i.e.*, at $\mathcal{O}(\alpha_S^2)$, using the `MiNNLOPS` framework. The resulting predictions are treated as pseudo-data for the purpose of our projections, with uncertainties estimated as explained in Section 3.1.1. Predictions for the $A_0 - A_2$ observable in the SMEFT framework have been calculated analytically as explained in Section 3.1 and are expressed as functions of $c^2_{\rm NP}/\Lambda^4$, where $c^2_{\rm NP}$ generically denotes the square of a Wilson coefficient of those appearing in Eq. (3.1).

The $A_0 - A_2$ observable in the presence of NP contributions can be expressed as:

$$(A_0 - A_2)^{\text{SMEFT}} = 4 - 10\left[\frac{\int d\sigma^{\text{SM}}(\cos\theta, \phi, m_{\ell\ell}, p_T^{\ell\ell})\, d\cos\theta\, d\phi\, (\cos^2\theta + \sin^2\theta\cos 2\phi)}{d\sigma^{\text{SM}}(m_{\ell\ell}, p_T^{\ell\ell}) + \left(\frac{c_{\text{NP}}}{\Lambda^2}\right)^2 d\tilde{\sigma}^{\text{NP}}(m_{\ell\ell}, p_T^{\ell\ell})}\right. \\ \left. + \left(\frac{c_{\text{NP}}}{\Lambda^2}\right)^2 \frac{\int d\tilde{\sigma}^{\text{NP}}(\cos\theta, \phi, m_{\ell\ell}, p_T^{\ell\ell})\, d\cos\theta\, d\phi\, (\cos^2\theta + \sin^2\theta\cos 2\phi)}{d\sigma^{\text{SM}}(m_{\ell\ell}, p_T^{\ell\ell}) + \left(\frac{c_{\text{NP}}}{\Lambda^2}\right)^2 d\tilde{\sigma}^{\text{NP}}(m_{\ell\ell}, p_T^{\ell\ell})}\right]. \tag{3.29}$$

Due to the NP contribution appearing in the denominator, the dependence of $(A_0 - A_2)^{\text{SMEFT}}$ on the Wilson coefficient is non-linear, as anticipated above, and also affects the normalization of the distribution. This is expected to potentially break the Gaussian assumption that would allow a straightforward χ^2 analysis. For this reason, we decided to extract the bounds using the log-likelihood-ratio (LLR) test-statistic, which represents a more robust procedure for non-linear χ^2 (non-Gaussian likelihood). The LLR is defined as:

$$t_{\boldsymbol{c}} = -2\ln\frac{\mathscr{L}_{H_0}}{\mathscr{L}_{H_1}(\boldsymbol{c})}\,, \tag{3.30}$$

where, generically, H_0 represents the null hypothesis and H_1 the alternative hypothesis, that depends on some parameters $\boldsymbol{c}$. In our specific case, H_0 is the SM-only hypothesis, H_1 the SMEFT hypothesis, and the $\boldsymbol{c}$ parameters are the relevant Wilson coefficients. For the two hypotheses, we assumed:

- bin-by-bin, the likelihood for the SM hypothesis is considered to be a gaussian distribution, centered around the $A_0 - A_2$ central value provided by the simulations, with a width fixed by the estimated uncertainty. The total likelihood is then:

$$\mathscr{L}_{H_0} = \prod_{i=1}^{n_{bins}} N\left((A_0 - A_2)_i^{\text{SM}}\,;\, \sigma_i\right)\,. \tag{3.31}$$

- analogously, the likelihood for the SMEFT hypothesis is constructed as the product of gaussian distributions that, bin-by-bin, are centered around the $(A_0 - A_2)^{\text{SMEFT}}$ central value, calculated as explained above, with the same variance as in the SM-only hypothesis:[9]

$$\mathscr{L}_{H_1}(c_{\text{NP}}) = \prod_{i=1}^{n_{bins}} N\left((A_0 - A_2)_i^{\text{SMEFT}}(c_{\text{NP}})\,;\, \sigma_i\right)\,, \tag{3.32}$$

  where we have made explicit the $(A_0 - A_2)^{\text{SMEFT}}$ dependence on the Wilson coefficients.

Each Wilson coefficient is constrained individually by setting all others to zero.[10] The analysis

[9] Here we assume that the uncertainty is not significantly affected by the presence of NP contributions. This is a reasonable assumption, which can possibly be relaxed, if needed, by assuming $\sigma_i = \sigma_i(c_{\text{NP}})$ in Eq. (3.32).

[10] Obviously, since we have an implementation of the full likelihood in Eq. (3.32), we could also perform a simulta-

is performed using two kinematic distributions: the transverse momentum of the lepton pair $p_T^{\ell\ell}$ for above $p_T^{\ell\ell} > 10$ GeV and $80\,\text{GeV} < m_{\ell\ell} < 100\,\text{GeV}$, and the di-lepton invariant mass $m_{\ell\ell}$ for above $m_{\ell\ell} > 100$ GeV and $p_T^{\ell\ell} > 10$ GeV. The binning schemes adopted for both cases are provided in Tables 3.7-3.14 of Appendix 3.E, and are inspired by the existing measurements of Refs. [96, 98].

To constrain each Wilson coefficient, we estimated the distribution of the test-statistic $t_{c_{\rm NP}}$ under the SM hypothesis H_0, for a fixed, reasonable, initial value of $c_{\rm NP}$, by computing its value on 10^4 pseudo-experiments drawn from $\mathscr{L}_{H_0}$. More explicitly, we used the available Monte Carlo to draw, with replacement, pseudo-data and used them to compute $t_{c_{\rm NP}}$ 10^4 times. This yields a distribution of values of $t_{c_{\rm NP}}$. From this distribution we can identify a threshold, denoted by $\bar{t}_{c_{\rm NP}}$, defining the 95% confidence level (CL) threshold for rejecting the null hypothesis H_0 when testing against the alternative hypothesis H_1 corresponding to the fixed value $c_{\rm NP}$.

Next, we compute the average value of $t_{c_{\rm NP}}$ over 10^3 samples drawn from $\mathscr{L}_{H_1}$. This corresponds to generating pseudo-experiments corresponding to the H_1 hypothesis with fixed $c_{\rm NP}$, computing $t_{c_{\rm NP}}$ for each of these pseudo-experiments, and taking the average.[11] The corresponding average value is denoted by $t^*_{c_{\rm NP}}$.

Finally, both the threshold value $\bar{t}_{c_{\rm NP}}$ and the average value $t^*_{c_{\rm NP}}$ are iteratively computed adjusting the value of $c_{\rm NP}$ (using a bisection method) until they are equal, within a fixed threshold

$$\delta_{c_{\rm NP}} = 2\left|\frac{t^*_{c_{\rm NP}} - \bar{t}_{c_{\rm NP}}}{t^*_{c_{\rm NP}} + \bar{t}_{c_{\rm NP}}}\right|, \qquad \text{with } \delta_{c_{\rm NP}} < 5\%. \tag{3.33}$$

This is equivalent to solving the optimization problem:

$$\tilde{c}_{\rm NP} = \arg\min_{c_{\rm NP}} \left|\delta_{c_{\rm NP}} - 5\%\right| . \tag{3.34}$$

where $\tilde{c}_{\rm NP}$ represents the Wilson coefficient value corresponding to the 95% CL bound.

This procedure is carried out to obtain both the upper and the lower bounds. Since NP contributions manifest themselves as $c^2_{\rm NP}/\Lambda^4$ and uncertainties are symmetric, the resulting constraints are expected to be symmetric around zero. Nevertheless, we extract both sides independently.[12] Results are presented and discussed in the next Section.

3.6 Results and discussion

In Tables 3.3 and 3.4 we report the 95% CL individual bounds for all the Wilson coefficients in Eq. (3.1), equally valid for both electrons and muons, fixing the NP scale to $\Lambda = 4$ TeV. For the

neous fit of some, or all coefficients, or a combination with other analyses.

[11] Comparing the average test-statistic under the alternative hypothesis with the distribution of the test-statistic under the null hypothesis is one possible procedure to compute a bound. Another option would be to compute the full distribution under the alternative hypothesis and compare the two distributions in terms of a "confusion matrix", or to require a given power of the test at fixed CL. These are arbitrary choices of hypothesis testing.

[12] In some cases the bounds on certain Wilson coefficients are not exactly symmetric. These small deviations arise from small numerical instabilities in the procedure used to estimate the LLR distribution and give an idea of the uncertainty in our numerical optimization procedure.

95%CL, $\Lambda = 4$ TeV	$\mathcal{L} = 300\,\text{fb}^{-1}$		$\mathcal{L} = 3000\,\text{fb}^{-1}$	
$p_T^{\ell\ell}$ distribution	$A_0 - A_2$	cross-section	$A_0 - A_2$	cross-section
$c_{\ell equ}^{(3)}$	[−19.1, 19.1]	[−15.6, 15.8]	[−12.9, 13]	[−14.1, 14.1]
$c_{\ell equ}^{(1)}$	[−352, 344]	[−363, 359]	[−346, 344]	[−359, 359]
$c_{\ell edq}$	[−388, 375]	[−400, 400]	[−375, 375]	[−400, 400]
c_{eW}	[−9.61, 9.61]	[−18, 17.6]	[−9.38, 9.38]	[−17.2, 17.2]
c_{eB}	[−17.5, 17.5]	[−32.8, 32]	[−17.5, 17.5]	[−32, 32]
c_{uW}	[−13.1, 13.4]	[−3.2, 3.28]	[−11.4, 11.6]	[−2.85, 2.86]
c_{uB}	[−23.8, 23.8]	[−5.63, 5.78]	[−20.9, 20.6]	[−5.16, 5.16]
c_{dW}	[−14.4, 14.4]	[−4.45, 4.42]	[−13.1, 13.1]	[−3.98, 3.98]
c_{dB}	[−25.8, 25.8]	[−7.97, 7.97]	[−23.4, 23.8]	[−7.03, 7.03]

Table 3.3: 95% CL individual bounds for the Wilson coefficients in Eq. (3.1) for integrated luminosity values of 300 fb^{-1} and 3000 fb^{-1}. For each luminosity value, in the left column are reported the bounds obtained looking at the $A_0 - A_2$ observable and in the right column are reported the bounds obtained looking at the cross-section in the $p_T^{\ell\ell}$ distribution, fixing the NP scale Λ as $\Lambda = 4$ TeV.

95%CL, $\Lambda = 4$ TeV	$\mathcal{L} = 300\,\text{fb}^{-1}$		$\mathcal{L} = 3000\,\text{fb}^{-1}$	
$m_{\ell\ell}$ distribution	$A_0 - A_2$	cross-section	$A_0 - A_2$	cross-section
$c_{\ell equ}^{(3)}$	[−0.297, 0.289]	[−0.0641, 0.0645]	[−0.141, 0.141]	[−0.0645, 0.0625]
$c_{\ell equ}^{(1)}$	[−1.64, 1.62]	[−0.152, 0.15]	[−0.703, 0.703]	[−0.149, 0.148]
$c_{\ell edq}$	[−2.11, 2.19]	[−0.205, 0.202]	[−0.898, 0.938]	[−0.199, 0.199]
c_{eW}	[−5.63, 5.63]	[−2.34, 2.34]	[−3.31, 3.28]	[−2.34, 2.29]
c_{eB}	[−9.38, 9.06]	[−3.63, 3.59]	[−5.31, 5.42]	[−3.59, 3.63]
c_{uW}	[−7.24, 7.27]	[−2.87, 2.86]	[−4.22, 4.34]	[−2.93, 2.81]
c_{uB}	[−6.09, 6.09]	[−2.34, 2.37]	[−3.57, 3.52]	[−2.34, 2.34]
c_{dW}	[−8.67, 8.91]	[−3.81, 3.84]	[−5.33, 5.33]	[−3.87, 3.87]
c_{dB}	[−7.59, 7.31]	[−3.13, 3.16]	[−4.5, 4.5]	[−3.17, 3.13]

Table 3.4: 95% CL individual bounds for the Wilson coefficients in Eq. (3.1) for integrated luminosity values of 300 fb^{-1} and 3000 fb^{-1}. For each luminosity value, in the left column are reported the bounds obtained looking at the $A_0 - A_2$ observable and in the right column are reported the bounds obtained looking at the cross-section in the $m_{\ell\ell}$ distribution, fixing the New Physics scale Λ as $\Lambda = 4$ TeV.

$m_{\ell\ell}$ analysis, the last bin, corresponding to $3 - 10$ TeV, was not included because the estimated uncertainty was too large, and the region was partly outside of EFT validity range. For each luminosity value, we also report, for comparison, the bounds obtained from the differential cross-sections. The procedure to extract the latter bounds is the same as the one described above, except that the $A_0 - A_2$ observable is replaced by the corresponding differential cross-section. Results in Tables 3.3 and 3.4 are obtained considering two benchmark values for the integrated luminosity, corresponding to 300 fb^{-1} and 3000 fb^{-1}. We have fixed the NP scale to $\Lambda = 4$ TeV, and we focused on the $A_0 - A_2$ observable and on the $p_T^{\ell\ell}$ and $m_{\ell\ell}$ differential cross-sections.

For a more immediate visual comparison, the same bounds, except for the four-fermion operators in the $p_T^{\ell\ell}$ distribution, are shown graphically in Figures 3.5 and 3.6. We have not included the constraints on the Wilson coefficient of the four-fermion operators in Figure 3.5, since the $p_T^{\ell\ell}$

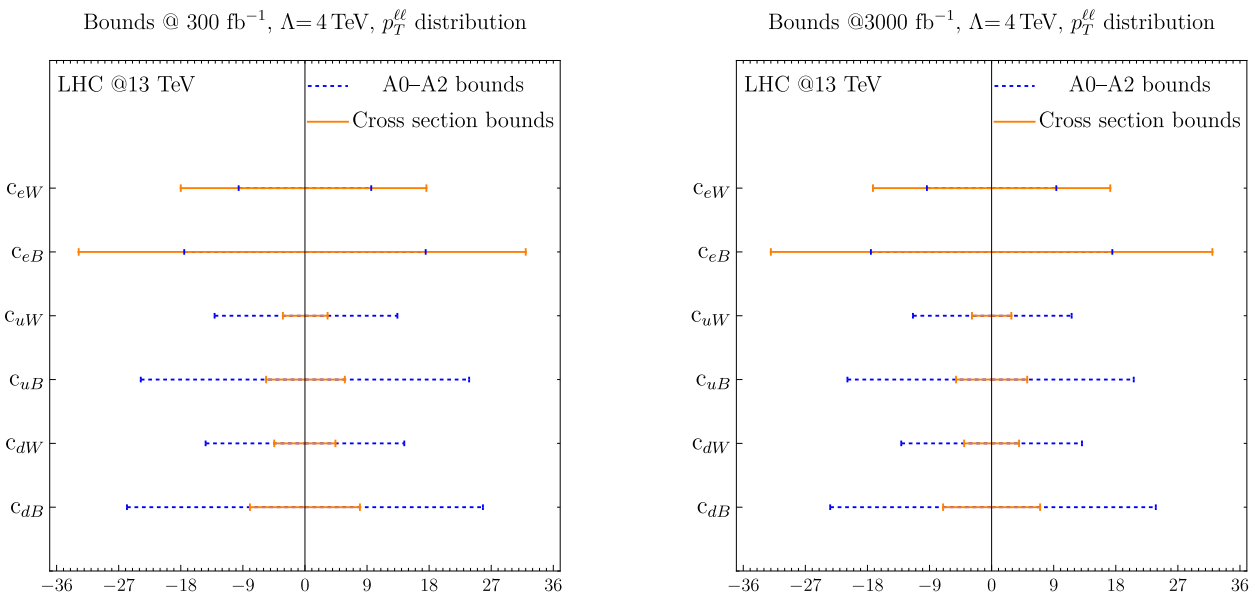


Figure 3.5: Comparison between the 95% CL boundaries obtained looking at the $A_0 - A_2$ observable (blue) and the cross-section (orange) in the $p_T^{\ell\ell}$ distribution. Left: 300 fb^{-1}. Right: 3000 fb^{-1}

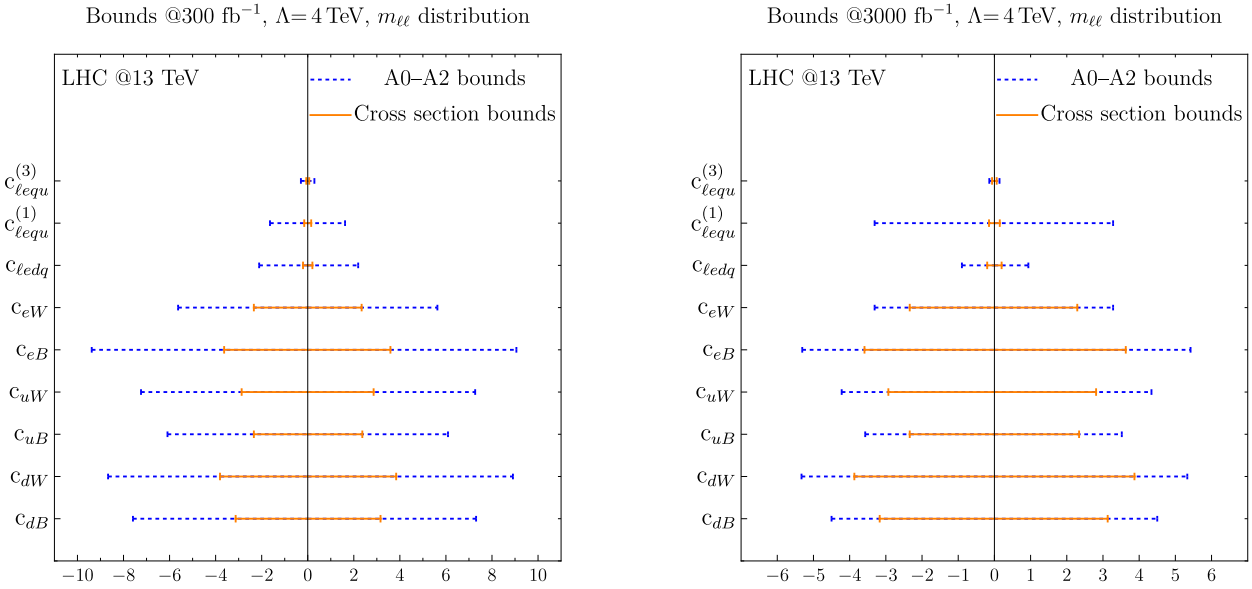


Figure 3.6: Comparison between the 95% CL boundaries obtained looking at the $A_0 - A_2$ observable (blue) and the cross-section (orange) in the $m_{\ell\ell}$ distribution. Left: 300 fb^{-1}. Right: 3000 fb^{-1}

distribution does not provide reasonably stringent bounds on them.[13]

Comparing the bounds shows that, in general, the constraints are tighter when looking at the cross-section rather than $A_0 - A_2$ observable (see Appendix 3.D for an intuitive explanation).

[13]This is due to the fact that the $p_T^{\ell\ell}$ analysis is done in a small window around the Z mass peak (80 GeV $< m_{\ell\ell} <$ 100 GeV), where the four-fermion operators do not profit of the Breit-Wigner enhancement of contributions involving the Z boson propagator.

However, our analysis neglects the leading operators contributing to the cross-section, namely the curret-current operators that interfere with the SM. In an analysis including all operators, those in Eq. (3.1) would be sub-leading and thus extremely hard to constrain. The $A_0 - A_2$ observable, on the other-hand, offers a complementary channel where the contribution of the current-current operators is highly suppressed, thanks to the interference with the SM and the Lam-Tung relation, and the dipole and scalar and tensor four-fermions operators provide the leading contributions. In the spirit of global fits and of breaking degeneracies between different (classes of) operators, the $A_0 - A_2$ observable becomes a very valuable tool, which gives independent and direct access to the chirality-breaking operators.[14]

From Tables 3.3 and 3.4, we can also see that the bounds projected for 300 fb^{-1} and for 3000 fb^{-1} are often comparable. This effect is more pronounced in the $p_T^{\ell\ell}$ analysis and for the differential cross-section observable. This behavior comes from the fact that the bound on those coefficients is dominated by the systematic uncertainty, which we assume not to improve with luminosity. As we discussed before, systematic uncertainty dominate the low-p_T and low-$m_{\ell\ell}$ regions, so that only operators with a shape significantly different from the SM in the high-p_T or high-$m_{\ell\ell}$ regions will see a reasonable improvement in their bounds when increasing the integrated luminosity. The different shapes shown in Figure 3.4 confirm this interpretation and show that the different growth in energy of the dipole and four-fermion operators leads not only to different sensitivities in the two kinematic distributions, but also to a different pattern of improvement of the bounds with the integrated luminosity.

As mentioned above, the $m_{\ell\ell}$ distribution, yields tighter bounds than the $p_T^{\ell\ell}$ distribution, indicating that higher energies enhance the sensitivity to both classes of operators. This can be seen very neatly in Figure 3.6, which makes apparent how the four-fermion operators are more tightly constrained than the dipole ones.

Finally, in Tables 3.5 and 3.6, we report the bounds obtained using the Taylor expansion of the $A_0 - A_2$ observable truncated at order $c_{\rm NP}^2/\Lambda^4$. Such bounds do not differ significantly from those obtained using the full expression in Table 3.3 and 3.4. This is the expected behavior in the region of validity of the EFT expansion, where higher-order terms are sub-leading corrections. In particular, the expansion parameter involves the ratio $d\sigma^{\rm SMEFT}/d\sigma^{\rm SM} \times (c_{\rm NP}/\Lambda^2)^2$, which remains sufficiently small for small enough Wilson coefficients and in most of the allowed phase space. The largest discrepancies between the two approaches appear in the regions of the phase space where the NP contributions are more significant, such as the tails of $p_T^{\ell\ell}$ and $m_{\ell\ell}$ distributions. In those regions, for relatively large values of the Wilson coefficients, the expansion parameter may become largish, making the expansion less reliable, and one could expect higher-order terms to become relevant. This is the reason why we believe that showing both sets of bounds is useful, as it gives an idea of the robustness of the results.

[14] As it was pointed out in Refs. [75, 108], the interference between dimension-eight operators and the SM is expected to violate the Lam-Tung relation at $\mathcal{O}(\alpha_{\rm S})$. These contributes are expected to be subdominant with respect to those of the dipole operators when probed close to the Z mass peak, and comparable with respect to the scalar/tensor operators in the high $m_{\ell\ell}$ region.

95% CL, $\Lambda = 4$ TeV $p_T^{\ell\ell}$ distribution	$\mathcal{L} = 300\,\text{fb}^{-1}$	$\mathcal{L} = 3000\,\text{fb}^{-1}$
$c_{\ell equ}^{(3)}$	[−18.8, 18.8]	[−12.9, 12.8]
$c_{\ell equ}^{(1)}$	[−359, 348]	[−344, 344]
$c_{\ell edq}$	[−388, 375]	[−375, 375]
c_{eW}	[−9.38, 9.38]	[−9.38, 9.3]
c_{eB}	[−17.5, 17.5]	[−17.2, 16.9]
c_{uW}	[−12.6, 12.5]	[−10.4, 10.3]
c_{uB}	[−23.1, 23.1]	[−18.9, 19.2]
c_{dW}	[−14.1, 14.1]	[−12.5, 12.5]
c_{dB}	[−25, 25.4]	[−22.7, 22.7]

Table 3.5: 95% CL individual bounds for the Wilson coefficients in eq. (3.1) for integrated luminosity values of 300 fb^{-1} and 3000 fb^{-1}. These are obtained looking at the $A_0 - A_2$ observable in the p_T distribution expanded at order $c_{\text{NP}}^2/\Lambda^4$, fixing the New Physics scale Λ as $\Lambda = 4$ TeV.

95% CL, $\Lambda = 4$ TeV $m_{\ell\ell}$ distribution	$\mathcal{L} = 300\,\text{fb}^{-1}$	$\mathcal{L} = 3000\,\text{fb}^{-1}$
$c_{\ell equ}^{(3)}$	[−0.227, 0.234]	[−0.133, 0.129]
$c_{\ell equ}^{(1)}$	[−1, 0.988]	[−0.563, 0.568]
$c_{\ell edq}$	[−1.31, 1.31]	[−0.738, 0.75]
c_{eW}	[−5.47, 5.31]	[−3.32, 3.28]
c_{eB}	[−8.75, 8.75]	[−5.31, 5.31]
c_{uW}	[−7.03, 7.03]	[−4.16, 4.24]
c_{uB}	[−6.04, 5.86]	[−3.52, 3.52]
c_{dW}	[−8.44, 8.44]	[−5.27, 5.27]
c_{dB}	[−7.03, 7.23]	[−4.45, 4.45]

Table 3.6: 95% CL individual bounds for the Wilson coefficients in Eq. (3.1) for integrated luminosity values of 300 fb^{-1} and 3000 fb^{-1}. These are obtained looking at the $A_0 - A_2$ observable in the $m_{\ell\ell}$ distribution expanded at order $c_{\text{NP}}^2/\Lambda^4$, fixing the New Physics scale Λ as $\Lambda = 4$ TeV.

3.7 Conclusions

In this work, we analyzed the contributions of chirality-breaking dimension-six operators, specifically dipole and four-fermion operators, to the Drell–Yan (DY) process.

Our focus was on the angular observable $A_0 - A_2$, which vanishes in the Standard Model (SM) up to $\mathcal{O}(\alpha_S^2)$ and does not receive contributions from dimension-six operators that interfere with the SM amplitude. In Ref. [75], bounds on dipole operators were derived at a center-of-mass energy of 8 TeV using the ATLAS analysis of Ref. [103]. Here, we extended that study by providing projected bounds at 13 TeV, considering both dipole and four-fermion operators, and using both the $p_T^{\ell\ell}$ and $m_{\ell\ell}$ distributions.

We performed SM Monte Carlo simulations at $\mathcal{O}(\alpha_S^2)$ with the `MiNNLOPS` tool, and included the NP contributions analytically to estimate their effect on the $p_T^{\ell\ell}$ and $m_{\ell\ell}$ distributions of the DY

cross-section, the angular coefficients A_ℓ, and the $A_0 - A_2$ observable. A pseudo-analysis was carried out for 300 fb^{-1} of integrated luminosity, corresponding the final LHC dataset, and for 3000 fb^{-1} of integrated luminosity, corresponding to the HL-LHC. We included estimates of both theoretical and experimental systematic uncertainties, based on current measurements and projections for future improvements.

We compared the 95% CL bounds on Wilson coefficients extracted from $A_0 - A_2$ with those obtained from the differential cross-section. The latter are found to be generally more stringent, which is not surprising given the much larger statistics and smaller experimental uncertainties available for cross-section measurements. However, the key point of this work is that the angular observable $A_0 - A_2$ offers a clean and independent probe of chirality-breaking operators, free from contamination by dimension-six operators that interfere with the SM. This is not the case for the cross-section, where SM contributions are never suppressed and the leading SMEFT effects arise from interference with current–current four-fermion operators. As such, $A_0 - A_2$ becomes particularly valuable in global SMEFT fits, where multiple operators are constrained simultaneously and degeneracies in parameter space must be resolved.

We stress that our study was conducted in a simplified theoretical setup and does not attempt to replicate a full experimental analysis. In particular, we assumed direct access to the angular coefficients and to the $A_0 - A_2$ observable, whereas real analyses typically extract these quantities via template fits. While we strived to provide realistic estimates of theoretical and experimental systematics, a dedicated experimental study will ultimately be required to fully assess the potential of $A_0 - A_2$ for probing chirality-breaking SMEFT operators at the LHC.

Appendix

3.A The Collins-Soper frame

Let the 4-vectors of the two incident beams in the laboratory frame be:

$$b_+ = (E_{b_+}, \vec{b}_+), \tag{3.35}$$

$$b_- = (E_{b_-}, \vec{b}_-). \tag{3.36}$$

We call b'_+ and b'_- the respective 4-vectors boosted in the di-lepton rest frame. The z axis of the CS frame is defined as the bisector of the unit vectors $\hat{b}'_+$ and $-\hat{b}'_-$, pointing such that its scalar product with the di-lepton 3-momentum, in the laboratory frame, is positive.
Another axis, called q, is defined as the one laying in the plane defined by $\hat{b}'_+$ and $\hat{b}'_-$, orthogonal to the z axis and pointing in the direction opposite to $\hat{b}'_+ + \hat{b}'_-$.
The angle θ is defined with respect to the z axis, while the angle ϕ is defined with respect to the q axis.
This orientation of the CS frame axes can provide, for certain events, angles that are shifted by π with respect to those defined in the original paper [74]. This can be taken into account by slightly modifying the original definitions, introducing a factor $p_z^{\ell\ell}/\mid p_z^{\ell\ell} \mid$ that accounts for axis orientation. Therefore, one has:

$$\cos\theta = \frac{p_{z,\ell\ell}}{\mid p_{z,\ell\ell} \mid} \frac{2(p_{z,\ell^+} E_{\ell^-} - p_{z,\ell^-} E_{\ell^+})}{m_{\ell\ell}\sqrt{m_{\ell\ell}^2 + p_T^{\ell\ell}}} \tag{3.37}$$

$$\tan\phi = \frac{p_{z,\ell\ell}}{\mid p_{z,\ell\ell} \mid} \frac{\sqrt{m_{\ell\ell}^2 + p_{T,\ell\ell}^2}}{m_{\ell\ell}} \frac{\Delta p_T^{\ell\ell} \cdot \hat{R}_T}{\Delta p_T^{\ell\ell} \cdot \hat{p}_{T,\ell\ell}} \tag{3.38}$$

where $\hat{p}_{T,\ell\ell}$ is a unit vector in the direction of $p_T^{\ell\ell}$ and $\hat{R}_T$ is the normalized cross product between $\hat{b}_+$ and $p_{z,\ell\ell}$.

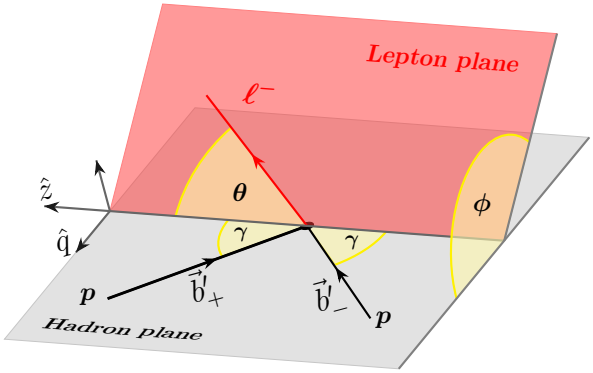


Figure 3.7: A graphic representation of the Collins-Soper frame.

3.B Angular observables from spherical-harmonics

In the following, we report the combination of spherical harmonics that provides the factor multiplying each A_l in Eq. (3.2).

$$
\begin{aligned}
1+\cos^2\theta &= \frac{4\sqrt{\pi}}{15}\left(10Y_0^0+\sqrt{5}Y_2^0\right), \\
\frac{1}{2}(1-3\cos^2\theta) &= -2\sqrt{\frac{\pi}{5}}Y_2^0, \\
\sin 2\theta\cos\phi &= 2\sqrt{\frac{2\pi}{15}}(Y_2^{-1}-Y_2^1), \\
\frac{1}{2}(\sin^2\theta\cos 2\phi) &= \sqrt{\frac{2\pi}{15}}(Y_2^{-2}+Y_2^2), \\
\sin\theta\cos\phi &= \sqrt{\frac{2\pi}{3}}(Y_1^{-1}-Y_1^1), \\
\cos\theta &= 2\sqrt{\frac{\pi}{3}}Y_1^0, \\
\sin^2\theta\sin 2\phi &= -2i\sqrt{\frac{2\pi}{15}}(Y_2^{-2}-Y_2^2), \\
\sin 2\theta\sin\phi &= -2i\sqrt{\frac{2\pi}{15}}(Y_2^{-1}+Y_2^1), \\
\sin\theta\sin\phi &= -i\sqrt{\frac{2\pi}{3}}(Y_1^{-1}+Y_1^1).
\end{aligned}
$$

One can verify that, apart from the first two, each of these combinations of spherical harmonics is orthogonal to the others.

3.C Differential cross-section and PDF integration

In this appendix we derive the expression for the fully differential unpolarized cross-section for the process $pp \to \ell^+\ell^- X$ at $\mathcal{O}(\alpha_S)$. This is achieved by splitting it in two subsequent processes: the first is $pp \to ZX$, which on the partonic level gives $q\bar{q} \to Zg$ or $qg \to Zq$; the second consists in the Z boson decaying into the lepton pair $Z \to \ell^+\ell^-$. Neglecting the quark masses the kinematic constraint $s + t + u = m_{\ell\ell}^2$ must hold, where $m_{\ell\ell}$ denotes the invariant mass of the lepton pair and s, t and u the partonic Mandelstam variables defined for the first subprocess.
In the CS frame the Z boson momentum reads $p_Z = \left(E_Z, p_T^{\ell\ell}, 0, p_l\right)$. All the kinematic quantities are expressed in terms of:

- the invariant mass of the lepton pair $m_{\ell\ell}$;
- the transverse momentum of the lepton pair $p_T^{\ell\ell}$;
- the energy of each hadron in the laboratory frame E_p;
- the rapidity $y = \log\left(\frac{E_Z + p_l}{E_Z - p_l}\right)$ of the Z boson;
- the fraction of hadronic momentum carried by the incoming partons x_a.

In terms of these quantities, the energy E_Z and longitudinal momentum p_l become

$$E_Z = \sqrt{m_{\ell\ell}^2 + (p_T^{\ell\ell})^2}\cosh y, \quad p_l = \sqrt{m_{\ell\ell}^2 + (p_T^{\ell\ell})^2}\sinh y, \tag{3.39}$$

and the Mandelstam variables can be rewritten as

$$s = 4x_a x_b E_p^2, \quad t = m_{\ell\ell}^2 - 2x_a E_p\sqrt{m_{\ell\ell}^2 + (p_T^{\ell\ell})^2}e^{-y}, \quad u = m_{\ell\ell}^2 - 2x_b E_p\sqrt{m_{\ell\ell}^2 + (p_T^{\ell\ell})^2}e^{y}. \tag{3.40}$$

The general expression for the hadronic cross-section $\sigma_{pp\to X\ell^+\ell^-}$ is

$$\begin{aligned}\sigma_{pp\to X\ell^+\ell^-} = \sum_{ab}\int \frac{d^3p_X}{(2\pi)^3 2E_X}\frac{d^3p_{\ell^+}}{(2\pi)^3 2E_{\ell^+}}\frac{d^3p_{\ell^-}}{(2\pi)^3 2E_{\ell^-}}\frac{1}{2s}f_{a/p}(x_a)f_{b/p}(x_b)dx_a dx_b \\ \times \langle|\mathcal{M}_{ab}|^2\rangle(2\pi)^4\delta^{(4)}(p_a + p_b - p_X - p_{\ell^+} - p_{\ell^-}),\end{aligned} \tag{3.41}$$

with $\mathcal{M}_{ab}$ the amplitudes for the partonic processes; the meaning of the other symbols is understood. Momentum conservation gives

$$\begin{aligned}d\sigma_{pp\to X\ell^+\ell^-} = \sum_{ab}\int d^4p_Z d\phi_{\ell\ell}^{(2)}\frac{d^3p_X}{(2\pi)^3 2E_X}\frac{1}{2s}f_{a/p}(x_a)f_{b/p}(x_b)dx_a dx_b \\ \times \langle|\mathcal{M}_{ab}|^2\rangle(2\pi)^4\delta^{(4)}(p_a + p_b - p_X - p_Z),\end{aligned} \tag{3.42}$$

where

$$d\phi_{\ell\ell}^{(2)} \equiv (2\pi)^4\,\delta^{(4)}\left(p_Z - p_{\ell^+} - p_{\ell^-}\right)\frac{d^3p_{\ell^+}}{(2\pi)^3 2E_{\ell^+}}\frac{d^3p_{\ell^-}}{(2\pi)^3 2E_{\ell^-}} \tag{3.43}$$

is the lepton pair phase space.
The integral on x_b is now performed using the identity $\frac{d^3p_X}{2E_X} = d^4p_X\delta(p_X^2)$: we can substitute in the Dirac δ-function $p_X^2 = s + t + u - m_{\ell\ell}^2$ and then use Eq. (3.40) to obtain

$$\delta(p_X^2) = \frac{1}{2E_p^2\,|2x_a - x_T e^y|}\delta\left(x_b - \frac{x_a x_T e^{-y} - \frac{m_{\ell\ell}^2}{2E_p^2}}{2x_a - x_T e^y}\right), \tag{3.44}$$

with $x_T \equiv \sqrt{m_{\ell\ell}^2 + (p_T^{\ell\ell})^2}/E_p$. Moreover, being $x_b < 1$, the above expression gives a lower limit on x_a that reads

$$x_a > x_a^{\min} \equiv \frac{x_T e^y - \frac{m_{\ell\ell}^2}{2E_p^2}}{2 - x_T e^{-y}}. \tag{3.45}$$

The integration over the the lepton pair momenta gives

$$\sigma_{pp\to\ell^+\ell^-} = \frac{1}{(2\pi)^6}\frac{1}{64E_p^4}\int\left[\sum_{ab}\int_{x_a^{\min}}^1 f_{a/p}(x_a)f_{b/p}(x_b)dx_a\frac{\langle|\mathcal{M}_{ab}|^2\rangle}{(2x_a - x_T e^y)\,x_a x_b^*}\right]dc_\theta d\phi d^4p_Z, \tag{3.46}$$

where the star in x_b^* keeps track of the constraint on x_b given by Eq. (3.44).
The 4-momentum of the outgoing Z boson can be parametrized as

$$p_Z = \left(\sqrt{m_{\ell\ell}^2 + (p_T^{\ell\ell})^2}\cosh y, p_T^{\ell\ell}\cos\alpha, p_T^{\ell\ell}\sin\alpha, \sqrt{m_{\ell\ell}^2 + (p_T^{\ell\ell})^2}\sinh y\right), \tag{3.47}$$

which gives

$$d^4p_Z = \frac{1}{2}p_T^{\ell\ell}d\alpha dm_{\ell\ell}^2 dp_T^{\ell\ell}dy. \tag{3.48}$$

Plugging this into Eq. (3.46) and integrating over α we find

$$\frac{d\sigma_{pp\to X\ell^+\ell^-}}{dm_{\ell\ell}^2 dp_T^{\ell\ell}dc_\theta d\phi} = \frac{1}{(2\pi)^5}\frac{\pi p_T^{\ell\ell}}{128E_p^4}\int_{y_{\min}}^{y_{\max}}\left[\sum_{ab}\int_{x_a^{\min}}^1 f_{a/p}(x_a)f_{b/p}(x_b^*)dx_a\frac{\langle|\mathcal{M}_{ab}|^2\rangle}{(2x_a - x_T e^y)\,x_a x_b^*}\right]dy, \tag{3.49}$$

where the upper and lower bounds

$$\begin{aligned} y_{\max} &\equiv \log\left[\frac{1}{2}\left(\frac{4E_p^2 + m_{\ell\ell}^2}{2E_p^2 x_T} + \sqrt{\left(\frac{4E_p^2 + m_{\ell\ell}^2}{2E_p^2 x_T}\right)^2 - 4}\right)\right], \\ y_{\min} &\equiv \log\left(\frac{x_T}{2}\right) \end{aligned} \tag{3.50}$$

follow from the fact that $x_a^{\min}$ is positive and lesser than 1.
The analytic results obtained using formula (3.49) were found to be in excellent agreement with those obtained with `Madgraph5` [109].

3.D Cross-section vs $A_0 - A_2$ bounds

In this appendix, we provide an intuitive explanation for why the bounds obtained from the cross-section are tighter than those obtained with the $A_0 - A_2$ observable. In particular we focus on the ratios:

$$r_1 = \left| \frac{\sigma^{\text{SMEFT}} - \sigma^{\text{SM}}}{\delta_{\sigma^{\text{SM}}}} \right| , \tag{3.51}$$

$$r_2 = \left| \frac{(A_0 - A_2)^{\text{SMEFT}} - (A_0 - A_2)^{\text{SM}}}{\delta_{(A_0 - A_2)^{\text{SM}}}} \right| , \tag{3.52}$$

$$R = \frac{r_1}{r_2} , \tag{3.53}$$

where $\delta_{\sigma^{\text{SM}}}$ and $\delta_{(A_0-A_2)^{\text{SM}}}$ indicate the uncertainty on the cross-section and the $A_0 - A_2$ observable, respectively. Clearly, at fixed c_{NP}, a larger value of the ratio indicates that the corresponding observable is more sensitive to the considered operator. Figure 3.8 shows the $r1$ (upper left), r_2 (upper right) and R (lower panel) ratios for the $p_T^{\ell\ell}$ distribution with fixed luminosity of 300 fb^{-1}. Each Wilson coefficient c_{NP} is set to 10 and Λ is fixed at 4 TeV.

It is immediate to notice that r_1 is significantly larger that r_2, as R ranges from a few units up to more than 1000. This behavior indicates that the cross-section is more sensitive to NP insertion, due to its smaller uncertainties. Furthermore, the R plot shows a peak in the central bins, where $A_0 - A_2$ increases and the statistical uncertainty is small thanks to the high statistics. This pattern does not occur for the c_{eW} and c_{eB} Wilson coefficients, which are, in fact, better constrained through the $A_0 - A_2$ observable. An analogous behavior is observed for the 3000 fb^{-1} integrated luminosity. Similar conclusions hold when considering the $m_{\ell\ell}$ distribution, where r_1 remains consistently larger than r_2.

3.E Numerical tables

In this appendix we report the numerical values of the observables appearing in Figures 3.1, 3.2, 3.3 with explicit separation of statistical and systematic uncertainties.

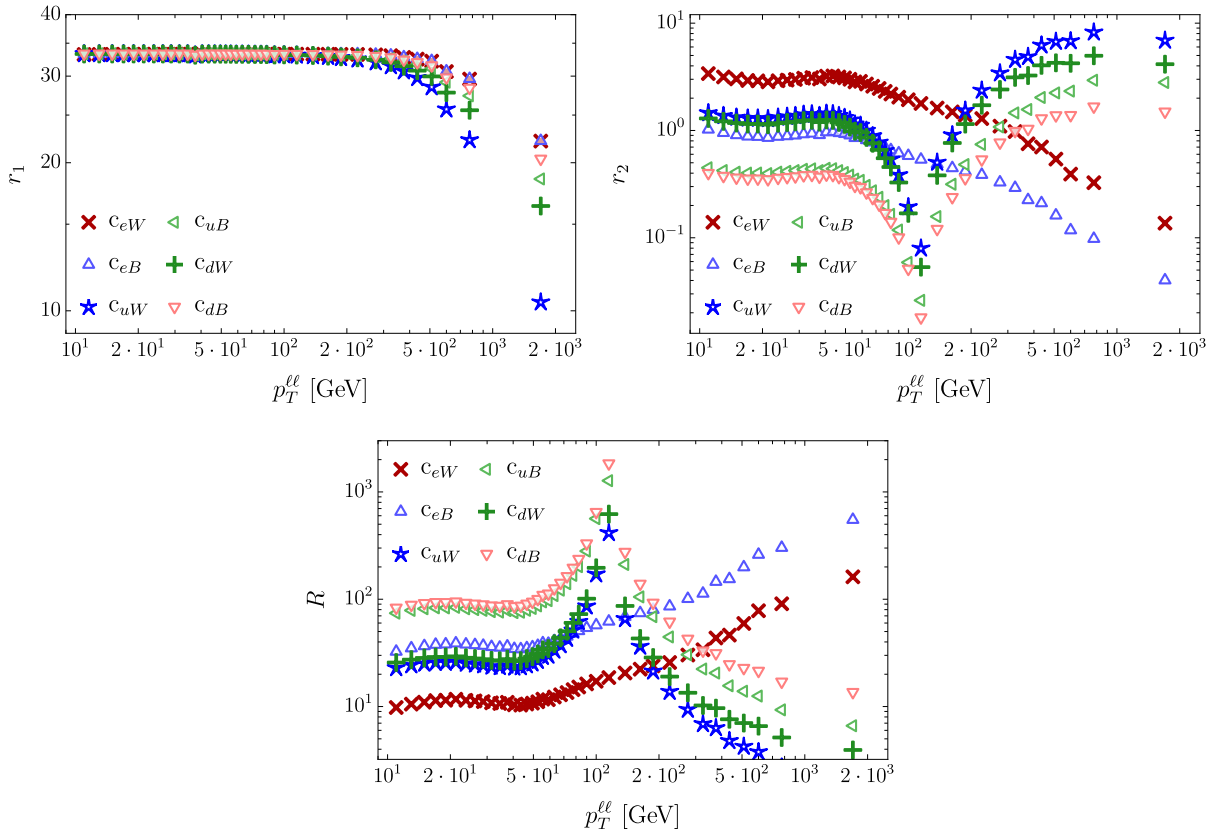


Figure 3.8: Plots of the ratios defined in equation (3.51) assuming $\mathcal{L} = 300\ \text{fb}^{-1}$ in the $p_T^{\ell\ell}$ distribution. Upper left: r_1, upper right: r_2, lower panel: R. The different behavior of c_{eW} and c_{eB}, that are better constrained by the $A_0 - A_2$ observable, is evident.

$p_T^{\ell\ell}$ [GeV]	N_{MC}	σ [pb]	$\delta\sigma_{\mathrm{stat}}^{300}$ [pb]	$\delta\sigma_{\mathrm{stat}}^{3000}$ [pb]	$\delta\sigma_{\mathrm{syst}}^{3\%}$ [pb]
10. – 12.	55355060	118.99	0.0462851	0.0146366	3.5697
12. – 14.	44432363	100.94	0.0350784	0.0110928	3.02821
14. – 16.	36325175	86.4391	0.0271705	0.00859205	2.59317
16. – 18.	30149734	74.566	0.0268787	0.00849979	2.23698
18. – 20.	25356520	64.7717	0.0211872	0.00669997	1.94315
20. – 22.5	26414219	69.6736	0.0345891	0.010938	2.09021
22.5 – 25.	21848818	59.3793	0.0205171	0.00648808	1.78138
25. – 27.5	18289822	51.0283	0.0207527	0.00656258	1.53085
27.5 – 30.	15450562	44.1082	0.0153048	0.00483982	1.32325
30. – 33.	15561084	45.4264	0.0155019	0.00490215	1.36279
33. – 36.	12982145	38.7237	0.0154616	0.00488938	1.16171
36. – 39.	10916808	33.2943	0.104793	0.0331385	0.998829
39. – 42.	9253016	28.593	0.0125402	0.00396556	0.857789
42. – 45.	7892013	24.7582	0.0112726	0.00356472	0.742747
45. – 48.	6768494	21.5125	0.0102497	0.00324124	0.645376
48. – 51.	5840548	18.7992	0.0102651	0.00324612	0.563975
51. – 54.	5053525	16.4331	0.009844	0.00311295	0.492994
54. – 57.	4394576	14.4348	0.0106579	0.00337034	0.433044
57. – 61.	5002474	16.5707	0.018601	0.00588215	0.497122
61. – 65.	4199710	14.0805	0.00805207	0.00254629	0.422415
65. – 70.	4330950	14.6711	0.0076631	0.00242328	0.440134
70. – 75.	3520638	12.0448	0.00867464	0.00274316	0.361345
75. – 80.	2883906	9.93928	0.00625729	0.00197873	0.298178
80. – 85.	2374827	8.24452	0.00648999	0.00205232	0.247335
85. – 95.	3608604	12.6469	0.021292	0.00673313	0.379406
95. – 105.	2524395	8.91138	0.00627343	0.00198383	0.267341
105. – 125.	3122264	11.1183	0.00677558	0.00214263	0.33355
125. – 150.	1999251	7.18747	0.00551796	0.00174493	0.215624
150. – 175.	1017474	3.68138	0.00427167	0.00135082	0.110441
175. – 200.	555846	2.01631	0.0030348	0.000959689	0.0604894
200. – 250.	518627	1.877	0.00313351	0.000990902	0.05631
250. – 300.	203615	0.735178	0.0018908	0.000597924	0.0220553
300. – 350.	89323	0.32241	0.00118899	0.000375991	0.00967229
350. – 400.	43580	0.156648	0.000841524	0.000266113	0.00469943
400. – 470.	28142	0.101251	0.000669394	0.000211681	0.00303754
470. – 550.	13814	0.0492041	0.000388475	0.000122847	0.00147612
550. – 650.	6735	0.0239324	0.000303223	0.0000958875	0.000717971
650. – 900.	4229	0.0148124	0.000228624	0.0000722974	0.000444371
900. – 2500.	791	0.00265501	0.0000894172	0.0000282762	0.0000796503

Table 3.7: Expected results of measurements of the transverse momentum spectrum at the LHC at 13 TeV with 300 fb^{-1} and 3 ab^{-1}. The columns show the bin range, the number of Monte Carlo events, the expected cross-section, the statistical uncertainty with 300 fb^{-1} and 3 ab^{-1}, and a systematic uncertainty of 3%.

$p_T^{\ell\ell}$ [GeV]	A_0	$(\delta A_0)^{300}_{\text{stat}}$	$(\delta A_0)^{3000}_{\text{stat}}$	$(\delta A_0)^{3\%}_{\text{syst}}$
10. – 12.	0.0454443	0.00114832	0.000363131	0.0186367
12. – 14.	0.0599507	0.00135674	0.000429039	0.0182015
14. – 16.	0.0762949	0.0014533	0.000459572	0.0177112
16. – 18.	0.0929306	0.00134978	0.000426839	0.0172121
18. – 20.	0.112002	0.0013841	0.000437691	0.01664
20. – 22.5	0.133622	0.001311	0.000414575	0.0159913
22.5 – 25.	0.158508	0.0011865	0.000375205	0.0152448
25. – 27.5	0.186841	0.00158303	0.000500597	0.0143948
27.5 – 30.	0.213909	0.00122356	0.000386924	0.0135827
30. – 33.	0.24182	0.00124103	0.000392448	0.0127454
33. – 36.	0.274575	0.0013978	0.000442023	0.0117627
36. – 39.	0.30449	0.00357953	0.00113195	0.0108653
39. – 42.	0.338655	0.00157738	0.00049881	0.00984036
42. – 45.	0.369891	0.00131081	0.000414515	0.00890328
45. – 48.	0.400941	0.00182146	0.000575997	0.00797177
48. – 51.	0.429686	0.00183945	0.000581685	0.00710942
51. – 54.	0.455141	0.00167453	0.000529532	0.00634577
54. – 57.	0.479131	0.00175029	0.00055349	0.00562608
57. – 61.	0.510621	0.00232018	0.000733707	0.00468139
61. – 65.	0.542085	0.00204103	0.000645431	0.00373744
65. – 70.	0.571995	0.00163757	0.000517846	0.00284015
70. – 75.	0.601204	0.00168406	0.000532546	0.00196388
75. – 80.	0.633106	0.00186528	0.000589854	0.00100681
80. – 85.	0.66316	0.00247316	0.000782083	0.0001052
85. – 95.	0.686713	0.0060731	0.00192048	0.000601351
95. – 105.	0.732296	0.00206278	0.000652307	0.00196888
105. – 125.	0.779334	0.00163382	0.000516661	0.00338001
125. – 150.	0.832182	0.0024437	0.000772765	0.00496547
150. – 175.	0.867126	0.00282857	0.000894471	0.00601376
175. – 200.	0.899789	0.0041151	0.00130131	0.00699372
200. – 250.	0.916585	0.00481605	0.00152297	0.00749755
250. – 300.	0.949971	0.00721618	0.00228196	0.00849908
300. – 350.	0.947089	0.0109943	0.00347669	0.00841257
350. – 400.	0.977413	0.0146018	0.0046175	0.0093229
400. – 470.	0.965547	0.0181728	0.00574673	0.00896602
470. – 550.	0.989361	0.0254853	0.00805915	0.00968054
550. – 650.	0.9903	0.0337669	0.010678	0.00970749
650. – 900.	0.978093	0.0496695	0.0157069	0.00934364
900. – 2500.	0.990086	0.111957	0.0354038	0.00968581

Table 3.8: Expected results of measurements of the A_0 observable at the LHC at 13 TeV with 300 fb^{-1} and 3 ab^{-1}. The columns show the bin range, the expected value of A_0, the statistical uncertainty with 300 fb^{-1} and 3 ab^{-1}, and a systematic uncertainty of 3%.

$p_T^{\ell\ell}$ [GeV]	A_2	$(\delta A_2)^{300}_{\text{stat}}$	$(\delta A_2)^{3000}_{\text{stat}}$	$(\delta A_2)^{3\%}_{\text{syst}}$
10. – 12.	0.0236693	0.00121024	0.000382711	0.000710079
12. – 14.	0.0252714	0.00147315	0.000465851	0.000758143
14. – 16.	0.0305507	0.00105898	0.00033488	0.00091652
16. – 18.	0.0313687	0.00101226	0.000320105	0.00094106
18. – 20.	0.0379034	0.000984646	0.000311372	0.0011371
20. – 22.5	0.0410952	0.00108569	0.000343324	0.00123286
22.5 – 25.	0.045506	0.000973089	0.000307718	0.00136518
25. – 27.5	0.0499428	0.000975623	0.000308519	0.00149828
27.5 – 30.	0.0537455	0.00112234	0.000354915	0.00161236
30. – 33.	0.0579562	0.00105402	0.000333309	0.00173869
33. – 36.	0.0626439	0.00116594	0.000368703	0.00187932
36. – 39.	0.0598799	0.00623964	0.00197315	0.00179644
39. – 42.	0.0652713	0.00123154	0.000389449	0.00195814
42. – 45.	0.0695346	0.00134794	0.000426257	0.00208604
45. – 48.	0.070112	0.00140319	0.000443728	0.00210336
48. – 51.	0.0707345	0.00135741	0.000429251	0.00212204
51. – 54.	0.0727296	0.00144449	0.000456788	0.00218189
54. – 57.	0.0774697	0.00252136	0.000797325	0.00232409
57. – 61.	0.0773344	0.00353121	0.00111667	0.00232003
61. – 65.	0.076637	0.00150062	0.000474538	0.00229911
65. – 70.	0.0759462	0.00163643	0.000517484	0.00227839
70. – 75.	0.0734238	0.00154593	0.000488867	0.00220272
75. – 80.	0.0749548	0.0019587	0.000619396	0.00224864
80. – 85.	0.0728264	0.00160893	0.000508789	0.0021848
85. – 95.	0.0776449	0.00432023	0.00136618	0.00232935
95. – 105.	0.0722412	0.00178587	0.000564742	0.00216723
105. – 125.	0.0682163	0.00171466	0.000542224	0.00204649
125. – 150.	0.0606084	0.00219335	0.000693597	0.00181825
150. – 175.	0.0549135	0.00265379	0.000839202	0.00164741
175. – 200.	0.0500717	0.00341425	0.00107968	0.00150216
200. – 250.	0.0438294	0.00375023	0.00118593	0.0013149
250. – 300.	0.0255768	0.00730141	0.00230891	0.000767309
300. – 350.	0.0257082	0.00878564	0.00277826	0.000771414
350. – 400.	0.046506	0.0139431	0.00440919	0.00139526
400. – 470.	0.0257966	0.0160086	0.00506237	0.000774092
470. – 550.	0.0126944	0.0248429	0.007856	0.000380961
550. – 650.	0.00365842	0.0297732	0.00941513	0.000110922
650. – 900.	0.0541991	0.0472457	0.0149404	0.00162197
900. – 2500.	0.0367986	0.0994578	0.0314513	0.00110695

Table 3.9: Expected results of measurements of the A_2 observable at the LHC at 13 TeV with 300 fb^{-1} and 3 ab^{-1}. The columns show the bin range, the expected value of A_2, the statistical uncertainty with 300 fb^{-1} and 3 ab^{-1}, and a systematic uncertainty of 3%.

$p_T^{\ell\ell}$ [GeV]	$A_0 - A_2$	$\delta(A_0 - A_2)^{300}_{\text{stat}}$	$\delta(A_0 - A_2)^{3000}_{\text{stat}}$	$\delta(A_0 - A_2)^{3\%}_{\text{syst}}$
10. – 12.	0.0283314	0.00219976	0.000695624	0.0186437
12. – 14.	0.0330746	0.002538	0.000802587	0.0182193
14. – 16.	0.0419759	0.00274345	0.000867555	0.0177411
16. – 18.	0.0415125	0.00280323	0.000886458	0.0172811
18. – 20.	0.0490277	0.00218055	0.000689551	0.0167469
20. – 22.5	0.0577262	0.00381117	0.0012052	0.0161526
22.5 – 25.	0.062618	0.00218661	0.000691467	0.0155138
25. – 27.5	0.0653363	0.00239468	0.000757264	0.0148491
27.5 – 30.	0.066978	0.00249069	0.000787625	0.0142801
30. – 33.	0.0764202	0.00216215	0.000683731	0.0136772
33. – 36.	0.0761219	0.00246108	0.000778261	0.0131836
36. – 39.	0.0853067	0.00485269	0.00153455	0.0127001
39. – 42.	0.0835377	0.00245159	0.000775261	0.0124663
42. – 45.	0.0908412	0.00268472	0.000848984	0.0122209
45. – 48.	0.0870325	0.00300801	0.000951216	0.0123383
48. – 51.	0.0925727	0.00348033	0.00110058	0.0123622
51. – 54.	0.089258	0.00357317	0.00112994	0.0126788
54. – 57.	0.0829	0.00435847	0.00137827	0.0131511
57. – 61.	0.094012	0.00298472	0.000943851	0.0133462
61. – 65.	0.0920686	0.00319159	0.00100927	0.0140083
65. – 70.	0.0915626	0.0031811	0.00100595	0.0146901
70. – 75.	0.0909227	0.0030153	0.000953521	0.0154339
75. – 80.	0.08731	0.00344346	0.00108892	0.0164048
80. – 85.	0.0913945	0.00421305	0.00133228	0.0171533
85. – 95.	0.0884333	0.00569044	0.00179947	0.0179584
95. – 105.	0.0861138	0.00424156	0.0013413	0.0194852
105. – 125.	0.0833509	0.00371347	0.0011743	0.0211513
125. – 150.	0.0839705	0.00447094	0.00141384	0.022989
150. – 175.	0.0834069	0.00632102	0.00199888	0.0242685
175. – 200.	0.0896245	0.00785136	0.00248282	0.0252912
200. – 250.	0.0792491	0.00902837	0.00285502	0.0262151
250. – 300.	0.0690518	0.0147523	0.0046651	0.0277607
300. – 350.	0.0671591	0.0205242	0.00649034	0.0277056
350. – 400.	0.07426	0.0326327	0.0103194	0.0286537
400. – 470.	0.0752987	0.0362228	0.0114547	0.0281729
470. – 550.	0.0679409	0.0490551	0.0155126	0.0292874
550. – 650.	0.0166015	0.0724604	0.022914	0.0307811
650. – 900.	0.142826	0.0871392	0.0275558	0.0267453
900. – 2500.	−0.109047	0.201254	0.063642	0.0343444

Table 3.10: Expected results of measurements of the $A_0 - A_2$ observable at the LHC at 13 TeV with 300 fb^{-1} and 3 ab^{-1}. The columns show the bin range, the expected value of A_2, the statistical uncertainty with 300 fb^{-1} and 3 ab^{-1}, and a systematic uncertainty of 3%.

$m_{\ell\ell}$ [GeV]	N_{MC}	σ [pb]	$\delta\sigma^{300}_{\mathrm{stat}}$ [pb]	$\delta\sigma^{3000}_{\mathrm{stat}}$ [pb]	$\delta\sigma^{3\%}_{\mathrm{syst}}$ [pb]
$100-105$	8204433	18.1485	0.0118372	0.00374325	0.544455
$105-110$	8349055	8.82233	0.00550703	0.00174147	0.26467
$110-115$	7067831	5.25065	0.00322271	0.00101911	0.15752
$115-120$	7159110	3.50436	0.00213849	0.000676251	0.105131
$120-126$	5813465	2.94182	0.0020837	0.000658923	0.0882547
$126-133$	5797184	2.37394	0.00193485	0.000611855	0.0712182
$133-141$	3755474	1.95064	0.00176662	0.000558653	0.0585193
$141-150$	3813026	1.56363	0.00158484	0.00050117	0.046909
$150-160$	3096496	1.24575	0.00119425	0.000377655	0.0373726
$160-171$	3143515	0.992093	0.00101408	0.000320682	0.0297628
$171-185$	1601319	0.882853	0.00124199	0.000392753	0.0264856
$185-200$	1631284	0.669961	0.000768605	0.000243054	0.0200988
$200-220$	1662754	0.607422	0.000784724	0.000248151	0.0182227
$220-243$	1701384	0.454496	0.000570742	0.000180485	0.0136349
$243-273$	867442	0.368791	0.000635313	0.000200904	0.0110637
$273-320$	890247	0.3257	0.000636502	0.000201279	0.00977101
$320-380$	917527	0.201384	0.000397336	0.000125649	0.00604152
$380-440$	709781	0.100279	0.000218025	0.0000689457	0.00300837
$440-510$	361260	0.0603857	0.000172991	0.0000547046	0.00181157
$510-600$	370583	0.0400292	0.000123339	0.0000390033	0.00120088
$600-700$	380320	0.0211948	0.0000512655	0.0000162116	0.000635844
$700-830$	155170	0.0127793	0.0000579316	0.0000183196	0.000383379
$830-1000$	158533	0.00713393	0.0000317332	0.0000100349	0.000214018
$1000-1500$	162895	0.00501588	0.0000223981	$7.1 \cdot 10^{-6}$	0.000150477
$1500-3000$	166879	0.000848871	$5.0 \cdot 10^{-6}$	$1.6 \cdot 10^{-6}$	0.0000254661
$3000-10000$	2710	$9.5 \cdot 10^{-6}$	$1.1 \cdot 10^{-6}$	$3.4 \cdot 10^{-7}$	$2.9 \cdot 10^{-7}$

Table 3.11: Expected results of measurements of the di-lepton invariant mass spectrum at the LHC at 13 TeV with 300 fb^{-1} and 3 ab^{-1}. The columns show the bin range, the number of Monte Carlo events, the expected cross-section, the statistical uncertainty with 300 fb^{-1} and 3 ab^{-1}, and a systematic uncertainty of 3%.

$m_{\ell\ell}$ [GeV]	A_0	$(\delta A_0)^{300}_{\text{stat}}$	$(\delta A_0)^{3000}_{\text{stat}}$	$(\delta A_0)^{3\%}_{\text{syst}}$
$100 - 105$	0.206069	0.00220099	0.000696013	0.0138179
$105 - 110$	0.198655	0.00390891	0.0012361	0.0140404
$110 - 115$	0.192236	0.0043067	0.0013619	0.0142329
$115 - 120$	0.18408	0.00459632	0.00145349	0.0144776
$120 - 126$	0.178093	0.00527618	0.00166847	0.0146572
$126 - 133$	0.16969	0.00676013	0.00213774	0.0149093
$133 - 141$	0.163594	0.00667648	0.00211129	0.0150922
$141 - 150$	0.157821	0.00815895	0.00258008	0.0152653
$150 - 160$	0.150422	0.0090537	0.00286303	0.0154873
$160 - 171$	0.145611	0.0109549	0.00346424	0.0156316
$171 - 185$	0.137872	0.00920768	0.00291172	0.0158639
$185 - 200$	0.126617	0.0122699	0.00388007	0.0162015
$200 - 220$	0.124962	0.0136978	0.00433163	0.0162512
$220 - 243$	0.106773	0.0130855	0.00413799	0.0167968
$243 - 273$	0.0990207	0.0193292	0.00611242	0.0170294
$273 - 320$	0.097449	0.0197648	0.00625017	0.0170767
$320 - 380$	0.0809004	0.0236554	0.00748048	0.017573
$380 - 440$	0.0768314	0.0331067	0.0104693	0.0176951
$440 - 510$	0.0528575	0.0400658	0.0126699	0.0184143
$510 - 600$	0.0517298	0.0535121	0.016922	0.018448
$600 - 700$	0.0508636	0.0685334	0.0216721	0.0184748
$700 - 830$	0.0460793	0.0885335	0.0279967	0.0186168
$830 - 1000$	0.0420531	0.117322	0.0371003	0.0187368
$1000 - 1500$	0.0289552	0.141289	0.0446794	0.0191276
$1500 - 3000$	0.0236164	0.41192	0.130261	0.0192946
$3000 - 10000$	-0.673888	6.51047	2.05879	0.0415498

Table 3.12: Expected results of measurements of the A_0 observable at the LHC at 13 TeV with 300 fb^{-1} and 3 ab^{-1}. The columns show the bin range, the expected value of A_0, the statistical uncertainty with 300 fb^{-1} and 3 ab^{-1}, and a systematic uncertainty of 3%.

$m_{\ell\ell}$ [GeV]	A_2	$(\delta A_2)^{300}_{\text{stat}}$	$(\delta A_2)^{3000}_{\text{stat}}$	$(\delta A_2)^{3\%}_{\text{syst}}$
$100 - 105$	0.0422219	0.00227595	0.000719719	0.00126666
$105 - 110$	0.0434971	0.00376089	0.0011893	0.00130493
$110 - 115$	0.0417116	0.00381904	0.00120769	0.00125134
$115 - 120$	0.038442	0.00474009	0.00149895	0.00115326
$120 - 126$	0.0355266	0.00459244	0.00145226	0.0010658
$126 - 133$	0.0373691	0.00624652	0.00197532	0.00112104
$133 - 141$	0.0345992	0.00571451	0.00180709	0.00103799
$141 - 150$	0.0364752	0.0083864	0.00265201	0.00109429
$150 - 160$	0.0364136	0.00710229	0.00224594	0.00109239
$160 - 171$	0.0291341	0.00948309	0.00299882	0.000874032
$171 - 185$	0.0356255	0.0085101	0.00269113	0.0010688
$185 - 200$	0.019759	0.0103417	0.00327033	0.000592756
$200 - 220$	0.0294363	0.0106759	0.00337601	0.000883093
$220 - 243$	0.0237177	0.0105273	0.00332902	0.000711509
$243 - 273$	0.0167273	0.0144559	0.00457135	0.000501912
$273 - 320$	0.0221835	0.0139955	0.00442577	0.000665533
$320 - 380$	0.0170735	0.0210878	0.00666854	0.000512299
$380 - 440$	0.0149437	0.0288257	0.00911549	0.00044819
$440 - 510$	0.00187905	0.0365037	0.0115435	0.0000557831
$510 - 600$	-0.00137813	0.0457439	0.0144655	0.0000413211
$600 - 700$	-0.00159732	0.0612091	0.019356	0.0000482714
$700 - 830$	-0.00557428	0.0672415	0.0212636	0.000167436
$830 - 1000$	0.0137169	0.10256	0.0324322	0.000410183
$1000 - 1500$	0.0317234	0.122617	0.038775	0.000951331
$1500 - 3000$	-0.00716862	0.396067	0.125248	0.000207753
$3000 - 10000$	2.86876	20.4434	6.46477	0.0919646

Table 3.13: Expected results of measurements of the A_2 observable at the LHC at 13 TeV with 300 fb^{-1} and 3 ab^{-1}. The columns show the bin range, the expected value of A_2, the statistical uncertainty with 300 fb^{-1} and 3 ab^{-1}, and a systematic uncertainty of 3%.

$m_{\ell\ell}$ [GeV]	$A_0 - A_2$	$\delta(A_0 - A_2)^{300}_{\text{stat}}$	$\delta(A_0 - A_2)^{3000}_{\text{stat}}$	$\delta(A_0 - A_2)^{3\%}_{\text{syst}}$
$100 - 105$	0.0489227	0.00379057	0.00119868	0.0146
$105 - 110$	0.0479573	0.00693351	0.00219257	0.0147503
$110 - 115$	0.0527973	0.00846707	0.00267752	0.0148349
$115 - 120$	0.0466155	0.00778599	0.00246215	0.0150535
$120 - 126$	0.0443988	0.0108978	0.0034462	0.0151961
$126 - 133$	0.0483033	0.0123165	0.00389482	0.0153476
$133 - 141$	0.0355806	0.0115509	0.00365273	0.0155731
$141 - 150$	0.0366276	0.0141871	0.00448637	0.0156924
$150 - 160$	0.0365152	0.0167127	0.00528503	0.0158598
$160 - 171$	0.0364686	0.0246929	0.00780859	0.0159708
$171 - 185$	0.0487894	0.0189455	0.00599109	0.0160874
$185 - 200$	0.0310485	0.0205385	0.00649484	0.0164532
$200 - 220$	0.0324652	0.0236061	0.00746492	0.0164864
$220 - 243$	0.0265552	0.0210694	0.00666273	0.0169683
$243 - 273$	0.0127414	0.0364795	0.0115358	0.0172249
$273 - 320$	0.039359	0.0307471	0.00972308	0.0171653
$320 - 380$	0.0332408	0.0434818	0.0137501	0.0176311
$380 - 440$	0.0179427	0.0558835	0.0176719	0.017783
$440 - 510$	-0.0132197	0.080222	0.0253684	0.0185206
$510 - 600$	0.0147696	0.1048	0.0331407	0.0184813
$600 - 700$	0.0223771	0.134543	0.0425464	0.0184946
$700 - 830$	-0.0145346	0.174328	0.0551275	0.0187056
$830 - 1000$	0.0364093	0.226636	0.0716687	0.0187375
$1000 - 1500$	-0.0000212533	0.269824	0.0853259	0.0191475
$1500 - 3000$	0.0905146	0.904677	0.286084	0.0193981
$3000 - 10000$	3.38865	37.2917	11.7927	0.146803

Table 3.14: Expected results of measurements of the $A_0 - A_2$ observable at the LHC at 13 TeV with 300 fb^{-1} and 3 ab^{-1}. The columns show the bin range, the expected value of $A_0 - A_2$, the statistical uncertainty with 300 fb^{-1} and 3 ab^{-1}, and a systematic uncertainty of 3%.

3.F UV models sensitive to our analysis

One of the goals of the SMEFT program is that of indirectly exploring the landscape of possible UV completions of the SM. Generally, this is done by constraining, through precision measurements, the Wilson coefficients of classes of SMEFT operators, and then match them to a specific BSM model, or, even better, to a class of BMS models. Indeed, it is essential to keep in mind that the SMEFT approach, being fully model-independent, gives little information on the UV, unless it is supplemented by additional hypotheses. For instance, constraints on electroweak precision observables give very different information on the UV if one assumes strongly interacting dynamics at the TeV scale, or weakly coupled extensions of the SM. In other words, the SMEFT gains all its power only when supplemented by reasonable hypotheses at least on the BSM framework that can give rise to the observables at hand.

An interesting question that one could ask on the present paper is which information on the UV one could gather from studying the operators that we consider. First of all, the fact that dipole and scalar/tensor four-fermion operators break chiral symmetry, rules out the possibility of leveraging our analysis to constrain minimal flavor violating (MFV) UV models. Indeed, if MFV is assumed, then selection rules imply that the Wilson coefficients of chirality-breaking operators are suppressed by (powers of) Yukawa couplings, making our bounds too weak to draw any relevant conclusion. Parity is the other selection rule that makes all constraints on parity-odd operators, both four-fermion and dipole operators, the latter corresponding to Electric Dipole Moments (EDMs) from high energy measurements essentially irrelevant. This is the reason why we considered only parity conserving operators. Concerning parity conserving dipole operators, corresponding to Magnetic Dipole Moments (MDMs), the impact of UV physics is quite well understood, since they often arise from one loop diagrams that contribute to the $\mathrm{U}(1)_Y$, $\mathrm{SU}(2)_L$ and $\mathrm{SU}(3)_c$ vertex correction, to which there is usually a limited set of contributing Feynman diagrams. The general structure of these operators is therefore relatively easy to understand. For instance, Ref. [110] studies the one-loop contributions to dipole moments in several UV theories, such as two-Higgs-doublet models, the minimal supersymmetric extension of the SM (MSSM), and scenarios with extra U(1) gauge symmetries, and performs the matching with the SMEFT operators that we consider: the general result is that, within the framework of the aforementioned BSM theories, suitable parameter choices can lead to sizable Wilson coefficients for the CP-even dipole operators. These parameter choices are often fine-tuned, which is expected for sizable chirality breaking contributions, but are still possible. In this respect, our bounds can be used to constrain such parameter choices, and to limit the possible fine-tuning available in some of these BSM scenarios.

Let us now briefly discuss four-fermion operators. As discussed, for instance, in Refs. [111, 112], they can be generated at tree level in BSM models which include leptoquark states. In particular the operators $Q^{(1)}_{\ell equ}$ and $Q^{(3)}_{\ell equ}$ can be generated by integrating out at tree level scalar leptoquarks, while $Q_{\ell edq}$ can be generated by integrating out at tree-level vector leptoquarks. Notice that, because of their chiral structure, the operators $Q_{\ell edq}$ and $Q^{(1)}_{\ell equ}$ introduce terms in the pion decay width

$\Gamma(\pi \to \ell\nu_\ell)$ that are proportional to the pion mass [113]. In particular, one finds that the ratio between the SM and the NP contribution scales as

$$\frac{\Gamma(\pi^+ \to \ell^+\nu_\ell)^{\rm SM}}{\Gamma(\pi^+ \to \ell^+\nu_\ell)^{\rm NP}} \propto \frac{m_\pi^2}{m_\ell(m_d + m_u)}. \tag{3.54}$$

This contribution is reasonable for the muon channel, but it is hugely enhanced for the electron channel. As a consequence, one can check that for Λ of $\mathcal{O}(1\text{ TeV})$, the present experimental measurement of the ratio [114]

$$R \equiv \frac{\Gamma(\pi^+ \to e^+\nu_e)}{\Gamma(\pi^+ \to \mu^+\nu_\mu)} = (1.230 \pm 0.004) \times 10^{-4}, \tag{3.55}$$

implies, for the electron channel, $c_{ledq} + c^{(1)}_{lequ} \sim \mathcal{O}(10^{-4})$.[15] For this reason, as already mentioned in the main text, our bounds on the $Q_{\ell edq}$ and $Q^{(1)}_{\ell equ}$ operators are only relevant for the second family.

[15] This was also pointed out in Refs. [115, 116].

Part II

A robust methodology for comparing test statistics in the validation of Generative Models

Chapter 4

From Traditional Simulations to Machine Learning Generators: The Need for Robust Validation

In this Chapter, we discuss the role of simulations in high energy physics (HEP) and why the community is starting to adopt machine-learning models to help the traditional Monte Carlo generators in providing simulated data. The advent of the High-Luminosity (HL) phase of the Large Hadron Collider (LHC) will result in higher experimental precision, which in turn demands more precise simulations. However, reaching such precision just with traditional generators can become computationally prohibitive. Machine learning generators offer a promising support to them, as they can produce large samples of simulated data much faster. Nonetheless, their reliability cannot be assumed a priori, therefore a validation procedure is necessary. Throughout this second part of the thesis we introduce a robust methodology, along with some new tools, that are designed for the validation of generative models in this context.

In Section 4.1, we explain why traditional Monte Carlo generators may become a computational bottleneck when trying to produce synthetic data matching the experimental accuracy, in particular in view of the requirements of the HL-LHC era. In Section 4.2 we provide a brief introduction to ML-based generators, going into some more details for few representative architectures. As usual, we do not aim to provide an in depth analysis, as extensive literature on the subject is already available [117–128]. In Section 4.3 and 4.4 we review how machine learning techniques can be employed in the event generation and detector simulation pipelines, respectively. In Section 4.4 we motivate the need for a rigorous validation procedure for ML-generated synthetic data and we argue that two-sample testing provides a natural framework for this task. Finally, in Section 4.6, we introduce two-sample testing from a technical perspective and describe how this statistical tool is used in Chapters 5 and 6.

4.1 Traditional generators and their limitations

As discussed in the first part of this thesis, the Standard Model (SM) is the theory describing our current understanding of particle physics. To test and validate it, its predictions has to be compared with experimental measurements. In the high energy physics (HEP) context, colliders, such as the LHC, are the primary sources of data. A crucial link between theory and experiment is provided by simulations. Traditional Monte Carlo generators, such as MadGraph [109, 129], Powheg [64, 84], PYTHIA8 [65], Sherpa [130] and others [129, 131, 132], are able to produce parton level events directly from the SM Lagrangian while also modeling subsequent phenomena such as electroweak and QCD parton showering. Other tools, such as GEANT4 [133], are able to simulate the detector response, enabling a complete comparison between theoretical predictions and experimental measurements. We provide a schematic description of the LHC simulation chain [134] that is pictorially represented in Figure 4.1.

- **Theory:** A lagrangian describing the underlying theory. It is often the Standard Model (SM), but it can also represent one of its extensions.
- **Scattering:** The partonic hard process that follows directly from the lagrangian of the theory, representing the most fundamental level of the interaction.
- **Decays:** Unstable particles produced in the scattering process decay into lighter ones according to the theory.
- **Perturbative higher order corrections:** Perturbative QCD and EW corrections are included to account for higher-order effects.
- **Shower:** The decay products undergo parton showering, emitting further quarks and gluons.
- **Fragmentation:** The partons in the shower "hadronize", forming stable hadrons such as baryons and mesons.
- **Detector:** The resulting particles interact with the detector, depositing energy and leaving traces of their passage.
- **Events:** The detector signals are reconstructed into physical observables to be compared with experimental data.

Although simulations play a crucial role and provide reliable results, they come with a high computational cost. In the specific case of the LHC, the amount of experimental data is expected to increase significantly in the near future, particularly with the transition to the High-Luminosity (HL) phase. Figure 4.2 shows the projected evolution of the peak and integrated luminosity of the LHC over the years, up to the end of its operation, around 2040. As evident from the figure, the integrated luminosity will grow to around a factor ten higher compared to current levels. Consequently, the

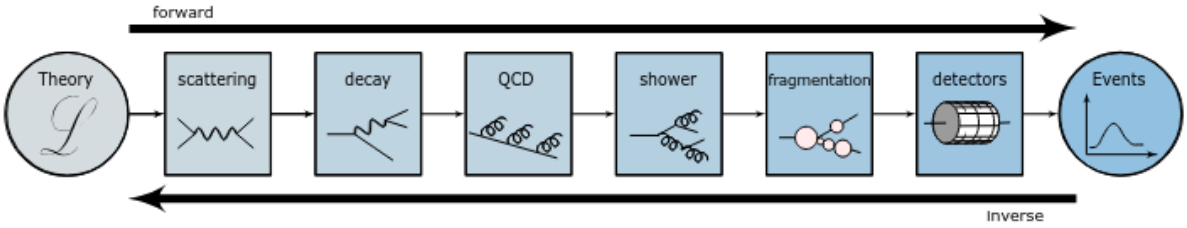


Figure 4.1: Illustration of the LHC simulation chain. Both the forward and backward direction are necessary to compare theory with data.

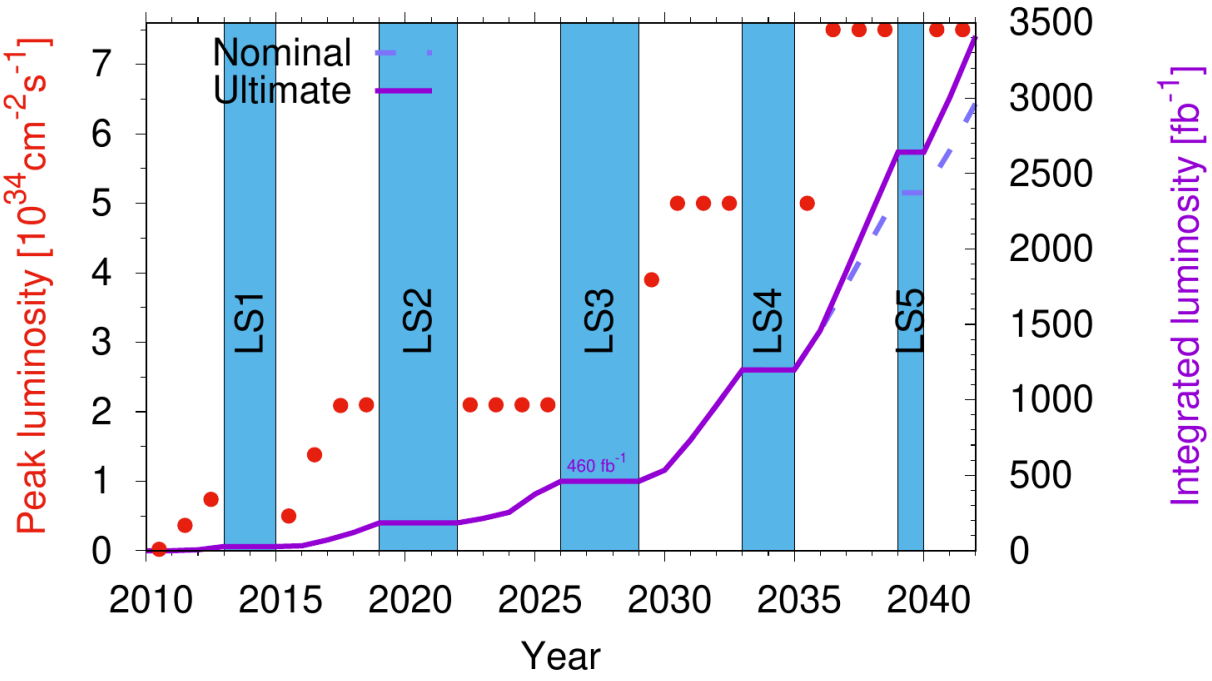


Figure 4.2: Peak and integrated luminosity in the different phases of the LHC, year by year

statistical uncertainties will approximately be reduced by a factor ~ 3:

$$\frac{\delta_{\text{stat,now}}}{\delta_{\text{stat,HL-LHC}}} \sim \frac{\frac{1}{\sqrt{N_{\text{now}}}}}{\frac{1}{\sqrt{N_{\text{HL-LHC}}}}} \sim \sqrt{\frac{\mathcal{L}_{\text{HL-LHC}}}{\mathcal{L}_{\text{now}}}} \sim 3\,. \tag{4.1}$$

Along with the increase in statistics, systematic uncertainties are also expected to improve, thanks both to higher data samples and to advancements in detector technologies and theoretical understanding. As a result, experimental uncertainties at and below the percent level will become standard [134]. Theoretical predictions, including those obtained from simulations, must achieve a corresponding level of precision in order to be compared with data. This requirement poses a serious challenge for the community, given that reaching such precisions in simulations implies exploiting a huge number of computational resources [135]. Figure 4.3 shows the projected Central Processing Units (CPU) requirements for the ATLAS (left) and CMS (right) experiments over the

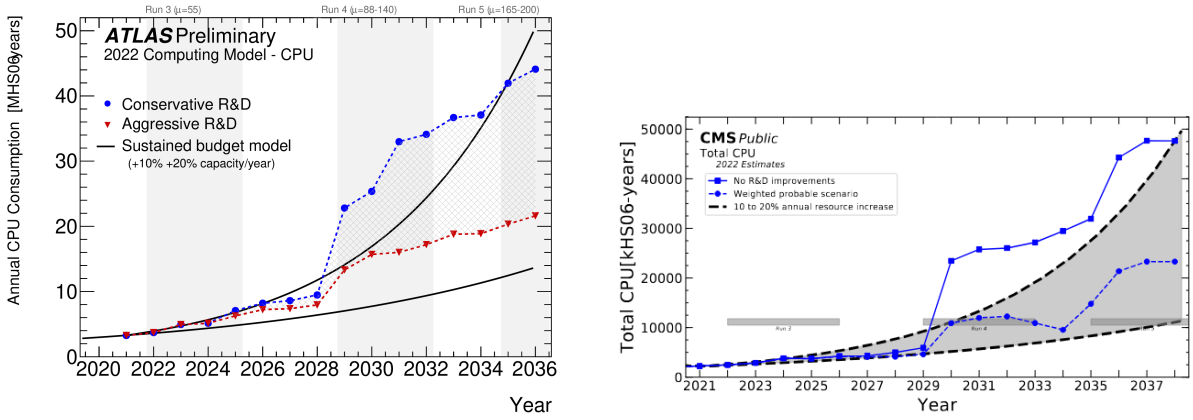


Figure 4.3: Projected CPU requirements for ATLAS [136] (left) and CMS [137] (right).

years until around the end of the LHC operations. Two hypotheses are considered, accounting for a conservative and a optimistic scenario, along with fixed annual increases of 10% and 20%. It is evident that relying solely on traditional simulation tools will become unfeasible in terms of both computational time and resource requirements. These projections, along with recent technology developments, pushed the HEP community to explore alternative approaches that can assist Monte Carlo generators in simulating events.

Machine Learning (ML), in particular, has gained significant importance in recent years, finding applications in a wide range of tasks such as classification, regression, and anomaly detection. In the context described above, generative models have attracted particular attention. In fact, their ability to produce synthetic data more efficiently than the high-fidelity Monte Carlo generators mentioned above, could represent a major development in HEP, providing a powerful tool to accelerate simulations. In particular, two aspects of the simulation chain can be sped up: the event generation and the detector simulation. These two topics will be discussed in the following Sections.

4.2 Machine Learning generative models in a nutshell

In this Section, we provide a brief introduction to some basic Machine Learning (ML) concepts. Our aim is just to provide some context for ML generators, which have become the focus of many HEP studies as a faster and less resource-intensive alternative to traditional Monte Carlo generators. We also briefly describe some of the most common ML–based generative models. Our goal is not to provide an exhaustive overview, since extensive literature exists for more in depth discussions both about ML in general [138–140] and generative models in particular [117, 118], but rather to offer an intuitive understanding of how these models works. We will focus on the Generative Adversarial Networks (GANs) [119, 120], Variational Autoencoders (VAE) [121, 122] and Normalizing Flows (NFs) [123, 124], while acknowledging that other possible generators have being explored and used in the HEP context [125–128].

ML can be regarded as an extension of traditional statistical data inference, as a set of methods

able to detect patterns and structures within the data. What distinguishes ML from conventional fitting techniques is the use of highly flexible, non-linear models (for instance neural networks) that can recognize intricate correlations within data without assuming a fixed functional form. Broadly speaking, ML tasks can be divided into two categories: *supervised* and *unsupervised* learning.

In supervised learning, each input data point is accompanied by a label. Formally, the dataset can be indicated with $\mathcal{D} = \{(\boldsymbol{x}_i, y_i)\}_{i=1}^{N}$, where $\mathcal{D}$ denotes the dataset, the index i runs over the N points within the sample, $\boldsymbol{x}_i$ is the i-th input point within the dataset, typically referred to as *feature*, (it can be a scalar, a vector, or something more complex like an image) and y_i is the corresponding label. The label y_i can either be a categorical variable (e.g., cat, dog, male, female, $\cdots$) or a real valued scalar. In the former case, the corresponding learning task is known as *classification*, while in the latter, it is called *regression*. For supervised learning, an objective is to find:

$$\hat{y} = \hat{f}(\boldsymbol{x}) = \arg\max_{c=1}^{C} p(y = c|\boldsymbol{x}, \mathcal{D}) \,, \tag{4.2}$$

where c is an index running over the C possible classes in the problem and $p(y = c|\boldsymbol{x}, \mathcal{D})$ denotes the probability distribution over possible labels, given the input vector $\boldsymbol{x}$ and training set $\mathcal{D}$. This corresponds to the Bayes-optimal classifier, which minimizes the expected misclassification error by choosing the most probable label given the data.

In unsupervised learning, instead, no labels are provided and the purpose is to find interesting patterns or structures within the data. In this case the dataset can be indicated as $\mathcal{D} = \{\boldsymbol{x}_i\}_{i=1}^{N}$. One possible task consists of estimating the probability density function (PDF) that had generated the data. Using a parametric model, the best parameters are obtained by maximizing the likelihood:

$$\hat{\theta} = \arg\max_{\theta} \prod_i p(\boldsymbol{x}_i|\theta) \,. \tag{4.3}$$

This procedure is called *density estimation*. More generally, unsupervised learning also includes other tasks, like clustering and dimensionality reduction, that aim at uncovering the latent structure of the data without using labels.

Generative models represent a specific class of unsupervised ML applications of great interest in HEP [141]. In practice, generative models aim to learn a model distribution $p_\theta(\boldsymbol{x})$ that approximates the true (and typically unknown) data distribution $p^*(\boldsymbol{x})$. Once $p_\theta(\boldsymbol{x})$ is available, it is possible to sample from it at will, allowing the generation of an arbitrary number of synthetic data points that are "similar" to the original ones. This capability is particularly promising in the context discussed in the previous Section, where traditional Monte Carlo generators can represent a computational bottleneck in the production of large simulated datasets. Different architectures implement this idea through distinct mechanisms: adversarial learning in GANs, latent variable modelling in VAEs, and exact likelihood optimization in NFs. In the following, we provide a brief description of how GANs, VAEs and NFs works.

GANs:

Generative Adversarial Networks [119, 120] are among the earliest deep generative models and became famous thanks to their ability to reproduce photorealistic images. They consist of two neural networks, a generator and a discriminator, that are trained in opposition to each other. The generator G maps latent vectors, $\boldsymbol{z} \sim p(\boldsymbol{z})$, into synthetic data that resemble the real ones. The discriminator D is trained to distinguish between true data, $\boldsymbol{x} \sim p(\boldsymbol{x})$, and the generator output $G(\boldsymbol{z})$. The two networks are updated alternately, improving at each step, until the generator is able to fool the discriminator which can no longer distinguish between $G(\boldsymbol{z})$ and the real data. Formally, the GAN objective can be written as the following minimax optimization problem:

$$\min_G \max_D \mathbb{E}_{\boldsymbol{x}\sim p(\boldsymbol{x})}[\ln D(\boldsymbol{x})] + \mathbb{E}_{\boldsymbol{z}\sim p(\boldsymbol{z})}[\ln(1 - D(G(\boldsymbol{z})))] \,. \tag{4.4}$$

GANs generally provide high quality synthetic data, but their training is unstable and computationally demanding, due to the adversarial nature of the optimization process.

VAEs:

An autoencoder is a network that learns to reconstruct data input data $\boldsymbol{x}$ by compressing them through an encoder and decompressing them through a decoder. In order to be able to generate new data, a Variational Autoencoder [121, 122] is used. Instead of just being compressed, the input $\boldsymbol{x}$ is mapped into a distribution. In particular, a generative model model $p_\theta(\boldsymbol{x}|\boldsymbol{z})$ is considered, with a prior $p_\theta(\boldsymbol{z})$ and posterior $p_\theta(\boldsymbol{z}|\boldsymbol{x})$. Learning the posterior is impractical due to the integral $p_\theta(\boldsymbol{x}) = \int p_\theta(\boldsymbol{x}|\boldsymbol{z})p_\theta(\boldsymbol{z})$. To overcome this, the true posterior is approximated by a parameterized function $q_\phi(\boldsymbol{z}|\boldsymbol{x})$. This structure resembles an autoencoder where $q_\phi(\boldsymbol{z}|\boldsymbol{x})$ takes a similar role as the encoder one, and in fact is called *probabilistic encoder*, while $p_\theta(\boldsymbol{x}|\boldsymbol{z})$, takes the role of the decoder, and is known as *probabilistic decoder*. The model is trained by maximizing the Evidence Lower Bound (ELBO):

$$\mathcal{L}(\theta, \phi; x) = \mathbb{E}_{\boldsymbol{z}\sim q_\phi(\boldsymbol{z}|\boldsymbol{x})}[\log p_\theta(\boldsymbol{x}|\boldsymbol{z})] - D_{\mathrm{KL}}(q_\phi(\boldsymbol{z}|\boldsymbol{x})||p_\theta(\boldsymbol{z})) \,, \tag{4.5}$$

where D_{KL} is the Kullback-Leibler divergence. VAEs provide a stable and interpretable generative framework, but the regularization imposed on the latent space can limit their expressive power.

NFs:

They idea behind Normalizing Flows [123, 124] is to use a sequence of bijective transformations to map a simple distribution $p_Z(\boldsymbol{z})$, such as a gaussian, into the probability density function of the real data $\boldsymbol{x}$. Consider an invertible, smooth function: $f : \mathbb{R}^d \to \mathbb{R}^d$, where d is the dimensionality of the data and apply this transformation to a random variable $\boldsymbol{z} \sim p_Z(\boldsymbol{z})$. The distribution of the transformed variable $\boldsymbol{y} = f(\boldsymbol{z})$ is then defined by the chain rule:

$$p_Y(\boldsymbol{y}) = p_Z(\boldsymbol{z}) \left| \det \frac{\partial f^{-1}}{\partial \boldsymbol{y}} \right| = p_Z(\boldsymbol{z}) \left| \det \frac{\partial f}{\partial \boldsymbol{y}} \right|^{-1} . \tag{4.6}$$

Applying a sequence of such invertible transformations, that depend on parameters $\boldsymbol{\theta}$, defines a chain called a normalizing flow. Denoting by $\boldsymbol{x}_k \sim p_{X_k}(\boldsymbol{x}_k)$, with $k = 1, \ldots, K$ the variable after the k-th transformation, and defining $\boldsymbol{x}_0 \equiv \boldsymbol{z}$, we can write:

$$\boldsymbol{x}_K = f_{K,\boldsymbol{\theta}} \circ \ldots f_{2,\boldsymbol{\theta}} \circ f_{1,\boldsymbol{\theta}}(\boldsymbol{x}_0)\,, \tag{4.7}$$

and the corresponding log-probability as:

$$\ln p_{X_k,\boldsymbol{\theta}}(\boldsymbol{x}_K) = \ln p_{X_0}(\boldsymbol{x}_0) - \sum_{k=1}^{K} \ln \left| \det \frac{\partial f_{k,\boldsymbol{\theta}}}{\partial \boldsymbol{x}_{k-1}} \right| . \tag{4.8}$$

The loss is simply the maximum likelihood on the real data $\boldsymbol{x}$:

$$\mathcal{L} = \max_{\boldsymbol{\theta}} \sum_{i=1}^{N} \ln p_{X_k,\boldsymbol{\theta}}(\boldsymbol{x})\,, \tag{4.9}$$

and the optimal parameters are:

$$\boldsymbol{\theta}^* = \arg\max_{\boldsymbol{\theta}} \sum_{i=1}^{N} \ln p_{X_k,\boldsymbol{\theta}}(\boldsymbol{x})\,. \tag{4.10}$$

Normalizing Flows offer an exact and tractable likelihood, enabling both sampling and density estimation. However, the requirement of invertibility and Jacobian computation make them computationally expensive and less flexible than other neutral networks.

4.3 The role of Machine Learning in event generation

As discussed previously, matching the precision requirements imposed by future experimental improvements demands more accurate simulations. In this Section, we focus specifically on how ML can assist in event generation, without including the detector effects.

Current Monte Carlo simulations rely on perturbative methods to model the hard process and QCD parton showers, while phenomenological models are used to describe the hadronization transition. Possible improvements can involve, for instance, the evaluation of higher orders tree-level and loop amplitudes. Machine Learning techniques can enhance all stages of the simulation chain, improving computational efficiency while maintaining or even increasing accuracy. In the following, we provide a brief overview of how ML can contribute to some steps in the event generation [134].

- **Phase space sampling:** To compute cross Sections, scattering amplitudes must be integrated over the phase space. Traditionally, this is achieved through importance sampling algorithms, such as VEGAS, that try to sample more points where the integrand, namely the matrix element, is large. Machine Learning methods can enhance this procedure in two ways. First, they can can be embedded within the traditional sampling framework, improving the mapping

from random numbers to phase space points. For example, a Normalizing Flow can provide a distribution that approximates the matrix element, allowing for more efficient sampling. The other possibility consists in training ML generative models directly on the events, both weighted and unweighted, that are provided by traditional generators, learning the event distribution directly from data. GANs or VAEs, for instance, can be used in this context, effectively acting as fast surrogate event generators.

- **Scattering amplitudes:** Computing scattering amplitudes is the first step that follows directly from theory, and provides the matrix elements that encode the dynamics of the underlying physics model. Traditionally, amplitudes are computed in perturbation theory through Feynman diagrams. Including higher order corrections rapidly increases both the number and the complexity of diagrams, making the computation of next-to-leading order (NLO) and next-to-next-to-leading order (NNLO) amplitudes highly demanding. Machine Learning techniques can help in this context, approximating matrix elements or accelerating amplitudes evaluation. Neural networks can be trained to learn the structure of complex scattering amplitudes, providing fast surrogate models that reproduce the dependence on the kinematic variables. Moreover, ML can also help optimizing diagram evaluation, symbolic simplification, and amplitude reduction.

- **Loop integrals:** Loop integrals typically present poles, due to ultraviolet or infrared divergences as well as other features, such as thresholds. Traditional tools handles these singularities, in particular when they lie on the real axis in Feynman parameter space, through a deformation of the integration contour into the complex plane. The choice of the contour is not unique, and the precision of the results is highly dependent on this choice. ML techniques can help in selecting an optimal contour, in order to enhance the accuracy of the computation. This can be particularly impactful when the integrand has a complicated form or in specific phase space region where the integral is hard to evaluate. Tools like normalizing flow can be employed to identify a more efficient parametrization of the integration domain, leading to better contour choices and more reliable results.

- **Parton shower:** Partons produced in the hard process undergo a cascade of successive splittings, emitting other partons until they reach the hadronization scale Λ_{QCD} and form stable particles. Traditional generators use a semi-classical approach to simulate this cascade, in which emissions are modelled through probabilistic QCD splitting functions. However, this approach is theoretically justified just when the daughter particles are emitted at small angles with respect to the mother. It is used, nonetheless, because computing the full distribution from first principles is computationally intractable. ML can help simulating the parton shower in different ways. They can be trained to learn correlations between the hard process and the final detector-level signatures, or they can be embedded within traditional shower algorithms to provide physically motivated corrections and refinements to the semi-classical description. Typically, the final state shows permutation invariance, that can be learned by architectures

such as Graph Neural Networks (GNN). Furthermore, ML can be used in low energy QCD experiments to extract improved, data-driven versions of the splitting kernels that govern the shower dynamics.

- **PDFs:** Parton distribution functions (PDFs) can be interpreted as probability densities to find a parton carrying a fraction x of the proton momentum. They are of fundamental importance since their convolution with partonic cross-sections provides the hadronic cross-section. In practice, PDFs are extracted from experimental data, as calculating them from first principles is unfeasible. Machine Learning techniques can assist in PDFs determination, as this task can be regarded as a pattern recognition problem, where the pattern is a probability distribution. Since both theoretical and experimental uncertainties affect the input data, the ML procedure yields an ensemble of possible PDFs as output, each consistent with the data within its uncertainties. The original idea was therefore to construct an ensemble of ML models, whose collective output represents the probability distribution of the possible PDFs. This idea led to the NNPDF[142] collaboration, which developed and optimized this methodology to produce reliable descriptions of proton structure. The main challenge today remains achieving optimal information extraction while avoiding overfitting, as well as extending PDF determinations into kinematic regions where few or no data exist. These regions are in fact particularly relevant for NP searches.

- **Fragmentation functions:** Fragmentation Functions (FFs), usually denoted with $D_i^h(z, Q^2)$, encode the probability for the parton i, produced in the hard scattering, to hadronize into a final state hadron h carrying a fraction z of the parton momentum at the energy scale Q^2. As for the PDFs, these functions cannot be calculated from first principles, but rather they have to be fitted from data, particularly from electron-positron collisions, semi-inclusive deep inelastic scattering and proton-proton collisions. Traditionally, a functional form for the FF is introduced at an initial scale Q_0, with its parameters fitted to experimental data. This procedure introduces a bias that can distort the final result. Feed forward neural networks (NNs) can be employed in this context to provide a universal, unbiased interpolator for $zD_i^h(z, Q^2)$, since no functional form is imposed and the network parameters are learned directly from data. As in the PDF case, different NNs can be trained on replicas of the same dataset fluctuated within its uncertainties, providing a probability distribution for the resulting FF.

Enhancing the performance in simulating different steps of the chain, as the ones discussed above, is a promising direction for employing ML in HEP simulations.

Another possibility that is being explored by the community, is to bypass the step-by-step approach. In this case, ML models are trained directly on complete final state events, with the aim of learning the underlying mapping between theoretical inputs and the corresponding final state observables. Such ML generators are designed to directly provide final events, without simulating each step of the chain separately. For this reason, they are referred to as *end-to-end generators* [134, 143, 144]. Typically, they are implemented using architectures such as GANs, VAEs and NFs, each

offering specific advantages and disadvantages, as we briefly described in the previous Section. These approaches enable much faster data generation and reduce computational requirements. Future developments may focus on including physics constraints (such as Lorentz invariance and momentum conservation) within the training process, as well as training these models on real data rather than on simulated samples.

4.4 Detector simulation

In this Section we briefly describe how the detector response, and specifically the calorimeter simulation, are currently handled in HEP and how ML can assist in providing more efficient results.

As shown in Figure 4.1, simulating the detector response is the last step of the full simulation chain. It is a fundamental stage, that enables the connection between theoretical particle-level predictions and the quantities that are actually measured in experiments. The final state particles provided by event generators act as primary particles entering the detector, interacting with its subparts. The interaction with the calorimeter material, in particular, leads them to undergo electromagnetic and hadronic showers, which determine both the amount of deposited energy and its spatial distribution. This step is among the most computationally demanding of the entire simulation chain, since each shower involves a large number of secondary interactions that must be simulated in detail, resulting in significant CPU usage. Traditionally, these interactions are modeled using GEANT4 [133], a framework that provides highly accurate description of particle–matter interactions but at a considerable computational cost. To overcome this bottleneck, several faster alternatives, usually referred to as *fast simulations* [145–148], have been developed over the years. These approaches rely on parametrized response of the calorimeter with respect to specific types of incoming particles, effectively avoiding the complex simulation of the particle shower that is carried on by GEANT4. Although such models provide a significant speedup in the calorimeter simulation, they typically lack the level of accuracy that is crucial, for the reasons we specified in 4.1, for the future challenges the community has to face.

Machine Learning, on the other hand, offers a valuable alternative, capable of combining the high accuracy of GEANT4 with the speed of fast-simulation techniques. Generative models, in particular, play a crucial role in this context: once trained on a sufficiently large set of examples of the detector response for a given type of incoming particle, they learn the underlying distribution of calorimeter showers and provide new, statistically consistent showers. This approach has already been studied in literature [149–151] using different architectures, such as GANs [119, 120], VAEs [121, 122], NFs [123, 124] and diffusion models [125, 126] and have shown promising performance. These results suggest that ML-based approaches may evolve into a viable and sustainable alternative for calorimeter simulation in the HL-LHC era, complementing full simulation in the most computationally demanding phases.

4.5 Model validation through two-sample testing

In the previous chapters, we explained why ML-based approaches can be game changers in modern particle physics, providing a significant speed-up in simulations when used to assist traditional Monte Carlo generators. We also highlighted the necessity for precise synthetic data, to match the growing experimental measurement accuracy, in particular in light of the HL-LHC era.

Although slow, traditional Monte Carlo generators can be sampled at will to provide arbitrarily high fidelity data, while no a priori guarantee exists that machine-learning generated data respects the accuracy standard necessary in this context. In factx, their statistical power is limited by the statistical uncertainty in the training data. For this reason, another important research direction has emerged, referred to as the validation of generative models. This field aims to ensure that the synthetic data produced by machine-learning based models meets the precision standards required for a given physics application. However, many existing methods lack a rigorous statistical foundation, making it difficult to provide robust and reliable evaluations, particularly in high-stakes scientific applications. In this second part of this thesis, we aim to provide a robust statistical approach in order to reliably test and validate synthetic data produced by ML-based generators against high-fidelity ones.

A natural statistical framework to evaluate generative models is two-sample hypothesis testing, which aims to determine whether two independent samples are drawn from the same distribution. These tests fall into two categories: *parametric* and *non-parametric* tests. Parametric tests, such as the (log-)likelihood ratio (LLR) test [152], identified by the Neyman-Pearson lemma as the most powerful for simple hypotheses, rely on specific assumptions about the underlying data distributions, like normality, and are highly effective when these assumptions hold. In contrast, non-parametric tests make no such assumptions, providing greater robustness at the cost of a potentially lower power. Although numerous non-parametric tests exist for both one- and multi-dimensional data, most are not computationally efficient in high-dimensional settings, where the curse of dimensionality poses a significant challenge.

Machine learning (ML) techniques, particularly classifiers that approximate the LLR or a related function, have proved promising in designing two-sample tests for evaluating generative models in HEP [153–158]. In particular, the model presented in Ref [153] and further developed and detailed in Ref.s [154, 156, 157] allows for a statistically robust evaluation as a goodness-of-fit and has been recently proposed for evaluating of generative models [159].

However, assessing the performance of these classifiers remains challenging. Many generative models developed for HEP have not yet achieved high accuracy beyond the first few moments (e.g. mean and covariance), leading to tests that may be overly sensitive to minor, less significant differences. As a result, interpreting strong rejections of models in these tests can be difficult. This highlights the need for a set of robust, simple, and interpretable two-sample tests that can serve as benchmarks for evaluating more advanced ML-based approaches.

For this reason, we focus at first on non-parametric tests based on univariate integral prob-

ability measures (IPMs). Our approach builds on two well-established methods for comparing one-dimensional, in short 1D, distributions, both based on the empirical Cumulative Distribution Function (eCDF): the Kolmogorov-Smirnov (KS) test [160, 161] and the Wasserstein (W) distance [162, 163]. Thanks to their simplicity, these metrics are particularly well suited for this purpose, as they are straightforward to interpret and computationally efficient. Once we exploited these simple metrics, we will focus on a kernel-based classifier, the New Physics Learning Machine (NPLM) [154, 156], in order to compare its performance with the ones of the simple 1D-based metrics.

4.6 Two-sample hypothesis testing

In this Section we provide a brief explanation of what is generally meant with two-sample testing and how we will use this technique in the framework of Chapter **??** and **??**.

Given two random variables X and Y, defined on a space $\mathcal{S} \subseteq \mathbb{R}^d$, let us consider two samples $\mathcal{X} = \{x_i\}$, where $i = 1, \ldots, n$, and $\mathcal{Y} = \{y_j\}$, where $j = 1, \ldots, m$. Using capital Latin letters $I, J = 1, \ldots, d$ to denote the components of d-dimensional vectors, we can indicate with x_{iI} the scalar value of the i-th element of sample $\mathcal{X}$ along the I-th dimension. We also assume that the two samples are independent and identically distributed according to the distributions p and q, respectively.

Two-sample testing aims at determining whether the null hypothesis at the population level,

$$H_0 : p = q, \tag{4.11}$$

can be rejected based on finite data. The alternative hypothesis is the negation of the null:

$$H_1 : p \neq q. \tag{4.12}$$

The test proceeds by selecting a *test statistic*, $t : (\mathcal{S})^n \times (\mathcal{S})^m \to \mathbb{R}$, and calculating its value on the observed data:

$$t_{\text{obs}} = t(\mathcal{X}, \mathcal{Y}). \tag{4.13}$$

To define a binary test, the observed value of the test statistic is compared to a threshold, t_α, where

$$\alpha = P(t \geq t_\alpha | H_0) = \int_{t_\alpha}^{\infty} f(t|H_0) dt, \tag{4.14}$$

with $f(t|H_0)$ being the distribution of the test statistic under the null hypothesis H_0, represents a preselected probability of a type-I error, i.e. the *rate of false positives*. The null hypothesis is not rejected if $t_{\text{obs}} < t_\alpha$. Alternatively, one can report the *p-value*, defined as the probability of obtaining a test statistic as extreme as t_{obs} under H_0:

$$\text{p}_{\text{obs}} = P(t \geq t_{\text{obs}} | H_0). \tag{4.15}$$

This quantity can also be mapped to a *Z-score* using the quantile function of a standard Gaussian distribution:

$$\mathrm{Z_{obs}} = \Phi^{-1}(1 - \mathrm{p_{obs}}). \tag{4.16}$$

Given α, the Neyman-Pearson construction provides a method to compare the performance of different tests. This consists in introducing a type-II error, or the *rate of false negatives*

$$\beta = P(t < t_\alpha | H_1) = \int_{-\infty}^{t_\alpha} f(t|H_1)dt, \tag{4.17}$$

where $f(t|H_1)$ is the distribution of the test statistic under the alternative hypothesis. Then the *power* of the test is defined as

$$\text{power} = P(t \geq t_\alpha | H_1) = 1 - \beta. \tag{4.18}$$

The best test is usually chosen as the one with the highest power for a given α, i.e., the one with the smallest rate of false negatives at fixed rate of false positives.

In this framework, it is not generally possible to guarantee, especially in finite-sample (non-asymptotic) settings, that the absence of detected discrepancies implies perfect agreement between two distributions. While certain tests may achieve consistency in the asymptotic limit—meaning they can detect any discrepancy with sufficient sample size—this property does not extend to practical scenarios with finite data.

To compute the quantities in eqs. (4.14) to (4.18), the distributions of the test statistic under both the null and alternative hypotheses, $f(t|H_0)$ and $f(t|H_1)$, must be known or estimated. In some cases, analytical approximations for finite sample sizes or asymptotic behavior can be used, as shown in Refs.[160, 161, 164] for the KS test, and in Refs.[71, 165] for likelihood ratio-based tests. In general, however, non-parametric testing often relies on empirical estimates.

We identify two common scenarios:

- The first case occurs when only two finite samples, $\mathcal{X}$ and $\mathcal{Y}$, representing two populations, are available. This scenario is common in the statistical literature on two-sample testing. To estimate $f(t|H_0)$ and $f(t|H_1)$, resampling methods, also known as randomization tests, such as permutation and bootstrap tests, are often used [166–168].

- In the second case, one or both distributions, while not known analytically, can still be sampled from. In other words, we have access to a *generator* for at least one of the distributions. This scenario commonly arises when evaluating the fidelity of a generator against a *reference* dataset, or when testing the compatibility of a dataset with a theoretical model represented by a Monte Carlo generator. Another common scenario involves comparing two data-generating methods, such as a highly accurate but computationally expensive Monte Carlo method against a faster but less accurate neural network-based model.

In practical applications, a generator might be available but limited by resource constraints,

preventing the generation of sufficient data to estimate the distribution of the test statistic accurately. In such cases, a hybrid approach can be used: generating as much data as possible and then applying resampling methods to estimate the distribution of the test statistic.

In the standard two-sample testing framework, as summarized in this Section, the two distributions in eqs. (4.11) and (4.12) are treated symmetrically, with no preference for p or q. However, this methodology can also be adapted for goodness-of-fit testing by designating one distribution as the *reference* one. This approach is useful when one distribution is considered the true or best approximation and the other is tested against it. For instance, in the example above, the fast but less accurate generative model would represent the alternative, while the slower but more accurate physics-based Monte Carlo generator the reference. Similarly, a dataset, treated as the alternative distribution, can be tested against a generator based on a well-verified theoretical model, taken as reference.

In this context, the hypotheses can be rephrased slightly. Under the null hypothesis H_0, both samples $\mathcal{X}$ and $\mathcal{Y}$ are drawn from the reference distribution p. Under the alternative hypothesis H_1, sample $\mathcal{X}$ is drawn from the reference p, while sample $\mathcal{Y}$ is drawn from an alternative distribution, $q \neq p$. The test statistic is then computed as in eq. (4.13), but the distribution $f(t|H_0)$ is estimated by testing the reference distribution against itself. This is done through repeated evaluations of the test statistic on pairs $(\mathcal{X}^{(i)}, \mathcal{X}^{(j)})$ of samples drawn from p. If p can be sampled, these samples are independent; otherwise, when only a single instance of data from the reference is available, resampling methods are used. Moreover, in this case the null hypothesis in eq. (4.11) becomes asymmetric between p and q, as we are specifically testing whether p, the distribution of the random variable x, also applies to y. The choice of which distribution to treat as the reference is part of the test design and can depend on specific biases or computational considerations.

In the following Chapters, we proceed by selecting a “true model" as the reference hypothesis. This reference could be either a generator or a finite dataset. We then introduce a precise methodology to compare the reference with alternative hypotheses, generated by systematically deforming the reference to simulate potential imperfections in a generative model. In our toy examples, where the PDF is known, we directly sample from both the reference and the alternative distributions. For more realistic data, such as those from the HEP literature, we rely on resampling methods. .

Chapter 5

Evaluating Two-Sample Tests for Validating Generators in Precision Sciences

In the previous Chapter, we explained why the High Energy Physics (HEP) community is focusing its attention on Machine Learning (ML) based generators to produce synthetic data. The significant speed-up offered by these approaches when combined with tradition Monte Carlo generators, is crucial in light of the advent of the High Luminosity Large Hadron Collider (HL-LHC) program, which will result in a huge volume of data. Within this context, we highlighted the need of a robust statistical validation framework to assess the reliability of ML-generated samples and to ensure their consistency with the underlying physical processes. A standard statistical tool for this purpose is the so-called two-sample test.

In this Chapter, we propose a robust methodology for evaluating two-sample tests, focusing on non-parametric tests based on univariate integral probability measures (IPMs). Our approach builds on two well-established methods for comparing one-dimensional, in short 1D, distributions, both based on the empirical Cumulative Distribution Function (eCDF): the Kolmogorov-Smirnov (KS) test [160, 161] and the Wasserstein (W) distance [162, 163]. We extend these to higher dimensions using three key test statistics: the sliced-Wasserstein (SW) distance, the mean of the KS statistics computed on each 1D marginal distribution, and a sliced version of the KS statistic. These tests are computationally efficient, as they can be evaluated in parallel, allowing for a fast and reliable estimate of their distribution under the null hypothesis.

For comparison with existing methods, we include in our analysis the unbiased Fréchet Gaussian Distance (FGD)[169, 170] and the unbiased quadratic Maximum Mean Discrepancy (MMD)[171, 172], computed using a quartic polynomial kernel. Additionally, when the underlying densities are known, we use the LLR test as a baseline to assess the effectiveness of the other tests.

We evaluate the proposed tests on a diverse set of scenarios. First, we use toy distributions with known probability density functions (PDFs), specifically correlated Gaussians and mixtures of

Gaussians in 5, 20, and 100 dimensions. These controlled benchmarks allow us to fully characterize the problem and compare the performance of the proposed non-parametric tests against the exact log-likelihood ratio test. Next, we apply our methodology to a particle physics dataset of gluon jets from the JetNet dataset [173], which includes both jet- and particle-level features. This real-world dataset enables us to test the robustness and versatility of our approach in a practical case where the underlying PDF is unknown and the data is finite.

For each dataset, we fix a reference distribution and we then introduce well-defined deformations, parameterized by a single constant, ϵ, to generate alternative hypotheses. We estimate the distribution of each test statistic under the null hypothesis by sampling, or resampling in the case of finite data, only from the reference distribution. Finally, we assess the sensitivity of the each test by testing the alternative hypothesis versus the reference one, and determining the minimum value of ϵ that leads to the rejection of the null hypothesis at a given Confidence Level (CL).[1]. See Figure 5.1 for a schematic representation. All results are accompanied by uncertainty estimates.

In this study, we examine a carefully selected basis of transformations that induce qualitatively different deformations in both the shape and correlation structure of the reference distributions. Since the magnitude of these deformations is not predetermined, this approach allows us to assess the strengths and weaknesses of each test in a fair manner. This approach ensures a transparent and controlled analysis while minimizing the introduction of model-dependent biases.

The Chapter is organized as follows. In Section 5.1, we describe the various test statistics considered in this study. Section 5.2 details our methodology for evaluating two-sample tests, while Section 5.3 describes the distributions, datasets, and deformations used in the analysis. In Section 5.4, we present the numerical results along with sample figures and tables, and in Section 5.5, we offer concluding remarks and discuss future directions. Appendix 5.A and 5.B provide additional material, including a full list of figures and tables.

Code, figures, and tables are available on GitHub [174–176] and full results are available on Zenodo [177, 178].

5.1 Test statistics

In this Section, we introduce the different test statistics considered in this work. With the exception of the FGD, all the test statistics are derived from or based on IPMs. An IPM can generally be expressed as:

$$d_{\mathcal{F}}(p, q) = \sup_{f \in \mathcal{F}} \left(\mathbb{E}_{x \sim p}[f(x)] - \mathbb{E}_{y \sim q}[f(y)] \right) , \tag{5.1}$$

where $\mathcal{F}$ is a class of real-valued measurable functions, $f : \mathcal{S} \to \mathbb{R}$.

One of the primary goals of this work is to compare the performance of these tests. To achieve this, we introduce several benchmark scenarios in Section 5.3, where the data-generating densities are fully known. In Section 5.1.5, we also introduce an exact LLR test, which is the most powerful

[1] This can be phrased as computing the upper value of ϵ excluded at a given CL.

Test statistic	Section
$t_{\mathrm{SW}} = \frac{1}{K} \sum_{\theta \in \Omega_K} \left(\frac{1}{n} \sum_{i=1}^{n} \mid \underline{x}_i^{\theta} - \underline{x}_i^{\prime\theta} \mid \right)$	5.1.1
$t_{\overline{\mathrm{KS}}} = \frac{1}{d} \sum_{I=1}^{d} \sqrt{\frac{nm}{n+m}} \sup_u \mid F_n^I(u) - G_m^I(u) \mid$	5.1.2
$t_{\mathrm{SKS}} = \frac{1}{K} \sum_{\theta \in \Omega_K} \sqrt{\frac{nm}{n+m}} \sup_u \mid F_n^{\theta}(u) - G_m^{\theta}(u) \mid$	5.1.2
$t_{\mathrm{MMD}} = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i}^{n} k(x_i, x_j) + \frac{1}{m(m-1)} \sum_{i=1}^{m} \sum_{j \neq i}^{m} k(y_i, y_j) - \frac{2}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} k(x_i, y_j)$, with: $k(x, x') = \left(\frac{1}{d} x^T x' + 1 \right)^4$	5.1.3
$t_{\mathrm{FGD}} = \lim_{n,m \to \infty} \sum_{I=1}^{d} (\mu_{1,n}^I - \mu_{2,m}^I)^2 + \mathrm{tr} \left(\Sigma_{1,n} + \Sigma_{2,m} - 2\sqrt{\Sigma_{1,n} \Sigma_{2,m}} \right)$	5.1.4
$t_{\mathrm{LLR}} = -2 \log \frac{\mathcal{L}_{H_0}}{\mathcal{L}_{H_1}}$	5.1.5

Table 5.1: Summary of the test statistics used in this work.

test for simple hypotheses, according to the Neyman-Pearson lemma [152]. This serves as a baseline for the absolute performance against which other tests can be compared. A summary of all the test statistics introduced in the following is given in Table 5.1.

5.1.1 Sliced Wasserstein distance

The SW distance [179, 180] is a computationally efficient variant of the W distance, derived from optimal transport theory [162, 181]. It involves averaging 1D projections of the W distance over all directions on the unit d-dimensional sphere, and it is a natural choice for two-sample testing in high-dimensional settings.

In this work, we focus on the 1-Wasserstein distance, commonly referred to as the *earth mover's distance*. In one dimension, it is defined as:

$$W_{n,m} = \int_{\mathbb{R}} |F_n(u) - G_m(u)| du, \tag{5.2}$$

where F_n and G_m are the eCDFs. This distance measures the minimal cost of transforming one distribution into another and depends linearly on the Euclidean distance between data points. It

can also be interpreted as an IPM, where the function class $\mathcal{F}$ is the space of 1-Lipschitz functions [163]. In the case of two 1D samples with an equal number of data points $m = n$, the quantity in eq. (5.2) can be simply computed as

$$W_n = \frac{1}{n} \sum_{i=1}^{n} | \, \underline{x}_i - \underline{x}'_i \, | \,, \tag{5.3}$$

where underlined variables represent elements in the set obtained by permuting the original sample with a permutation $\mathcal{P}$ that sorts the points:

$$\{\underline{x}\} = \mathcal{P}(\{x\} \mid \underline{x}_1 \leq \ldots \leq \underline{x}_n) \,. \tag{5.4}$$

The SW test statistic takes the following form:

$$t_{\text{SW}} = \frac{1}{K} \sum_{\theta \in \Omega_K} W_n^\theta = \frac{1}{K} \sum_{\theta \in \Omega_K} \left(\frac{1}{n} \sum_{i=1}^{n} | \, \underline{x}_i^\theta - \underline{x}_i'^\theta \, | \right) . \tag{5.5}$$

where Ω_K is a set of K directions selected uniformly at random on the unit sphere $\Omega = \{\theta \in \mathbb{R}^d \setminus ||\theta|| = 1\}$, and $\{\underline{x}_i^\theta\}_{i=1}^n = \{\theta^T x_i\}_{i=1}^n$ are the sorted data points projected on the direction θ. It is important to note that, in the asymptotic limit $m, n \to \infty$ with $m/n \to c \neq 0, \infty$, the distribution of the test statistic in eq. (5.5) under the null hypothesis will depend on the underlying data distribution. This means that the asymptotic behavior of the test statistic is not distribution-free [182].

5.1.2 Kolmogorov-Smirnov inspired test statistics

The KS test is a widely used non-parametric method for both goodness-of-fit and two-sample testing [160, 161]. It measures the largest absolute difference between the two eCDFs of the samples. The KS test can be viewed as an IPM, where the function class $\mathcal{F}$ consists of indicator functions $\mathbb{1}_{(-\infty,t]}$ for all $t \in \mathbb{R}$.

The KS test statistic is defined as

$$t_{\text{KS}} = \sqrt{\frac{nm}{n+m}} \sup_u | \, F_n(u) - G_m(u) \, | \,. \tag{5.6}$$

where $F_n(u)$ and $G_m(u)$ are the eCDFs of the two samples. The prefactor ensures that, under the null hypothesis (i.e., the two samples are drawn from the same distribution) and as $m, n \to \infty$ with $m/n \to c \neq 0, \infty$, the test statistic follows the Kolmogorov distribution. The Kolmogorov distribution has the following CDF:

$$F_K(x) = 1 - 2 \sum_{k=1}^{\infty} (-1)^{k-1} e^{-2k^2 x^2} \,, \tag{5.7}$$

and PDF

$$f_K(x) = \frac{d}{dx} F_K(x) = 8x \sum_{k=1}^{\infty} (-1)^{k-1} k^2 e^{-2k^2x^2} \,. \tag{5.8}$$

While the KS test is widely used for 1D data, its application in higher dimensions is limited due to the curse of dimensionality [183, 184]. To address this, we consider two efficient multivariate extensions: the mean KS test and the sliced KS test.

Mean KS The mean KS test extends the KS test to higher dimensions by averaging the KS statistics computed along each dimension of the data. This test was originally introduced in Ref. [185], using the median, and subsequently applied in Ref. [186]. Its variation based on the mean was later introduced in Ref. [187]. The test statistic is defined by the average

$$t_{\overline{\text{KS}}} = \frac{1}{d} \sum_{I=1}^{d} t_{\text{KS}}^I = \frac{1}{d} \sum_{I=1}^{d} \sqrt{\frac{nm}{n+m}} \sup_u |F_n^I(u) - G_m^I(u)|, \tag{5.9}$$

where $F_n^I(u)$ and $G_m^I(u)$ are the eCDFs of the 1D marginals along the I-th dimension. This approach makes the KS test computationally feasible in multivariate settings. Since the mean KS test is uniquely defined by the 1D marginals, it is not expected to be directly sensitive to correlations between dimensions. To improve the sensitivity to such correlations, we introduce the sliced KS test.

Sliced KS Similar to the SW distance introduced in eq. (5.5), the sliced KS (SKS) test extends the KS test to higher dimensions by projecting the original d-dimensional data onto 1D subspaces. These projections are taken along K random directions sampled from the unit sphere. For each direction θ, the KS test statistic is computed as:

$$t_{\text{KS}}^\theta = \sqrt{\frac{nm}{n+m}} \sup_u \mid F_n^\theta(u) - G_m^\theta(u) \mid \,. \tag{5.10}$$

where $F_n^\theta(u)$ and $G_m^\theta(u)$ are the eCDFs of the projected samples along direction θ. The SKS test statistic is then defined as the average of the KS statistics across the K random directions:

$$t_{\text{SKS}} = \frac{1}{K} \sum_{\theta \in \Omega_K} t_{\text{KS}}^\theta \,. \tag{5.11}$$

This approach leverages 1D KS tests over multiple projections, making it computationally feasible for higher-dimensional data and potentially sensitive to correlations between dimensions. To the best of our knowledge, the SKS test has not been previously introduced in the literature.

5.1.3 Maximum Mean Discrepancy

MMD is a statistical measure of the distance between two probability distributions, introduced in Ref.s [171, 172]. It is an example of an IPM, where the function class $\mathcal{F}$ is the unit ball in a reproducing kernel Hilbert space (RKHS).

An unbiased empirical estimate of MMD, following Ref. [172], is given by

$$\begin{aligned} t_{\mathrm{MMD}} =& \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i}^{n} k(x_i, x_j) + \frac{1}{m(m-1)} \sum_{i=1}^{m} \sum_{j \neq i}^{m} k(y_i, y_j) \\ & - \frac{2}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} k(x_i, y_j) \,, \end{aligned} \tag{5.12}$$

where $k(x, x')$ is the kernel function defining the RKHS.

In Ref. [188], a fourth-order polynomial kernel was used:

$$k(x, x') = \left(\frac{1}{d} x^T x' + 1 \right)^4 . \tag{5.13}$$

This kernel is not *characteristic*, meaning that $k(x, x')$ is not a true metric on the space of probability measures. Specifically, for this kernel, the condition $p = q$ is sufficient for $t_{\mathrm{MMD}}(p, q) = 0$, but it is not necessary. For example, this kernel cannot distinguish between distributions that differ beyond their fourth moment (see also example 3 in Ref. [189]). Because the polynomial kernel is not characteristic, this instance of MMD is a *pseudo-metric*. In contrast, characteristic kernels, such as Gaussian and Laplacian kernels, are capable of fully distinguishing between different distributions. However, these kernels require tuning hyperparameters, like the kernel bandwidth, based on the data. Following Ref. [188], we will use the fourth-order polynomial kernel described in eq. (5.13), as it has the advantage of not requiring hyperparameter tuning. A comparison of its performance relative to characteristic kernels could be the subject of future work. Finally, computing the MMD between two datasets of size n has a computational cost of $\mathcal{O}(n^2)$ because it requires storing the full kernel matrix K, where each element is given by $K_{ij} = k(x_i, x_j)$. This makes MMD computationally expensive, especially in large-scale scenarios or when the test needs to be evaluated multiple times.

5.1.4 Fréchet Gaussian Distance

The FGD is a pseudo-metric, specifically the Fréchet distance or 2-Wasserstein distance, between two multivariate Gaussian distributions. These distributions are characterized by means μ_1, $\mu_2 \in \mathbb{R}^d$ and covariance matrices $\Sigma_1, \Sigma_2 \in \mathbb{R}^{d \times d}$. The FGD between two samples of sizes n and m is given by:

$$\mathrm{FGD}_{n,m} = \sum_{I=1}^{d} (\mu_{1,n}^I - \mu_{2,m}^I)^2 + \mathrm{tr} \left(\Sigma_{1,n} + \Sigma_{2,m} - 2\sqrt{\Sigma_{1,n} \Sigma_{2,m}} \right) , \tag{5.14}$$

where $\mu_{1,n}^I$ and $\mu_{2,m}^I$ represent the I-th components of the sample means $\mu_{1,n}$ and $\mu_{2,m}$, respectively.

As noted in Ref.[188], the FGD is biased when computed on finite samples [169]. To mitigate this bias, an unbiased asymptotic extrapolation can be introduced, as proposed in Ref.[170]. This asymptotic value, denoted as

$$t_{\mathrm{FGD}} \coloneqq \mathrm{FGD}_\infty = \lim_{n,m\to\infty} \mathrm{FGD}_{m,n}\,, \tag{5.15}$$

is estimated by fitting a linear model to FGD values computed at different finite sample sizes. For simplicity, in the following we refer to FGD_∞ simply as FGD.

5.1.5 Likelihood-ratio

Detecting deviations from a reference model can be framed as a goodness-of-fit test between two competing statistical models. For simple hypotheses, the Neyman–Pearson lemma shows that the most powerful test in this scenario is the likelihood-ratio test [152]. In this work, we introduce the likelihood-ratio test using a slightly non-standard approach, following the two-sample testing framework outlined in Section **??**.

The likelihood function for the datasets $\mathcal{X}$ and $\mathcal{Y}$ under the null hypothesis (where both samples follow the reference distribution p) is written as

$$\mathcal{L}_{H_0} = \prod_{x\in\mathcal{X}} p(x) \prod_{y\in\mathcal{Y}} p(y). \tag{5.16}$$

Under the alternative hypothesis (where the sample $\mathcal{Y}$ follows a different distribution q), the likelihood is:

$$\mathcal{L}_{H_1} = \prod_{x\in\mathcal{X}} p(x) \prod_{y\in\mathcal{Y}} q(y). \tag{5.17}$$

The ratio of the likelihoods under the null and alternative hypotheses is then given by

$$\Lambda = \frac{\mathcal{L}_{H_0}}{\mathcal{L}_{H_1}} = \prod_{y\in\mathcal{Y}} \frac{p(y)}{q(y)}, \tag{5.18}$$

The test statistic for the LLR test is then defined as:

$$t_{\mathrm{LLR}} = -2\log\Lambda. \tag{5.19}$$

Notice that, from eq. (5.18), the value of the test statistic on the observed data does not depend on sample $\mathcal{X}$, since p was selected as the reference distribution. As discussed in Section **??**, this effectively transforms the problem into a goodness-of-fit test between the reference distribution p and the sample $\mathcal{Y}$ in a two-sample testing framework. To estimate the distribution of the test statistic under the null hypothesis, we evaluate eq. (5.19) on several reference-distributed samples $\{\mathcal{X}^{(i)}\}$.

5.2 Methodology

This work builds on recent contributions, such as Ref. [188], by providing an informative and fair comparison between evaluation metrics. Specifically, we establish a well-defined and reproducible methodology aimed at comparing test statistics close to a meaningful decision boundary.

We begin by using the true PDF of the data, denoted as p, as the reference distribution, which defines the null hypothesis. We then introduce various deformations of p to generate alternative hypotheses with distinct characteristics. These deformations, described in Section 5.3.3, are controlled by a unique scalar parameter ϵ. The PDF of each deformed distribution is referred to as q_ϵ.

Following the procedure outlined in Section **??**, to estimate the test statistic t under the null hypothesis, denoted by t_0, we use a Monte Carlo approach based on repeated sampling. Multiple random pairs of samples are drawn from p, and t_0 is computed for each pair. This generates an empirical distribution with eCDF $F(t_0)$ and ePDF $f(t_0)$, which can be used to establish a threshold t_0^α, corresponding to a preselected false positive rate.

We define two confidence levels corresponding to $1 - \alpha = 0.95, 0.99$. For each null distribution $f(t_0)$, we identify a threshold t_0^α such that:

$$\alpha = \int_{t_0^\alpha}^{\infty} dF(t_0) = \int_{t_0^\alpha}^{\infty} f(t_0) dt_0. \tag{5.20}$$

In the Monte Carlo approach, this integral is empirically estimated by repeated sampling:

$$\alpha = 1 - F(t_0^\alpha) = \frac{\text{Number of simulated } t_0 \text{ values } \geq t_0^\alpha}{\text{Total number of simulations}}. \tag{5.21}$$

In this setup and for all tests considered here, with the exception of the LLR, the distribution of the test statistic under the null hypothesis depends only on the reference distribution p, and not on the alternative q_ϵ. Therefore, the null distribution needs to be estimated only once for each reference p and sample size.

Once the threshold values t_0^α are determined, two-sample tests are conducted between samples generated by the true model p and samples from the deformed models q_ϵ. The basic idea is to perform the tests for different values of ϵ until the test statistic under the alternative hypothesis, $t(\epsilon)$, reaches the threshold t_0^α, i.e. $t(\epsilon) = t_0^\alpha$. The value of ϵ for which the null hypothesis is rejected at the CL α is denoted by ϵ_α. This is found by solving the following optimization problem:

$$\epsilon_\alpha = \arg\min_\epsilon |t(\epsilon) - t_0^\alpha|, \tag{5.22}$$

using a simple bisection method over a specified range of ϵ. This iterative method repeatedly halves the range until a sufficiently accurate value of ϵ_α is found.

In practice, to ensure a robust statistical interpretation, each test under the alternative hypothesis is repeated 100 times. For each test, we calculate the average test statistic, $\mu_{t(\epsilon)}$, and its standard

deviation, $\sigma_{t(\epsilon)}$. We then evaluate $\mu_{t(\epsilon)}$, $\mu_{t(\epsilon)} - \sigma_{t(\epsilon)}$, and $\mu_{t(\epsilon)} + \sigma_{t(\epsilon)}$. For each of these quantities, if it is above the threshold, the middle of the ϵ range is shifted left; if it is below, it is shifted right. The range is halved at each step, and the procedure is repeated until both the range of ϵ and the range of the absolute difference between t_0^α and $\mu_{t(\epsilon)}$, $\mu_{t(\epsilon)} - \sigma_{t(\epsilon)}$, and $\mu_{t(\epsilon)} + \sigma_{t(\epsilon)}$ fall within a predefined tolerance. For convergence, we fixed a tolerance of 10^{-2}. In cases where the test statistic is particularly noisy (affecting convexity of the optimization problem), we increase the tolerance to 5×10^{-2}, which does not affect the overall results. The value of ϵ for which $\mu_{t(\epsilon)}$ converges to t_0^α is identified with ϵ_α. The values of ϵ for which $\mu_{t(\epsilon)} - \sigma_{t(\epsilon)}$ and $\mu_{t(\epsilon)} + \sigma_{t(\epsilon)}$ converge to t_0^α are used to establish the one standard deviation uncertainty bounds on ϵ_α, denoted as $\epsilon_{\alpha-\text{low}}$ and $\epsilon_{\alpha-\text{up}}$.

The LLR test requires a different approach since it is a parametric test that explicitly depends on both the reference distribution p and the alternative distribution q_ϵ. As a result, also the test statistic under the null hypothesis, denoted by $t_0(\epsilon)$, now varies with ϵ, and so does its ePDF $f(t_0(\epsilon))$. This means that for each value of ϵ, we must compute the null distribution and re-evaluate the threshold for the chosen significance level α. The same procedure used for the non-parametric tests is followed here, but with the additional step of generating the null distribution for each value of ϵ. The optimization problem for the LLR then becomes

$$\epsilon_\alpha = \arg\min_\epsilon |t(\epsilon) - t_0^\alpha(\epsilon)|, \tag{5.23}$$

where $t(\epsilon)$ is the test statistic under the alternative hypothesis and $t_0^\alpha(\epsilon)$ is the threshold value for the same significance level α and value of ϵ. This optimization is more computationally expensive than for non-parametric tests since the null distribution must be computed for each value of ϵ. Nevertheless, it is carried out using the same bisection method. In cases where the LLR test is highly sensitive to the value of ϵ, such as in high-dimensional scenarios or when the sample size is large, we apply a slightly larger tolerance than 10^{-2} on the function value during optimization.

It is important to note that the LLR test can only be applied when the PDFs of both the true and alternative distributions are known. While this holds true for our reference distributions, when it is known analytically, it is not always the case for deformed distributions. Specifically, for random deformations like $\Sigma_{i\neq j}$, $\mathcal{N}$, and $\mathcal{U}$, the analytical forms of the PDFs q_ϵ cannot be obtained. This situation mimics what often occurs in real-world applications, where samples can be drawn from a distribution even though the exact PDF is unknown.[2] For these cases, we only perform and compare the non-parametric tests.

Our methodology can be directly applied to scenarios where the full details of the problem are known (see Section 5.3.1). However, when working with the finite set of simulated gluon jets from the JetNet dataset [173], we must make some adjustments. In this case, we do not have access to the underlying PDFs that generate the data, even for the reference model. Nevertheless, we can still apply the same deformations to the numerical samples from the dataset, as we do with the toy models. We therefore use a bootstrap-based resampling method, where multiple instances are

[2] A typical example is exactly the one in which the output of the true PDF is “smeared" by some probability distribution determined, for instance, by a detector response.

samples with replacement from the full dataset, to estimate the PDF of the test statistic $f(t_0)$. This allows us to test the alternative hypothesis as described earlier. Obviously, since we do not know the analytical form of the PDFs for the physics datasets, we cannot perform the LLR test in these cases.

We applied this methodology across different dimensionalities (number of features) and sample sizes for each model (see Section 5.3). Each metric was tested under various conditions, allowing for a fair comparison of the performance of each test statistic in relation to the specific deformations and the complexity of the problem.

5.3 Datasets and analysis setup

This Section describes the distributions, datasets, and deformations used in our analysis, along with details about the dimensionalities and sample sizes considered.

5.3.1 Toy models

We define the PDFs of the reference distributions analytically, using multivariate Gaussian distributions and mixtures of multivariate Gaussian distributions. Having direct access to these PDFs offers two key advantages: first, it allows us to apply the LLR test for most of the considered deformations, providing optimal benchmarks according to the Neyman-Pearson lemma; second, it enables us to generate as many samples as needed with minimal computational effort. This ensures that we can always build the distribution of the test statistic under the null hypothesis, $f(t_0)$, with the statistics required by the hypothesis we want to test. As a result, the tail regions corresponding to $\alpha = 0.05$ and $\alpha = 0.01$ are well populated, supporting a robust and fair comparison when evaluating the alternative hypothesis.

Mixture of Gaussians (MoG): We consider mixtures of q multivariate Gaussians (q components) in d dimensions, referred to as MoG models, each with diagonal covariance matrices. This allows us to study cases in which the PDF has more than one local maxima manifesting themselves as multiple peaks in the marginals. The reference model is constructed as follows: for a given q and d, we randomly generate means from a uniform distribution in the range $[-5, 5]$ and standard deviations in the range $[0, 1]$. The Gaussian components are then mixed according to a categorical distribution with randomly assigned probabilities P_i, where $i = 1, \ldots, q$. Each component has a different probability, and all dimensions of a given multivariate Gaussian share the same probability. For our analysis, we consider three MoG models: $q = 3$ components in $d = 5$ dimensions, $q = 5$ components in $d = 20$ dimensions, and $q = 10$ components in $d = 100$ dimensions. It is important to note that, although the covariance matrix for each Gaussian component is diagonal, the overall mixture model exhibits off-diagonal elements of order one in the correlation matrix.

Correlated Gaussians (CG): We also consider correlated d-dimensional multivariate Gaussian distributions, referred to as CG models, defined by their mean vector and covariance matrix. For a given dimensionality d, the reference model is constructed by randomly generating means from a uniform distribution in the range $[-5, 5]$. Full covariance matrix is taken from the correlation matrix of the MoG model with the same dimensionality, with order-one off-diagonal elements. In our analysis, we consider $d = 5, 20, 100$.

For both the MoG and CG models, we compute $f(t_0)$ for each metric by drawing 10^4 pairs of samples and performing two-sample tests between them, with varying sample sizes $n = m = (1, 2, 5, 10) \times 10^4$. The alternative hypothesis is tested using the procedure outlined in Section 5.2, with 100 pairs of samples drawn from the reference and deformed distributions for each PDF, sample size, and deformation.

5.3.2 JetNet Dataset

To explore a scenario where we lack access to the underlying PDFs, and to check our results on a dataset relevant for HEP, we also consider, in our experiments, a dataset of simulated gluon jets within the JetNet dataset. We consider two sets of data: one labeled as particle_features, which includes features of individual particles within each jet, and another labeled as jet_features, which contains the overall features of the jets.

Out of a dataset containing a total of $n_{\text{jet}} = 177252$ gluon jets, we use, in our analysis, the following dataset structures:

- Particle features dataset: this consists of tensors with shapes $[n_{\text{jet}}, n_{\text{part}}, 3]$, where n_{part} is the maximum number of particles per jet. The last dimension corresponds to three physical features: the transverse momentum p_T, the pseudorapidity η, and the azimuthal angle ϕ with respect to the jet direction;

- Jet features dataset: this consists of tensors with shapes $[n_{\text{jet}}, 3]$, where the last dimension contains the transverse momentum p_T, the pseudorapidity η, and the jet mass m_j.

These are the the typical kinematic variables describing jets and sub-jet partons in hadronic collisions. In this study, for the particle features dataset, we focus on the case with $n_{\text{part}} = 30$, and we discuss why we do not extend the analysis beyond this limit.

Since we do not know the PDFs generating the JetNet datasets, we cannot generate unlimited samples to build $f(t_0)$. To address this, we use a bootstrap method for resampling on the available data: we shuffle the dataset and split it in half, extracting sub-samples of size n from each half and performing two-sample tests between them.

This procedure is repeated until each metric is evaluated 10^3 times, ensuring that there are enough points available in the $\alpha < 0.01$ region, which allows for a reasonable estimate of the 99% CL threshold. To test the alternative hypothesis, we apply the same method, where one half of

the dataset is deformed, and the other half is kept as the reference. This procedure is repeated 10^2 times for each metric, sample size, and deformation.

All deformations are performed by first standardizing the dataset, which involves subtracting the mean and dividing by the standard deviation. After applying the deformation, we multiply by the original standard deviation and add back the original mean to ensure consistency in scale.

As in the toy examples, to thoroughly understand the behavior of the test statistics in different complexity scenarios, we selected sample sizes of $n = m = (1, 2, 5) \times 10^4$. Additionally, while testing the particle features dataset, we performed tests both on data with the original scale, where p_T and m_j can be quite large when measured in GeV, and on scaled data, with zero mean and unit variance. This approach allows us to test the robustness of the metrics to changes in the data scale and helps us determine the best approach for each metric.

5.3.3 Deformations

The deformed models are obtained by applying parametric transformations to the reference models. Specifically, given a random variable X distributed according to the reference PDF p, the deformation acts as a parametric function of ϵ, transforming X into a new random variable $Y = g(X; \epsilon)$, where Y is distributed according to the deformed PDF q_ϵ.

When the transformation function g is bijective, we can express q_ϵ analytically in terms of p through the change of variables formula:

$$q_\epsilon(Y) = p(g^{-1}(Y; \epsilon))|\det J_{g^{-1}}| = p(g^{-1}(Y; \epsilon))|\det J_g|^{-1}, \tag{5.24}$$

where $J_g = \partial g / \partial X$ is the Jacobian matrix of the transformation g. In these cases, the likelihood $q_\epsilon(Y)$ can be computed, and so can the LLR test statistic.

However, when g is not bijective or not invertible (e.g. random transformations, such as a smearing), the PDF q_ϵ cannot be expressed analytically, and the LLR cannot be computed. Nonetheless, we can still generate samples from q_ϵ by drawing samples from p and applying the transformation $g(X; \epsilon)$ to obtain the deformed samples.

When working with real datasets like JetNet, where only samples drawn from the reference distribution are available, the deformed points are obtained by applying the transformation g directly to the sample points. In such cases, neither p nor q_ϵ are known analytically, and the LLR test cannot be performed.

Below, we describe each deformation as a function of the transformation g and its action on the design matrix x_{iI} of a sample. We also highlight when the function is invertible, and when it is not.

1. **μ-deformation**
   A random vector $\delta\mu_I$, with each entry drawn from a uniform distribution in the interval $[-\epsilon, \epsilon]$, is added to each point from the reference model, modifying its mean from μ_I to $\mu_I + \delta\mu_I$. The

transformation g is bijective and can be written as:

$$\begin{aligned} y_{iI} &= g(x_{iI}; \delta\mu_I(\epsilon)) = x_{iI} + \delta\mu_I(\epsilon)\,, \\ x_{iI} &= g^{-1}(y_{iI}; \delta\mu_I(\epsilon)) = y_{iI} - \delta\mu_I(\epsilon)\,, \\ \delta\mu_I(\epsilon) &\sim \mathcal{U}_{[-\epsilon,\epsilon]}\,, \end{aligned} \tag{5.25}$$

with $\epsilon \geq 0$.

2. **$\boldsymbol{\Sigma_{II}}$-deformation**
   A random vector $\delta\sigma_I$, with each entry drawn from a uniform distribution in the interval $[1, 1+\epsilon]$, is used to scale the standard deviations of the model from σ_I to $\sigma_I \cdot \delta\sigma_I$, leaving the correlation matrix unchanged. The mean is kept fixed by subtracting the model mean before scaling and adding the mean back after scaling. In formulae:

$$\begin{aligned} y_{iI} &= g(x_{iI}; \delta\sigma_I) = \mu_I + \delta\sigma_I(x_{iI} - \mu_I)\,, \\ x_{iI} &= g^{-1}(y_{iI}, \delta\sigma_I) = \mu_I + \delta\sigma_I^{-1}(y_{iI} - \mu_I)\,, \\ \delta\sigma_I &\sim \mathcal{U}_{[1,1+\epsilon]}\,, \end{aligned} \tag{5.26}$$

   with $\epsilon \geq 0$. The transformation is bijective, so that the LLR can be computed.

3. **$\boldsymbol{\Sigma_{I\neq J}}$-deformation**
   A fraction ϵ of each component of the reference model is shuffled independently relative to the others. In formulae:

$$y_{iI} = g\left(x_{jI}; P_{ij}^I(\epsilon)\right) = \sum_j P_{ij}^I(\epsilon) x_{jI}\,, \tag{5.27}$$

   with $\epsilon \geq 0$. P_{ij}^I is a tensor describing I permutation matrices in the (i, j) space (the sum in the formula above does not run on the index I). The P_{ij}^I permutation matrix has all entries equal to zero but one entry equal to one for each row and column. A fraction $1 - \epsilon$ of the ones are on the diagonal, while a fraction ϵ are off the diagonal. The positions of the ones are random, and different for each I.

   Shuffling points independently for each feature has the effect of modifying the covariance matrix without changing the marginal distributions. As ϵ increases, the correlations between features decreases, leading to a diagonal covariance matrix when $\epsilon = 1$. For $\epsilon > 1$, this deformation is followed by the Σ_{II} deformation, which scales the standard deviations further. The inverse transformation and the PDF of the deformed distribution are not known, so that the LLR cannot be computed for this deformation.

4. **pow$^+$-deformation**
   Each component of each point is modified by raising its absolute value to the power $(1 + \epsilon)$, while keeping the sign fixed. This deformation alters the model beyond just its mean and

covariance matrix. The transformation is bijective, and can be written as

$$\begin{aligned} y_{iI} &= g(x_{iI};\epsilon) = \text{sign}(x_{iI})|x_{iI}|^{1+\epsilon}\,, \\ x_{iI} &= g^{-1}(y_{iI};\epsilon) = \text{sign}(y_{iI})|y_{iI}|^{1/(1+\epsilon)}\,, \end{aligned} \tag{5.28}$$

with $\epsilon \geq 0$.

5. **pow$^-$-deformation**
   Similar to the previous deformation, but with the power $(1-\epsilon)$. In formulae:

$$\begin{aligned} y_{iI} &= g(x_{iI};\epsilon) = \text{sign}(x_{iI})|x_{iI}|^{1-\epsilon}\,, \\ x_{iI} &= g^{-1}(y_{iI};\epsilon) = \text{sign}(y_{iI})|y_{iI}|^{1/(1-\epsilon)}\,, \end{aligned} \tag{5.29}$$

   with $\epsilon \geq 0$.

6. **$\mathcal{N}$-deformation**
   Each component of each point is modified by adding a random value δ_{iI}, drawn from a normal distribution $\mathcal{N}_{0,\epsilon}$ with zero mean and standard deviation ϵ. The transformation can be written as:

$$\begin{aligned} y_{iI} &= x_{iI} + \delta_{iI}(\epsilon)\,, \\ \delta_{iI}(\epsilon) &\sim \mathcal{N}_{0,\epsilon}\,, \end{aligned} \tag{5.30}$$

   with $\epsilon \geq 0$. The transformation is not invertible, and the PDF of the deformed distribution is not known, so that the LLR cannot be evaluated.

7. **$\mathcal{U}$-deformation**
   Each component of each point is modified by adding a random value δ_{iI}, drawn from a uniform distribution in the $[-\epsilon,\epsilon]$ interval. In formulae:

$$\begin{aligned} y_{iI} &= x_{iI} + \delta_{iI}(\epsilon)\,, \\ \delta_{iI}(\epsilon) &\sim \mathcal{U}_{[-\epsilon,\epsilon]}\,, \end{aligned} \tag{5.31}$$

   with $\epsilon \geq 0$. Like the $\mathcal{N}$ deformation, the transformation is not invertible, and the PDF of q_ϵ is not known.

5.4 Results

In this Section we present the results for all the models discussed above. All results have been obtained with an implementation of the metrics in TensorFlow2 available in the GitHub repository [174]. All figures and tables documenting the results are available in the GitHub repositories [175, 176] and listed and linked in Appendices 5.A and 6.C.

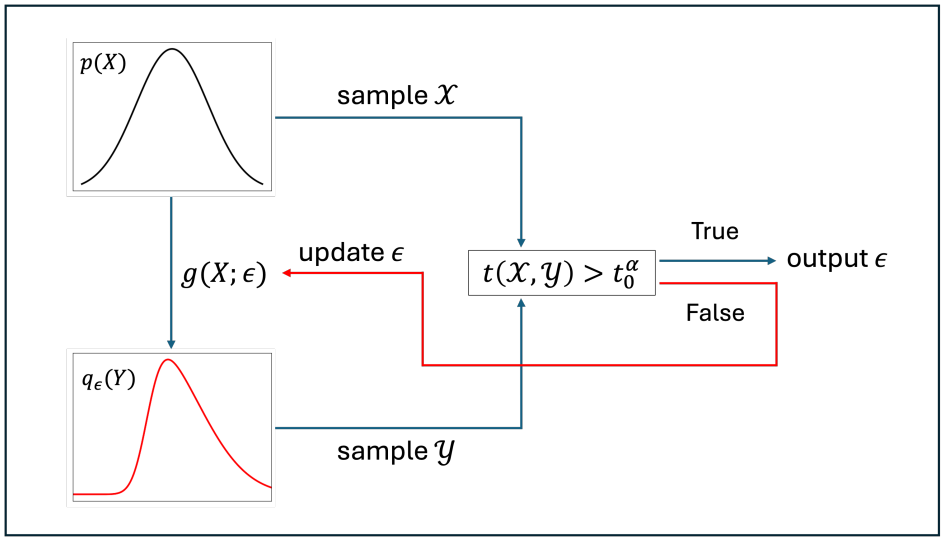


Figure 5.1: Schematic view of the methodology.

For all the considered distributions, the figures include corner plots showing the 1D and 2D marginal probability distributions for the reference and deformed distributions, color plots that give a pictorial representation of the correlation matrices for the reference and deformed distributions, and the distributions of the test statistics under the null hypothesis for all metrics but the LLR and for all sample sizes. The available tables report the results in terms of the values of ϵ for which the null hypothesis is rejected (in short, the upper bounds on ϵ) at the 95% and 99% Confidence Level (CL) for each metric, sample size, and deformation.

5.4.1 Toy models

Let us start discussing results for the two classes of toy distributions that we consider in this work, namely the MoG and CG models. Since the full list of results for the different dimensionalities $d = 5, 20, 100$, and sample sizes $n = m = (1, 2, 5, 10) \cdot 10^4$ is reported in the Appendices, we show here only a small subset of figures and tables, which helps understanding the methodology and the overall picture of the results that we discuss below.

Figures 5.2 and 5.3 show corner plots for the MoG and CG models with $d = 5$ and Σ_{ii}-deformation with $\epsilon = 0.5$. The plots show the 1D and 2D marginal probability distributions for the reference and deformed distributions. In the figure we clearly see the effect of the Σ_{ii}-deformation, which changes the vector of standard deviations, while leaving the vector of means and the correlation matrix unchanged. The plots show, in the 2D marginal distributions, contours for the highest density intervals corresponding to a probability of 68.27%, 95.45%, and 99.73%.[3] Similar plots are available for all the other deformations and dimensionalities considered, and are referenced in Appendix 5.A.

Figure 5.4 shows a color plot for the correlation matrix for the MoG model with $d = 20$ and $q = 5$

[3]These probabilities correspond to one, two, and three standard deviations for a univariate Gaussian distribution.

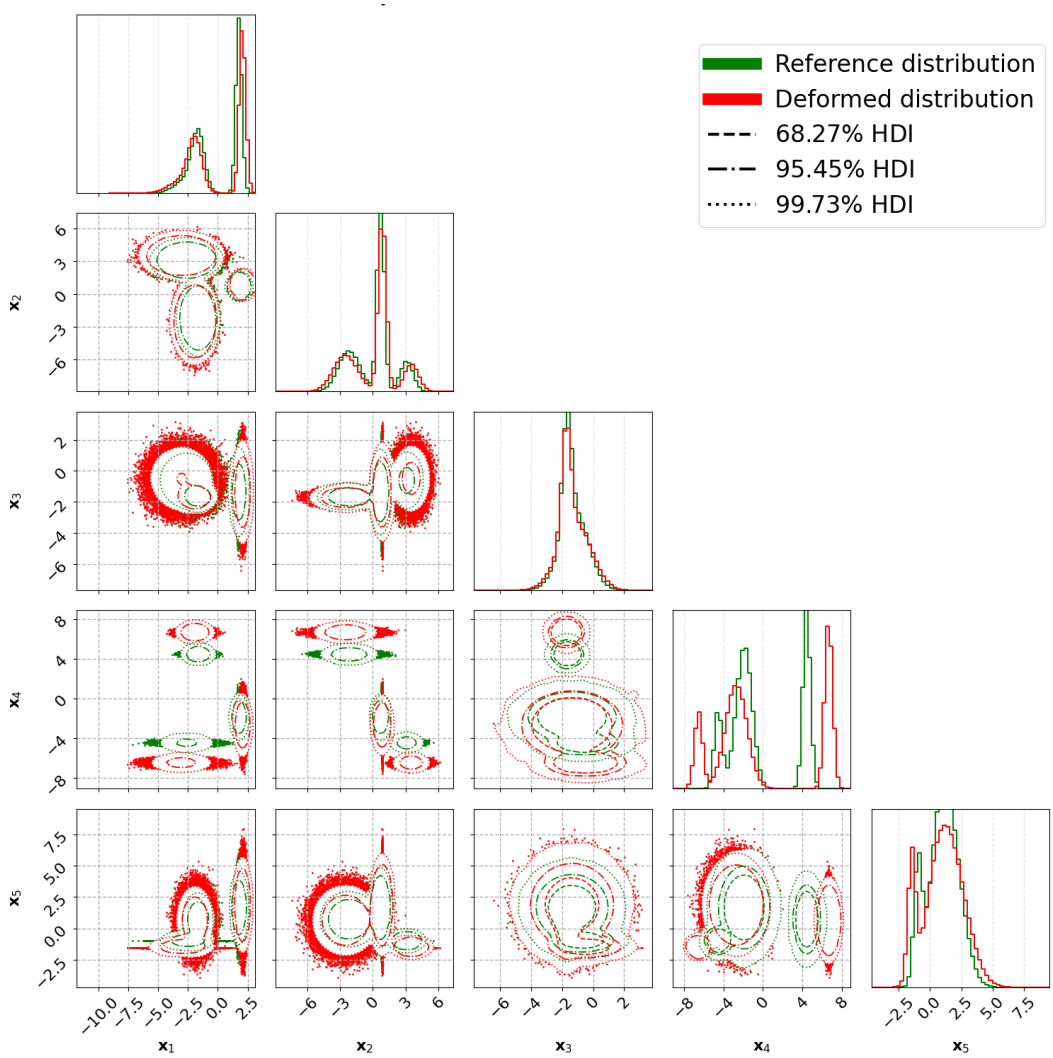


Figure 5.2: Corner plot showing the 1D and 2D marginal probability distributions for the reference and deformed distributions for the MoG model with $d = 5$ and $q = 3$, and Σ_{ii}-deformation with $\epsilon = 0.5$. The plots are made with 10^6 points per sample.

(left panel), and the corresponding $\Sigma_{i\neq j}$ deformation with $\epsilon = 0.5$ (right plot). The figure makes very intuitive the effect of the $\Sigma_{i\neq j}$ deformation: it decreases the correlation between the features until the covariance matrix becomes diagonal (for $\epsilon > 1$ the covariance matrix is kept diagonal and the standard deviation is modified according to the Σ_{ii}-deformation). Similar plots are available for all the other deformations and dimensionalities considered and are refereced in Appendix 5.A.

Figure 5.5 shows the empirical PDF and CDF of the various test statistics under the null hypothesis for the MoG model with $d = 20$ and $q = 5$, and $n = m = 5 \cdot 10^4$ samples. The histograms are built with 10^4 iterations for each test statistic. In blue, red, and green, we highlight the thresholds corresponding to 68%, 95%, and 99% CLs, respectively. The five panels represent the SW (upper left), $\overline{\text{KS}}$ (upper right), SKS (middle left), FGD (middle right), and MMD (lower) metrics, respectively. The plots corresponding to KS-inspired tests, namely $\overline{\text{KS}}$ and SKS, include a line showing,

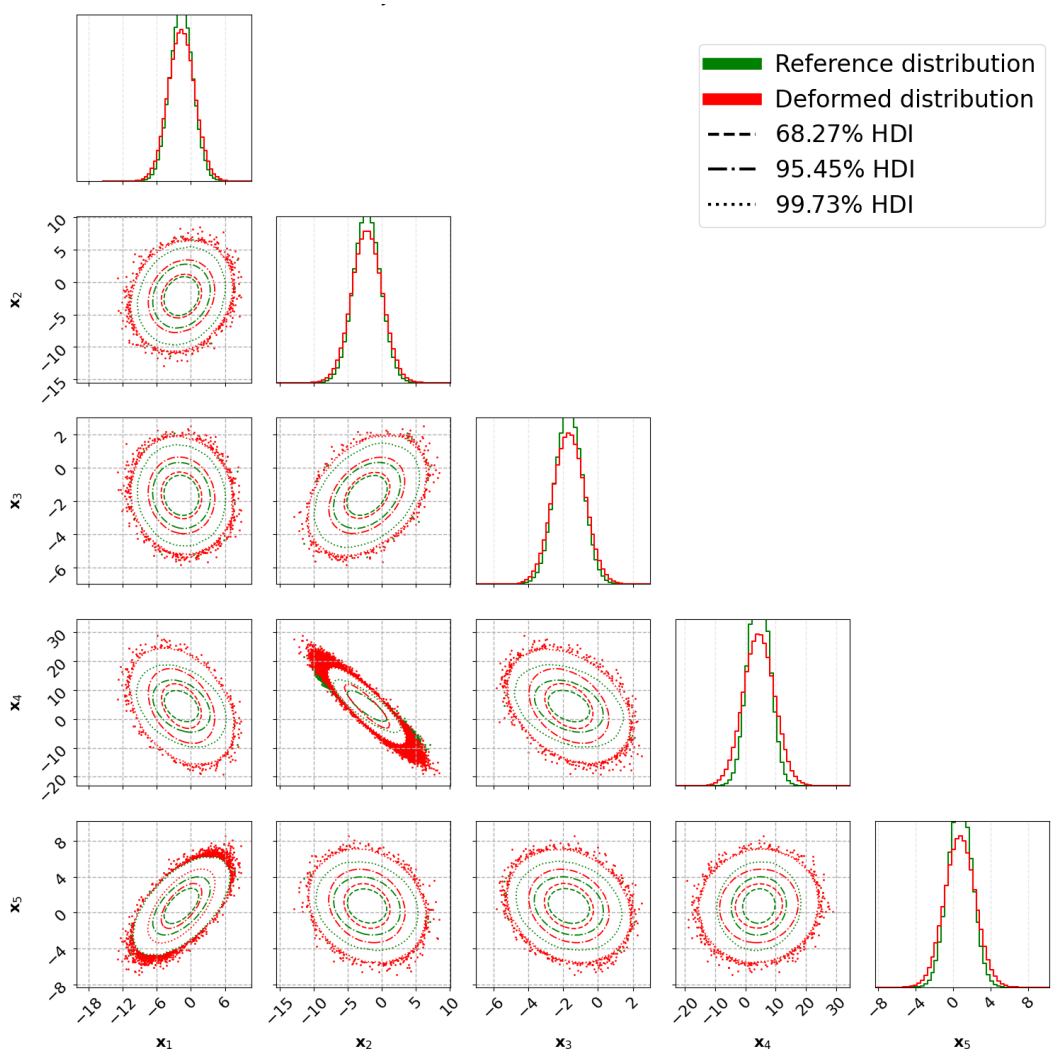


Figure 5.3: Corner plot showing the 1D and 2D marginal probability distributions for the reference and deformed distributions for the CG model with $d = 5$. The plots are made with 10^6 points per sample.

for comparison, the (parameter free) Kolmogorov distribution expected as asymptotic distribution for the univariate (original) KS test. Similar plots are available for all the other dimensionalities and sample sizes considered and are referenced in Appendix 5.A.

Finally, results for each model are summarized in tables and referenced for all dimensionalities and sample sizes considered in Appendix 6.C. As an instance, Table 5.2 shows the upper bound on ϵ at the 95% and 99% CL for the MoG model with $d = 20$, $q = 5$, and $n = m = 5 \cdot 10^4$ samples, while Table 5.3 shows the same for the CG model with $d = 20$ and $n = m = 5 \cdot 10^4$ samples. The tables report the results for all the metrics and all the deformations considered.

Inspecting results for MoG and CG in different dimensions and with different sample sizes, we can draw a few general conclusions:

- for the μ deformation, and for all dimensions and sample sizes, all metrics perform similarly,

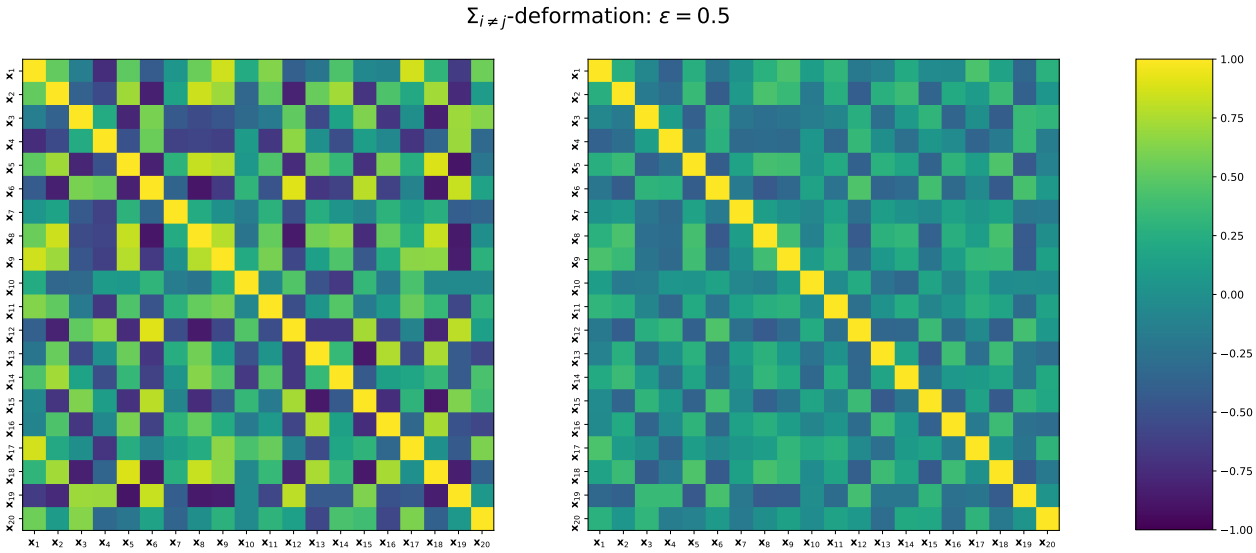


Figure 5.4: Color plot showing the correlation matrix for the reference (left) and deformed (right) distributions for the MoG model with $d = 20$ and $q = 5$, and $\Sigma_{i\neq j}$-deformation with $\epsilon = 0.5$. The figure is identical in the case of the CG model, since the same correlation matrix is used for both models. The plots are made with 10^6 points per sample.

within uncertainties, for the CG models, with a slightly better performance of $\overline{\text{KS}}$ for low dimensionalities, and of MMD for large dimensionalities; concerning the MoG models, the situation changes, with $\overline{\text{KS}}$ always performing better than the other metrics by around one order of magnitude; the LLR is always the most sensitive for the μ-deformation, with at least one order of magnitude better performance than the best of the other metrics;

- for the Σ_{ii} deformation, all metrics perform similarly, within uncertainties, for the CG models, with a slightly better performance of FGD and SW; concerning the MoG models, the situation changes, with $\overline{\text{KS}}$ always performing better than the other metrics by around one order of magnitude; the LLR outperforms the other metrics by at least one, and often two, orders of magnitude;

- for the $\Sigma_{i\neq j}$ deformation, FGD is the clear winner, with sensitivities better than other metrics by up to one order of magnitude across all dimensions and sample sizes, and both for the CG and the MoG models; notice that the $\overline{\text{KS}}$ is not sensitive, by construction, to this deformation;

- for the pow$^+$ and pow$^-$ deformations, $\overline{\text{KS}}$ is again the most sensitive for the MoG models, often by one order of magnitude, while MMD is the leader for the CG models, even thought the sensitivity is often only about one or two standard deviations better than that of other metrics; indeed, for the CG models, all metrics perform similarly, within uncertainties, for the pow$^+$ and pow$^-$ deformations; the LLR is again around one order of magnitude more sensitive than the best metric;

MoG model with d = 20, q = 5, and n = m = $5 \cdot 10^4$						
	μ-deformation			Σ_{ii}-deformation		
Statistic	$\epsilon_{95\%\text{CL}}$	$\epsilon_{99\%\text{CL}}$	t (s)	$\epsilon_{95\%\text{CL}}$	$\epsilon_{99\%\text{CL}}$	t (s)
t_{SW}	$0.04957^{+0.018}_{-0.02}$	$0.06694^{+0.017}_{-0.017}$	3023	$0.01679^{+0.005}_{-0.0063}$	$0.02315^{+0.0045}_{-0.005}$	3197
$t_{\overline{\text{KS}}}$	$\mathbf{0.00482^{+0.0013}_{-0.0018}}$	$\mathbf{0.00667^{+0.0011}_{-0.0013}}$	2966	$\mathbf{0.00175^{+0.00052}_{-0.00068}}$	$\mathbf{0.00248^{+0.00042}_{-0.00052}}$	3185
t_{SKS}	$0.03647^{+0.011}_{-0.014}$	$0.04821^{+0.011}_{-0.012}$	**2899**	$0.01329^{+0.003}_{-0.0043}$	$0.01759^{+0.0025}_{-0.003}$	**3022**
t_{FGD}	$0.05778^{+0.026}_{-0.027}$	$0.0787^{+0.023}_{-0.021}$	4047	$0.01945^{+0.0063}_{-0.0081}$	$0.02651^{+0.0053}_{-0.0056}$	4507
t_{MMD}	$0.04425^{+0.019}_{-0.018}$	$0.06215^{+0.017}_{-0.015}$	10204	$0.00923^{+0.0058}_{-0.0051}$	$0.01305^{+0.0053}_{-0.0044}$	11217
t_{LLR}	$0.00021^{+0.00013}_{-0.00014}$	$0.0003^{+0.00013}_{-0.00014}$	5911	$0.00007^{+0.00005}_{-0.00004}$	$0.0001^{+0.00005}_{-0.00004}$	6304
	$\Sigma_{i\neq j}$-deformation			pow_+-deformation		
Statistic	$\epsilon_{95\%\text{CL}}$	$\epsilon_{99\%\text{CL}}$	t (s)	$\epsilon_{95\%\text{CL}}$	$\epsilon_{99\%\text{CL}}$	t (s)
t_{SW}	$0.02162^{+0.0056}_{-0.008}$	$0.02935^{+0.0045}_{-0.0055}$	**3410**	$0.00581^{+0.0017}_{-0.0022}$	$0.00798^{+0.0015}_{-0.0017}$	**3157**
$t_{\overline{\text{KS}}}$	$1.00146^{+0.00074}_{-0.00031}$	$1.00238^{+0.00055}_{-0.00031}$	3967	$\mathbf{0.0004^{+0.00015}_{-0.00017}}$	$\mathbf{0.00059^{+0.00013}_{-0.00014}}$	3363
t_{SKS}	$0.02306^{+0.0071}_{-0.0088}$	$0.03079^{+0.0062}_{-0.0072}$	3553	$0.0043^{+0.0009}_{-0.0013}$	$0.00565^{+0.00074}_{-0.0009}$	3193
t_{FGD}	$\mathbf{0.00551^{+0.0015}_{-0.002}}$	$\mathbf{0.00748^{+0.0013}_{-0.0013}}$	6327	$0.00702^{+0.0021}_{-0.0028}$	$0.00965^{+0.0016}_{-0.0019}$	4870
t_{MMD}	$0.01723^{+0.008}_{-0.0072}$	$0.02431^{+0.0069}_{-0.0064}$	11450	$0.00332^{+0.0018}_{-0.0017}$	$0.00467^{+0.0017}_{-0.0014}$	11801
t_{LLR}	-	-	-	$0.00002^{+0.00001}_{-0.00001}$	$0.00002^{+0.00001}_{-0.00001}$	6877
	pow_-deformation			$\mathcal{N}$-deformation		
Statistic	$\epsilon_{95\%\text{CL}}$	$\epsilon_{99\%\text{CL}}$	t (s)	$\epsilon_{95\%\text{CL}}$	$\epsilon_{99\%\text{CL}}$	t (s)
t_{SW}	$0.00604^{+0.0017}_{-0.0023}$	$0.00825^{+0.0016}_{-0.0018}$	**3051**	$0.19318^{+0.025}_{-0.039}$	$0.22704^{+0.019}_{-0.026}$	**2403**
$t_{\overline{\text{KS}}}$	$\mathbf{0.00042^{+0.00015}_{-0.00018}}$	$\mathbf{0.00061^{+0.00013}_{-0.00015}}$	3372	$\mathbf{0.00751^{+0.002}_{-0.0024}}$	$\mathbf{0.00993^{+0.0018}_{-0.002}}$	2934
t_{SKS}	$0.00441^{+0.00092}_{-0.0014}$	$0.00574^{+0.00077}_{-0.00094}$	3324	$0.15874^{+0.023}_{-0.034}$	$0.18473^{+0.019}_{-0.023}$	2726
t_{FGD}	$0.00722^{+0.0021}_{-0.0027}$	$0.00987^{+0.0016}_{-0.0019}$	4892	$0.18095^{+0.023}_{-0.038}$	$0.21269^{+0.016}_{-0.02}$	3756
t_{MMD}	$0.00353^{+0.0016}_{-0.0015}$	$0.00494^{+0.0014}_{-0.0012}$	11418	$0.43531^{+0.066}_{-0.11}$	$0.51609^{+0.045}_{-0.054}$	8642
t_{LLR}	$0.00002^{+0.00001}_{-0.00001}$	$0.00002^{+0.00001}_{-0.00001}$	6991	-	-	-
	$\mathcal{U}$-deformation			Timing		
Statistic	$\epsilon_{95\%\text{CL}}$	$\epsilon_{99\%\text{CL}}$	t (s)	t^{null} (s)		
t_{SW}	$0.33394^{+0.044}_{-0.068}$	$0.39248^{+0.033}_{-0.044}$	**2354**	338		
$t_{\overline{\text{KS}}}$	$\mathbf{0.01211^{+0.003}_{-0.0035}}$	$\mathbf{0.01575^{+0.0027}_{-0.003}}$	2835	**155**		
t_{SKS}	$0.27395^{+0.041}_{-0.059}$	$0.3188^{+0.033}_{-0.04}$	2601	509		
t_{FGD}	$0.31409^{+0.04}_{-0.07}$	$0.36919^{+0.027}_{-0.036}$	3643	2795		
t_{MMD}	$0.75353^{+0.12}_{-0.18}$	$0.89336^{+0.078}_{-0.098}$	7700	13860		
t_{LLR}	-	-	-	-		

Table 5.2: Upper bound on ϵ at the 95% and 99% CL for all the considered metrics and deformations in the MoG model with $d = 20$, $q = 5$, and $n = m = 5 \cdot 10^4$ samples. $t^{\text{null}}(s)$ is the total time in seconds to compute the values of the test statistic under the null hypothesis. $t(s)$ is the total time in seconds needed for the optimization in ϵ.

- finally, for the random deformations $\mathcal{N}$ and $\mathcal{U}$, $\overline{\text{KS}}$ is the most sensitive for the MoG models, with a gap of around one order of magnitude or even more, while FGD is the most sensitive for the CG models, up to about a factor of two compared to the other metrics.

The overall picture singles out the $\overline{\text{KS}}$ as generally the most sensitive metric for the MoG models, and the FGD as the most sensitive for the CG models. The great sensitivity of the $\overline{\text{KS}}$ for the MoG models is likely due to the fact that the KS-test is very sensitive for multi-modal distributions since a slight relative difference between the modes reflects very strongly on the difference in the cumulative distribution functions. The FGD, on the other hand, is likely the most sensitive for the CG models because it is designed as a distance between multivariate Gaussians, with sensitivity to

CG model with d = 20 and n = m = $5 \cdot 10^4$						
	μ-deformation			Σ_{ii}-deformation		
Statistic	$\epsilon_{95\%\text{CL}}$	$\epsilon_{99\%\text{CL}}$	t (s)	$\epsilon_{95\%\text{CL}}$	$\epsilon_{99\%\text{CL}}$	t (s)
t_{SW}	$0.04948^{+0.022}_{-0.021}$	$0.06621^{+0.021}_{-0.02}$	571	$0.02059^{+0.0066}_{-0.0078}$	$0.02732^{+0.0061}_{-0.0065}$	617
$t_{\overline{\text{KS}}}$	$\mathbf{0.04811^{+0.022}_{-0.021}}$	$0.06605^{+0.021}_{-0.02}$	**407**	$0.02898^{+0.011}_{-0.012}$	$0.04029^{+0.0097}_{-0.01}$	**434**
t_{SKS}	$0.04841^{+0.021}_{-0.021}$	$\mathbf{0.06372^{+0.02}_{-0.02}}$	655	$0.02623^{+0.0087}_{-0.01}$	$0.03417^{+0.0082}_{-0.0086}$	694
t_{FGD}	$0.05029^{+0.026}_{-0.022}$	$0.06539^{+0.024}_{-0.02}$	1886	$\mathbf{0.01695^{+0.007}_{-0.007}}$	$\mathbf{0.02215^{+0.0065}_{-0.0059}}$	1994
t_{MMD}	$0.0596^{+0.028}_{-0.02}$	$0.08041^{+0.026}_{-0.02}$	7733	$0.02325^{+0.011}_{-0.0079}$	$0.03109^{+0.01}_{-0.0079}$	8173
t_{LLR}	$0.00556^{+0.0031}_{-0.003}$	$0.00795^{+0.003}_{-0.003}$	2441	$0.00153^{+0.001}_{-0.00098}$	$0.0022^{+0.00098}_{-0.00099}$	3081
	$\Sigma_{i\neq j}$-deformation			pow_{+}-deformation		
Statistic	$\epsilon_{95\%\text{CL}}$	$\epsilon_{99\%\text{CL}}$	t (s)	$\epsilon_{95\%\text{CL}}$	$\epsilon_{99\%\text{CL}}$	t (s)
t_{SW}	$0.02783^{+0.0077}_{-0.0099}$	$0.03884^{+0.0064}_{-0.0076}$	**1073**	$0.0046^{+0.0017}_{-0.0019}$	$0.00614^{+0.0016}_{-0.0017}$	642
$t_{\overline{\text{KS}}}$	$1.02831^{+0.015}_{-0.015}$	$1.04211^{+0.0046}_{-0.012}$	1401	$0.00602^{+0.002}_{-0.0024}$	$0.00806^{+0.0019}_{-0.0019}$	**459**
t_{SKS}	$0.03839^{+0.011}_{-0.014}$	$0.05106^{+0.01}_{-0.012}$	1172	$0.00505^{+0.0017}_{-0.002}$	$0.00646^{+0.0016}_{-0.0017}$	747
t_{FGD}	$\mathbf{0.00483^{+0.0012}_{-0.0014}}$	$\mathbf{0.00631^{+0.0011}_{-0.0011}}$	3433	$0.00419^{+0.0019}_{-0.0018}$	$0.0054^{+0.0017}_{-0.0015}$	2765
t_{MMD}	$0.03094^{+0.017}_{-0.013}$	$0.04245^{+0.016}_{-0.013}$	8963	$\mathbf{0.00358^{+0.0018}_{-0.0012}}$	$\mathbf{0.00483^{+0.0016}_{-0.0013}}$	8839
t_{LLR}	-	-	-	$0.00042^{+0.00025}_{-0.00026}$	$0.00061^{+0.00025}_{-0.00025}$	2919
	pow_-deformation			$\mathcal{N}$-deformation		
Statistic	$\epsilon_{95\%\text{CL}}$	$\epsilon_{99\%\text{CL}}$	t (s)	$\epsilon_{95\%\text{CL}}$	$\epsilon_{99\%\text{CL}}$	t (s)
t_{SW}	$0.00455^{+5}_{-0.0017}$	$0.00609^{+5}_{-0.0015}$	587	$0.28641^{+0.047}_{-0.065}$	$0.33654^{+0.037}_{-0.046}$	535
$t_{\overline{\text{KS}}}$	$0.00575^{+0.002}_{-0.0022}$	$0.00773^{+0.0018}_{-0.0019}$	**461**	$0.32182^{+0.055}_{-0.08}$	$0.3832^{+0.045}_{-0.054}$	**393**
t_{SKS}	$0.00487^{+0.0017}_{-0.0019}$	$0.00632^{+0.0016}_{-0.0017}$	750	$0.28237^{+0.046}_{-0.066}$	$0.32811^{+0.038}_{-0.048}$	612
t_{FGD}	$0.00411^{+0.0017}_{-0.0015}$	$0.0054^{+0.0015}_{-0.0014}$	2758	$\mathbf{0.16992^{+0.02}_{-0.03}}$	$\mathbf{0.1944^{+0.014}_{-0.018}}$	2132
t_{MMD}	$\mathbf{0.00346^{+0.0019}_{-0.0014}}$	$\mathbf{0.00477^{+0.0018}_{-0.0014}}$	8990	$0.73852^{+0.086}_{-0.091}$	$0.85602^{+0.075}_{-0.062}$	5790
t_{LLR}	$0.00042^{+0.00025}_{-0.00026}$	$0.0006^{+0.00025}_{-0.00025}$	2930	-	-	-
	$\mathcal{U}$-deformation			Timing		
Statistic	$\epsilon_{95\%\text{CL}}$	$\epsilon_{99\%\text{CL}}$	t (s)	t^{null} (s)		
t_{SW}	$0.49513^{+0.079}_{-0.11}$	$0.5818^{+0.063}_{-0.078}$	512	313		
$t_{\overline{\text{KS}}}$	$0.55562^{+0.096}_{-0.14}$	$0.65585^{+0.083}_{-0.089}$	**378**	**127**		
t_{SKS}	$0.48849^{+0.085}_{-0.11}$	$0.56476^{+0.072}_{-0.079}$	582	480		
t_{FGD}	$\mathbf{0.2926^{+0.036}_{-0.05}}$	$\mathbf{0.33697^{+0.025}_{-0.034}}$	2042	3821		
t_{MMD}	$1.28521^{+0.15}_{-0.17}$	$1.49004^{+0.11}_{-0.12}$	6502	13843		
t_{LLR}	-	-	-	-		

Table 5.3: Upper bound on ϵ at the 95% and 99% CL for all the considered metrics and deformations in the CG model with $d = 20$ and $n = m = 5 \cdot 10^4$ samples. $t^{\text{null}}(s)$ is the total time in seconds to compute the values of the test statistic under the null hypothesis. $t(s)$ is the total time in seconds needed for the optimization in ϵ.

differences in both the mean and the covariance matrix.

Even though $\overline{\text{KS}}$ and FGD sometimes outperform other metrics, in most cases the results are comparable within uncertainties, which suggests, depending on the case, the use of the metric that is most efficient to compute, especially in optimization and model selection phases where many models need to be confronted with a reference or among each others. The computational efficiency can be read from the time needed to estimate the test statistic distribution under the null hypothesis, while the time reported for each deformation is the time needed for the optimization in ϵ, that means to estimate the upper bound on ϵ and its uncertainty. From the results for the null hypotheses we see that, in the case of the toy models, for which the data generating PDF is known and can

be efficiently sampled, the $\overline{\text{KS}}$ and SW metrics are generally the fastest for the computation of the distribution under the null hypothesis, respectively in low and high dimensionality, followed closely by the SKS. In contrast, the FGD and MMD are typically much slower. In particular, FGD slows down considerably with increasing dimensionality (large d), while MMD suffers more from increasing the sample size (large n, m).

The toy MoG and CG models are designed to highlight the differences that we discussed above. To better understand the performance of the metrics in a real-world scenario, where the data generating PDFs are not known, we now move to the JetNet dataset, with features corresponding to kinematic distributions of jets or particles within jets.

5.4.2 JetNet Dataset

We now present the results for the JetNet dataset, focusing on the particle and jet features datasets introduced in Section 5.3.2. As in the case of the toy models, we show only a small subset of figures and tables, which helps understanding the methodology and the overall picture of the results. In order to also check the dependence of the metrics performance on the scale of data, in this case we repeated the tests with original data and with data scaled to have zero mean and unit variance.

Before discussing the results, let us mention that, in the particle level dataset, we consider $n_{\text{part}} = 30$ particles per jet and we do not go above this number, even though the dataset contains up to 150 particles per jet. This is because as the number of particles increases, so does the number of soft particles, with a p_T distribution more and more similar to a δ-function peaked at zero momentum (likely also due to padding when the constituents generated by the Monte Carlo is less than 150). As some of the features become so narrow, then some metrics become more and more sensitive to smaller and smaller deformations that modify the features. This makes our numerical procedure ineffective. Indeed, if the reference distribution is a δ-function, than even a metric like KS is theoretically infinitely sensitive to any deformation. We find that stopping at $n_{\text{part}} = 30$ is a good compromise between a large number of features (three for each particle gives a total of 90 features) and robustness with respect to the aforementioned issue.

In Figure 6.9 we show the effect of the various deformations on the 1D marginal distributions in the jet kinematic variables for the jet level dataset. The plots show the distributions for the reference dataset and for the dataset deformed with all the deformations considered and with $\epsilon = 0.5$. The figures give an idea of how the kinematic distributions are affected by the different deformations.

To give an idea of how the correlations are also affected, we show in Figures 5.7 and 5.8 corner plots for the 1D and 2D marginal distributions corresponding to the μ and Σ_{ii} deformations, respectively, with $\epsilon = 0.5$ for the jet level dataset.

Even though figures are too large to be shown here, and are references in Appendix 5.A, the behavior of the kinematic variables and of the deformations in the case of the particle level dataset is very similar to what we have seen for the jet level dataset.

As for the toy models, the results for the JetNet dataset are summarized in tables and referenced in Appendix 6.C. We show here only a small subset of the result tables. Starting from the jet level

Jet features with $\mathbf{n = m = 5 \cdot 10^4}$						
	μ-deformation			Σ_{ii}-deformation		
Statistic	$\epsilon_{95\%\text{CL}}$	$\epsilon_{99\%\text{CL}}$	t (s)	$\epsilon_{95\%\text{CL}}$	$\epsilon_{99\%\text{CL}}$	t (s)
t_{SW}	$0.03049^{+0.019}_{-0.013}$	$0.04713^{+0.015}_{-0.015}$	**1108**	$0.04623^{+0.017}_{-0.025}$	$0.06323^{+0.019}_{-0.015}$	**1141**
$t_{\overline{\text{KS}}}$	$\mathbf{0.01585^{+0.0043}_{-0.0063}}$	$\mathbf{0.01927^{+0.0043}_{-0.0056}}$	17004	$\mathbf{0.02085^{+0.0064}_{-0.008}}$	$\mathbf{0.02567^{+0.006}_{-0.0075}}$	21589
t_{SKS}	$0.02815^{+0.013}_{-0.014}$	$0.03444^{+0.012}_{-0.014}$	35328	$0.04838^{+0.018}_{-0.019}$	$0.06304^{+0.016}_{-0.02}$	27128
t_{FGD}	$0.03986^{+0.025}_{-0.013}$	$0.06157^{+0.019}_{-0.017}$	11779	$0.04333^{+0.028}_{-0.023}$	$0.05934^{+0.027}_{-0.022}$	18470
t_{MMD}	$0.04941^{+0.034}_{-0.021}$	$0.0712^{+0.03}_{-0.022}$	78077	$0.07669^{+0.072}_{-0.035}$	$0.11237^{+0.068}_{-0.035}$	71427
	$\Sigma_{i\neq j}$-deformation			pow_+-deformation		
Statistic	$\epsilon_{95\%\text{CL}}$	$\epsilon_{99\%\text{CL}}$	t (s)	$\epsilon_{95\%\text{CL}}$	$\epsilon^{\text{pow}_+}_{99\%\text{CL}}$	t (s)
t_{SW}	$0.30801^{+0.08}_{-0.11}$	$0.45956^{+0.071}_{-0.063}$	**1033**	$0.02535^{+0.0077}_{-0.011}$	$0.03745^{+0.0066}_{-0.0084}$	**1028**
$t_{\overline{\text{KS}}}$	$1.01892^{+0.0084}_{-0.01}$	$1.02245^{+0.011}_{-0.0035}$	19934	$\mathbf{0.0232^{+0.0074}_{-0.011}}$	$\mathbf{0.02698^{+0.01}_{-0.0092}}$	35049
t_{SKS}	$0.2959^{+0.12}_{-0.12}$	$0.40074^{+0.11}_{-0.054}$	32727	$0.02709^{+0.014}_{-0.012}$	$0.03452^{+0.017}_{-0.012}$	28409
t_{FGD}	$\mathbf{0.22063^{+0.053}_{-0.082}}$	$\mathbf{0.29862^{+0.045}_{-0.052}}$	13459	$0.02454^{+0.015}_{-0.014}$	$0.0321^{+0.017}_{-0.012}$	11640
t_{MMD}	$0.80374^{+0.26}_{-0.28}$	$1.05932^{+0.078}_{-0.1}$	31136	$0.02933^{+0.019}_{-0.015}$	$0.03749^{+0.021}_{-0.016}$	54684
	pow_--deformation			$\mathcal{N}$-deformation		
Statistic	$\epsilon_{95\%\text{CL}}$	$\epsilon^{\text{pow}_-}_{99\%\text{CL}}$	t (s)	$\epsilon_{95\%\text{CL}}$	$\epsilon^{\mathcal{N}}_{99\%\text{CL}}$	t (s)
t_{SW}	$0.02553^{+0.0078}_{-0.0088}$	$0.03665^{+0.0074}_{-0.0068}$	**1080**	$0.12904^{+0.029}_{-0.034}$	$0.16235^{+0.02}_{-0.025}$	**981**
$t_{\overline{\text{KS}}}$	$\mathbf{0.02125^{+0.01}_{-0.0092}}$	$\mathbf{0.02649^{+0.0074}_{-0.009}}$	15925	$\mathbf{0.10579^{+0.014}_{-0.019}}$	$\mathbf{0.11672^{+0.012}_{-0.016}}$	28786
t_{SKS}	$0.02682^{+0.012}_{-0.012}$	$0.03607^{+0.01}_{-0.012}$	47622	$0.11163^{+0.022}_{-0.023}$	$0.12765^{+0.017}_{-0.023}$	38615
t_{FGD}	$0.02511^{+0.017}_{-0.012}$	$0.03353^{+0.016}_{-0.01}$	18451	$0.16887^{+0.046}_{-0.052}$	$0.19783^{+0.043}_{-0.036}$	13634
t_{MMD}	$0.03^{+0.02}_{-0.014}$	$0.04112^{+0.021}_{-0.012}$	39156	$0.25305^{+0.085}_{-0.11}$	$0.29551^{+0.081}_{-0.073}$	52861
	$\mathcal{U}$-deformation			Timing		
Statistic	$\epsilon_{95\%\text{CL}}$	$\epsilon^{\mathcal{U}}_{99\%\text{CL}}$	t (s)	t^{null} (s)		
t_{SW}	$0.22631^{+0.05}_{-0.064}$	$0.27734^{+0.044}_{-0.039}$	**916**	**129**		
$t_{\overline{\text{KS}}}$	$\mathbf{0.18246^{+0.022}_{-0.032}}$	$\mathbf{0.19931^{+0.018}_{-0.027}}$	32276	1907		
t_{SKS}	$0.18837^{+0.033}_{-0.048}$	$0.21334^{+0.027}_{-0.029}$	38491	4382		
t_{FGD}	$0.27796^{+0.1}_{-0.074}$	$0.34469^{+0.068}_{-0.062}$	19098	1794		
t_{MMD}	$0.49303^{+0.16}_{-0.18}$	$0.57279^{+0.12}_{-0.11}$	55838	3504		

Table 5.4: Upper bound on ϵ at the 95% and 99% CL for all the considered metrics and deformations in the jet level dataset with $n = m = 5 \cdot 10^4$ samples using original features. $t^{\text{null}}(s)$ is the total time in seconds to compute the values of the test statistic under the null hypothesis. $t(s)$ is the total time in seconds needed for the optimization in ϵ.

dataset, Table 5.4 shows the upper bound on ϵ at the 95% and 99% CL for the jet level dataset with $n = m = 5 \cdot 10^4$ samples.

From the table we see that, except for the $\Sigma_{i\neq j}$-deformation, to which $\overline{\text{KS}}$ is not sensitive by construction, $\overline{\text{KS}}$ is the most sensitive to all other deformations. As expected from the toy results, for the $\Sigma_{i\neq j}$ deformation FGD is the most sensitive test. Despite $\overline{\text{KS}}$ being generally the most sensitive metric, one can see that, within uncertainties, results from all metrics are comparable and the gap among them is not particularly significant.

In contrast with the results in Table 5.4, we show in Table 5.5 the results obtained performing the tests on data scaled to zero mean and unit variance. As expected, the results for $\overline{\text{KS}}$ remain unchanged, since the KS test is "scale invariant", while the results for the other metrics are affected. Moreover, while FGD continues to be the most sensitive metric for the $\Sigma_{i\neq j}$ deformation, now with a large gap (one order of magnitude better than the best of the other metrics), the other metrics

Scaled Jet features with $\mathbf{n = m = 5 \cdot 10^4}$						
	μ-deformation			Σ_{ii}-deformation		
Statistic	$\epsilon_{95\%\text{CL}}$	$\epsilon_{99\%\text{CL}}$	t (s)	$\epsilon_{95\%\text{CL}}$	$\epsilon_{99\%\text{CL}}$	t (s)
t_{SW}	$0.01623^{+0.0045}_{-0.0069}$	$0.02098^{+0.0049}_{-0.0059}$	12410	$0.02089^{+0.0073}_{-0.008}$	$0.02834^{+0.0077}_{-0.0079}$	**1054**
$t_{\overline{\text{KS}}}$	$0.01585^{+0.0043}_{-0.0063}$	$0.01927^{+0.0043}_{-0.0056}$	17174	$\mathbf{0.02085^{+0.0064}_{-0.008}}$	$\mathbf{0.02567^{+0.006}_{-0.0075}}$	38871
t_{SKS}	$\mathbf{0.0113^{+0.0044}_{-0.005}}$	$\mathbf{0.0141^{+0.0037}_{-0.0045}}$	32620	$0.02254^{+0.0074}_{-0.0099}$	$0.02773^{+0.0073}_{-0.0089}$	28803
t_{FGD}	$0.02106^{+0.0062}_{-0.0079}$	$0.02659^{+0.0058}_{-0.0069}$	**11583**	$0.02133^{+0.0078}_{-0.0097}$	$0.02741^{+0.0071}_{-0.008}$	14254
t_{MMD}	$0.06739^{+0.013}_{-0.021}$	$0.08802^{+0.013}_{-0.011}$	46972	$0.0318^{+0.015}_{-0.0083}$	$0.04328^{+0.014}_{-0.012}$	28709
	$\Sigma_{i\neq j}$-deformation			pow$_+$-deformation		
Statistic	$\epsilon_{95\%\text{CL}}$	$\epsilon_{99\%\text{CL}}$	t (s)	$\epsilon_{95\%\text{CL}}$	$\epsilon^{\text{pow}_+}_{99\%\text{CL}}$	t (s)
t_{SW}	$0.0503^{+0.016}_{-0.019}$	$0.07052^{+0.015}_{-0.014}$	**1008**	$0.02465^{+0.011}_{-0.0081}$	$0.03314^{+0.0099}_{-0.0095}$	**1025**
$t_{\overline{\text{KS}}}$	$1.02009^{+0.0072}_{-0.001}$	$1.02812^{+0.003}_{-0.008}$	16410	$0.0232^{+0.0074}_{-0.011}$	$0.02698^{+0.01}_{-0.0092}$	35198
t_{SKS}	$0.06201^{+0.02}_{-0.029}$	$0.07573^{+0.02}_{-0.024}$	35383	$0.0402^{+0.015}_{-0.015}$	$0.04921^{+0.015}_{-0.015}$	47807
t_{FGD}	$\mathbf{0.00627^{+0.0016}_{-0.0018}}$	$\mathbf{0.00809^{+0.0015}_{-0.0018}}$	14008	$0.02237^{+0.013}_{-0.011}$	$0.0281^{+0.011}_{-0.0084}$	24967
t_{MMD}	$0.0794^{+0.039}_{-0.031}$	$0.112^{+0.031}_{-0.026}$	29620	$\mathbf{0.01898^{+0.012}_{-0.0094}}$	$\mathbf{0.02472^{+0.012}_{-0.0076}}$	66075
	pow$_-$-deformation			$\mathcal{N}$-deformation		
Statistic	$\epsilon_{95\%\text{CL}}$	$\epsilon^{\text{pow}_-}_{99\%\text{CL}}$	t (s)	$\epsilon_{95\%\text{CL}}$	$\epsilon^{\mathcal{N}}_{99\%\text{CL}}$	t (s)
t_{SW}	$0.02527^{+0.011}_{-0.011}$	$0.03513^{+0.0084}_{-0.01}$	**993**	$0.11836^{+0.027}_{-0.028}$	$0.14062^{+0.018}_{-0.026}$	**910**
$t_{\overline{\text{KS}}}$	$\mathbf{0.02125^{+0.01}_{-0.0092}}$	$\mathbf{0.02649^{+0.0074}_{-0.009}}$	16472	$0.10579^{+0.014}_{-0.019}$	$0.11672^{+0.012}_{-0.016}$	31727
t_{SKS}	$0.03986^{+0.013}_{-0.017}$	$0.04873^{+0.013}_{-0.013}$	27407	$0.08577^{+0.024}_{-0.028}$	$0.10148^{+0.021}_{-0.026}$	25899
t_{FGD}	$0.02163^{+0.015}_{-0.0097}$	$0.02954^{+0.014}_{-0.0087}$	12892	$\mathbf{0.07833^{+0.0094}_{-0.019}}$	$\mathbf{0.08847^{+0.0084}_{-0.0069}}$	13246
t_{MMD}	$0.02133^{+0.013}_{-0.0086}$	$0.02924^{+0.011}_{-0.0081}$	68458	$0.26032^{+0.037}_{-0.057}$	$0.29897^{+0.028}_{-0.036}$	42149
	$\mathcal{U}$-deformation			Timing		
Statistic	$\epsilon_{95\%\text{CL}}$	$\epsilon^{\mathcal{U}}_{99\%\text{CL}}$	t (s)	t^{null} (s)		
t_{SW}	$0.20487^{+0.042}_{-0.048}$	$0.2434^{+0.032}_{-0.035}$	**877**	**123**		
$t_{\overline{\text{KS}}}$	$0.18018^{+0.024}_{-0.035}$	$0.19884^{+0.018}_{-0.027}$	25630	1913		
t_{SKS}	$0.14529^{+0.04}_{-0.056}$	$0.1719^{+0.035}_{-0.048}$	42277	4383		
t_{FGD}	$\mathbf{0.13545^{+0.014}_{-0.032}}$	$\mathbf{0.15299^{+0.015}_{-0.012}}$	12782	1787		
t_{MMD}	$0.45177^{+0.066}_{-0.091}$	$0.52083^{+0.05}_{-0.047}$	56078	3504		

Table 5.5: Upper bound on ϵ at the 95% and 99% CL for all the considered metrics and deformations in the jet level dataset with $n = m = 5 \cdot 10^4$ samples using features scaled to zero mean and unit variance. $t^{\text{null}}(s)$ is the total time in seconds to compute the values of the test statistic under the null hypothesis. $t(s)$ is the total time in seconds needed for the optimization in ϵ.

become now more sensitive and compete with $\overline{\text{KS}}$ for the other deformations. Nevertheless, results for the other deformations are comparable for the different metrics within uncertainties.

We can now proceed to discuss the results for the particle level dataset. Table 5.6 shows the upper bound on ϵ at the 95% and 99% CL for the particle level dataset with $n = m = 5 \cdot 10^4$ samples. Zeros in the table represent values smaller than 10^{-5} obtained from the optimization within a specified tolerance.

From the table we start to observe the effect discussed at the beginning of the Section, where $\overline{\text{KS}}$ becomes much more sensitive than the other tests (for all deformations except $\Sigma_{i\neq j}$). This increased sensitivity is due to the presence of several features with a narrow distribution, for which $\overline{\text{KS}}$ is particularly sensitive. Moreover, despite a few exceptions, all other metrics generally perform equally well within uncertainties.

Particle features with $\mathbf{n = m = 5 \cdot 10^4}$						
	μ-deformation			Σ_{ii}-deformation		
Statistic	$\epsilon_{95\%\text{CL}}$	$\epsilon_{99\%\text{CL}}$	t (s)	$\epsilon_{95\%\text{CL}}$	$\epsilon_{99\%\text{CL}}$	t (s)
t_{SW}	$0.02633^{+0.0098}_{-0.013}$	$0.03714^{+0.0084}_{-0.0097}$	**849**	$0.02913^{+0.012}_{-0.0079}$	$0.04108^{+0.0093}_{-0.011}$	**824**
$t_{\overline{\text{KS}}}$	$\mathbf{0.0^{+0.0045}_{-0.0}}$	$\mathbf{0.00771^{+0.0022}_{-0.0068}}$	49525	$\mathbf{0.0^{+0.013}_{-0.0}}$	$\mathbf{0.01904^{+0.0086}_{-0.011}}$	55017
t_{SKS}	$0.01592^{+0.0046}_{-0.0061}$	$0.02334^{+0.0058}_{-0.0059}$	17572	$0.02735^{+0.0049}_{-0.01}$	$0.03362^{+0.0081}_{-0.0071}$	24987
t_{FGD}	$0.04749^{+0.012}_{-0.024}$	$0.06462^{+0.013}_{-0.013}$	30820	$0.04004^{+0.017}_{-0.012}$	$0.0556^{+0.018}_{-0.016}$	25551
t_{MMD}	$0.1396^{+0.1}_{-0.065}$	$0.21274^{+0.071}_{-0.065}$	18527	$0.06988^{+0.048}_{-0.031}$	$0.0986^{+0.037}_{-0.036}$	33217
	$\Sigma_{i\neq j}$-deformation			pow_+-deformation		
Statistic	$\epsilon_{95\%\text{CL}}$	$\epsilon_{99\%\text{CL}}$	t (s)	$\epsilon_{95\%\text{CL}}$	$\epsilon^{\text{pow}_+}_{99\%\text{CL}}$	t (s)
t_{SW}	$0.04883^{+0.012}_{-0.015}$	$0.06979^{+0.0097}_{-0.016}$	**1966**	$0.02745^{+0.011}_{-0.0075}$	$0.03872^{+0.0088}_{-0.011}$	**806**
$t_{\overline{\text{KS}}}$	$0.99933^{+0.0085}_{-0.014}$	$1.01732^{+0.006}_{-0.0079}$	11225	$\mathbf{0.0^{+0.0066}_{-0.0}}$	$\mathbf{0.01141^{+0.0073}_{-0.011}}$	46010
t_{SKS}	$0.04267^{+0.018}_{-0.012}$	$0.06018^{+0.014}_{-0.013}$	29568	$0.03594^{+0.021}_{-0.016}$	$0.05069^{+0.011}_{-0.014}$	24821
t_{FGD}	$\mathbf{0.02641^{+0.0058}_{-0.012}}$	$\mathbf{0.03966^{+0.006}_{-0.0079}}$	28408	$0.02459^{+0.013}_{-0.012}$	$0.03501^{+0.013}_{-0.012}$	25798
t_{MMD}	$0.25965^{+0.068}_{-0.086}$	$0.35327^{+0.073}_{-0.056}$	16061	$0.02054^{+0.014}_{-0.0071}$	$0.02657^{+0.013}_{-0.0091}$	26195
	pow_--deformation			$\mathcal{N}$-deformation		
Statistic	$\epsilon_{95\%\text{CL}}$	$\epsilon^{\text{pow}_-}_{99\%\text{CL}}$	t (s)	$\epsilon_{95\%\text{CL}}$	$\epsilon^{\mathcal{N}}_{99\%\text{CL}}$	t (s)
t_{SW}	$0.02745^{+0.011}_{-0.014}$	$0.03872^{+0.0088}_{-0.01}$	**809**	$0.10733^{+0.022}_{-0.026}$	$0.13357^{+0.016}_{-0.016}$	**691**
$t_{\overline{\text{KS}}}$	$\mathbf{0.0^{+0.0095}_{-0.0}}$	$\mathbf{0.01323^{+0.0069}_{-0.0085}}$	45685	$\mathbf{0.0656^{+0.016}_{-0.053}}$	$\mathbf{0.08707^{+0.013}_{-0.016}}$	7484
t_{SKS}	$0.03777^{+0.015}_{-0.017}$	$0.04837^{+0.014}_{-0.015}$	15966	$0.08456^{+0.013}_{-0.013}$	$0.09935^{+0.0089}_{-0.01}$	18276
t_{FGD}	$0.02241^{+0.019}_{-0.01}$	$0.0353^{+0.019}_{-0.011}$	26549	$0.14608^{+0.034}_{-0.038}$	$0.1758^{+0.023}_{-0.021}$	23330
t_{MMD}	$0.02077^{+0.017}_{-0.01}$	$0.02939^{+0.016}_{-0.0089}$	20263	$0.33827^{+0.088}_{-0.089}$	$0.37964^{+0.091}_{-0.073}$	19908
	$\mathcal{U}$-deformation			Timing		
Statistic	$\epsilon_{95\%\text{CL}}$	$\epsilon^{\mathcal{U}}_{99\%\text{CL}}$	t (s)	t^{null} (s)		
t_{SW}	$0.1889^{+0.038}_{-0.046}$	$0.2351^{+0.028}_{-0.04}$	**625**	**150**		
$t_{\overline{\text{KS}}}$	$\mathbf{0.10693^{+0.022}_{-0.086}}$	$\mathbf{0.14193^{+0.021}_{-0.019}}$	13565	2126		
t_{SKS}	$0.14453^{+0.02}_{-0.028}$	$0.17795^{+0.0057}_{-0.023}$	17723	4818		
t_{FGD}	$0.25168^{+0.053}_{-0.065}$	$0.30289^{+0.04}_{-0.036}$	23243	7351		
t_{MMD}	$0.58039^{+0.17}_{-0.15}$	$0.6876^{+0.14}_{-0.14}$	25557	3880		

Table 5.6: Upper bound on ϵ at the 95% and 99% CL for all the considered metrics and deformations in the particle level dataset with $n = m = 5 \cdot 10^4$ samples using original features. $t^{\text{null}}(s)$ is the total time in seconds to compute the values of the test statistic under the null hypothesis. $t(s)$ is the total time in seconds needed for the optimization in ϵ.

Finally, Table 5.7 shows the results obtained using features from the particle level dataset scaled to zero mean and unit variance. As in the case of the jet level dataset, the results for $\overline{\text{KS}}$ remain unchanged, while the results for the other metrics are affected by the scaling of the data. The FGD metric becomes very sensitive to the $\Sigma_{i\neq j}$ deformation, outperforming the other metrics by more than one order of magnitude. Despite all other tests showing an increased sensitivity, they still struggle to compete with $\overline{\text{KS}}$, except in the case of the random deformations, where FGD is the most sensitive. Even in these cases, the gap is not particularly significant, and, at least compared to $\overline{\text{KS}}$, lies within uncertainties.

In summary, also for the JetNet dataset, the $\overline{\text{KS}}$ test and the FGD test are the most sensitive for most of the deformations and for most values of the sample size. With the exception of some special cases, the SW test gives similar results, within uncertainties, with a testing time that is up to fifty times smaller than that of the other tests, and always below one hour, even for sample sizes

Scaled Particle features with $\mathbf{n = m = 5 \cdot 10^4}$						
	μ-deformation			Σ_{ii}-deformation		
Statistic	$\epsilon_{95\%\text{CL}}$	$\epsilon_{99\%\text{CL}}$	t (s)	$\epsilon_{95\%\text{CL}}$	$\epsilon_{99\%\text{CL}}$	t (s)
t_{SW}	$0.01334^{+0.0038}_{-0.0046}$	$0.01815^{+0.0037}_{-0.0029}$	**1116**	$0.0166^{+0.0059}_{-0.0063}$	$0.02125^{+0.006}_{-0.0034}$	**1079**
$t_{\overline{\text{KS}}}$	$\mathbf{0.0^{+0.0045}_{-0.0}}$	$\mathbf{0.00771^{+0.0022}_{-0.0049}}$	58835	$\mathbf{0.0^{+0.013}_{-0.0}}$	$0.01904^{+0.0086}_{-0.011}$	62555
t_{SKS}	$0.01275^{+0.0034}_{-0.0043}$	$0.01734^{+0.0036}_{-0.0028}$	18356	$0.02131^{+0.007}_{-0.0073}$	$0.02899^{+0.006}_{-0.0047}$	26542
t_{FGD}	$0.01627^{+0.003}_{-0.006}$	$0.02025^{+0.0024}_{-0.0047}$	39057	$0.01469^{+0.0034}_{-0.0057}$	$\mathbf{0.01805^{+0.0043}_{-0.0051}}$	27175
t_{MMD}	$0.01613^{+0.0049}_{-0.0058}$	$0.02141^{+0.0032}_{-0.0035}$	22841	$0.01606^{+0.0074}_{-0.0066}$	$0.02089^{+0.0055}_{-0.0061}$	33730
	$\Sigma_{i\neq j}$-deformation			pow$_+$-deformation		
Statistic	$\epsilon_{95\%\text{CL}}$	$\epsilon_{99\%\text{CL}}$	t (s)	$\epsilon_{95\%\text{CL}}$	$\epsilon^{\text{pow}_+}_{99\%\text{CL}}$	t (s)
t_{SW}	$0.03128^{+0.0068}_{-0.011}$	$0.04152^{+0.0061}_{-0.0067}$	**1424**	$0.01894^{+0.0055}_{-0.0072}$	$0.02425^{+0.0068}_{-0.0039}$	**1006**
$t_{\overline{\text{KS}}}$	$0.99134^{+0.016}_{-0.0078}$	$1.01532^{+0.008}_{-0.004}$	11987	$\mathbf{0.0^{+0.0066}_{-0.0}}$	$\mathbf{0.01141^{+0.0073}_{-0.011}}$	49091
t_{SKS}	$0.03809^{+0.011}_{-0.013}$	$0.0515^{+0.011}_{-0.0083}$	27313	$0.03552^{+0.0055}_{-0.013}$	$0.04366^{+0.01}_{-0.0066}$	15487
t_{FGD}	$\mathbf{0.0026^{+0.00076}_{-0.00089}}$	$\mathbf{0.00345^{+0.00052}_{-0.00076}}$	33338	$0.01534^{+0.0052}_{-0.0062}$	$0.01886^{+0.0045}_{-0.0055}$	24241
t_{MMD}	$0.01919^{+0.011}_{-0.0079}$	$0.02614^{+0.0089}_{-0.0065}$	20604	$0.01896^{+0.0074}_{-0.008}$	$0.02428^{+0.0068}_{-0.0071}$	27198
	pow$_-$-deformation			$\mathcal{N}$-deformation		
Statistic	$\epsilon_{95\%\text{CL}}$	$\epsilon^{\text{pow}_-}_{99\%\text{CL}}$	t (s)	$\epsilon_{95\%\text{CL}}$	$\epsilon^{\mathcal{N}}_{99\%\text{CL}}$	t (s)
t_{SW}	$0.01909^{+0.0073}_{-0.0077}$	$0.02693^{+0.0061}_{-0.0068}$	**1006**	$0.10868^{+0.02}_{-0.017}$	$0.1277^{+0.011}_{-0.022}$	**886**
$t_{\overline{\text{KS}}}$	$\mathbf{0.0^{+0.0095}_{-0.0}}$	$\mathbf{0.01323^{+0.0069}_{-0.0085}}$	45323	$0.0656^{+0.016}_{-0.049}$	$0.08707^{+0.013}_{-0.019}$	22186
t_{SKS}	$0.0356^{+0.0093}_{-0.013}$	$0.04726^{+0.007}_{-0.011}$	22261	$0.10733^{+0.022}_{-0.017}$	$0.13357^{+0.016}_{-0.026}$	24344
t_{FGD}	$0.01543^{+0.007}_{-0.0065}$	$0.01852^{+0.0068}_{-0.0042}$	24968	$\mathbf{0.04853^{+0.0071}_{-0.0075}}$	$\mathbf{0.05702^{+0.0051}_{-0.006}}$	24273
t_{MMD}	$0.01859^{+0.0085}_{-0.0081}$	$0.02501^{+0.0083}_{-0.0064}$	27960	$0.26953^{+0.035}_{-0.052}$	$0.30333^{+0.029}_{-0.011}$	19782
	$\mathcal{U}$-deformation			Timing		
Statistic	$\epsilon_{95\%\text{CL}}$	$\epsilon^{\mathcal{U}}_{99\%\text{CL}}$	t (s)	t^{null} (s)		
t_{SW}	$0.19116^{+0.035}_{-0.03}$	$0.22462^{+0.02}_{-0.039}$	**774**	**133**		
$t_{\overline{\text{KS}}}$	$0.10693^{+0.022}_{-0.086}$	$0.14193^{+0.021}_{-0.019}$	10646	1972		
t_{SKS}	$0.19116^{+0.035}_{-0.03}$	$0.22462^{+0.02}_{-0.039}$	16154	4379		
t_{FGD}	$\mathbf{0.08969^{+0.006}_{-0.016}}$	$\mathbf{0.10215^{+0.0042}_{-0.016}}$	21825	6689		
t_{MMD}	$0.48398^{+0.032}_{-0.088}$	$0.5512^{+0.022}_{-0.058}$	14676	3605		

Table 5.7: Upper bound on ϵ at the 95% and 99% CL for all the considered metrics and deformations in the particle level dataset with $n = m = 5 \cdot 10^4$ samples features scaled to zero mean and unit variance. $t^{\text{null}}(s)$ is the total time in seconds to compute the values of the test statistic under the null hypothesis. $t(s)$ is the total time in seconds needed for the optimization in ϵ.

of $5 \cdot 10^4$.

Notice that, while the $\overline{\text{KS}}$ metric is extremely fast to evaluate in the case of the toy models, for most of the configurations in terms of dimensions and number of samples, it becomes much less efficient for the JetNet dataset. This is due to the fact that even though $\overline{\text{KS}}$ is highly parallelizable, we could obtain a very efficient implementation only when the data generating distribution is available, while we had more troubles getting the most out of parallel computation with resampling techniques. This is not the same for the SW test, which remains extremely efficient also in the case of resampling.

5.5 Conclusions and Outlook

In this work, we have developed and thoroughly tested a comprehensive framework for comparing non-parametric two-sample tests in the context of the evaluation of high-dimensional generative models. This type of statistical tests are designed to detect deformation in the shape or correlation structure of a distribution from samples of finite size. Our analysis has focused on several key metrics, including the sliced Wasserstein distance, the mean of 1D Kolmogorov-Smirnov tests, a sliced variation of the Kolmogorov-Smirnov test, the unbiased quadratic Maximum Mean Discrepancy, and the unbiased Fréchet Gaussian Distance. Where possible, we have also benchmarked these non-parametric methods against the parametric log-likelihood-ratio test.

Our results demonstrate that non-parametric tests based on 1D marginal distributions or 1D projections can achieve sensitivity comparable to more complex multivariate metrics. This is particularly significant in high-dimensional settings where multivariate tests often struggle due to the curse of dimensionality, in contrast with the need for computational efficiency. Tests such as the sliced Wasserstein distance and the Kolmogorov-Smirnov-inspired tests show strong performance while maintaining lower computational costs, making them well suited for tasks involving large sample sizes and high-dimensional data, especially in the phase of model selection and optimization of generative models.

As expected, the log-likelihood-ratio test consistently outperforms non-parametric methods when the underlying probability densities of both the reference and deformed distributions are known. However, in practical applications where exact probability densities are not available, such as for the JetNet dataset we considered, non-parametric approaches are indispensable.

Our methodology has proven to be flexible, scalable, and capable of accommodating various deformations across a wide range of dimensions and sample sizes. The combination of toy models and real-world datasets used in this study provides a valuable benchmark, reflecting the theoretical and practical challenges faced in generative model validation.

Furthermore, with few exceptions, each metric exhibits similar sensitivity to the value of ϵ across different datasets for the same deformation. This consistency demonstrates that our procedure provides a robust evaluation of the tests themselves, rather than merely assessing the specific models and deformations considered.

Looking ahead, there are several promising directions for future research and applications of this methodology:

1. Extension to classifier-based metrics: Various classifier-based two-sample tests have been proposed in the literature, demonstrating significant potential for HEP applications. However, a thorough study of their relative and absolute performance remains absent. The methodology presented here offers a straightforward approach for their evaluation and validation.

2. State-of-the-art generative models: Our study shows that evaluation metrics based on one-dimensional tests can be both competitive and efficient, making them well-suited for model

selection. The next step would be to apply these metrics to advanced generative models, including Generative Adversarial Networks, Variational Autoencoders, and Normalizing Flows, particularly in scientific applications where precision and reliability are essential.

3. Uncertainty quantification: Although our methodology provides a well-defined procedure for assigning uncertainty to test results, future work could explore more comprehensive approaches. For instance, in the case of generative models where deformations are not parameterized by a single parameter ϵ, the full distribution of the test statistic under the alternative hypothesis could be estimated, enabling a more detailed assessment of the test power. This could also allow for the computation of the complete confusion matrix, providing deeper insights into the statistical properties of the tests.

4. Enhancement of computational efficiency: While our current implementation of the proposed metrics in TensorFlow2 is computationally efficient, there remains room for further optimization. Improvements in computational efficiency would make these methods more accessible and scalable for even larger datasets, facilitating their application in real-world problems.

In conclusion, we believe that the methodology developed in this work represents a step forward in the key challenge of evaluating generative models, particularly in scientific applications where precision and reliability are key requirements.

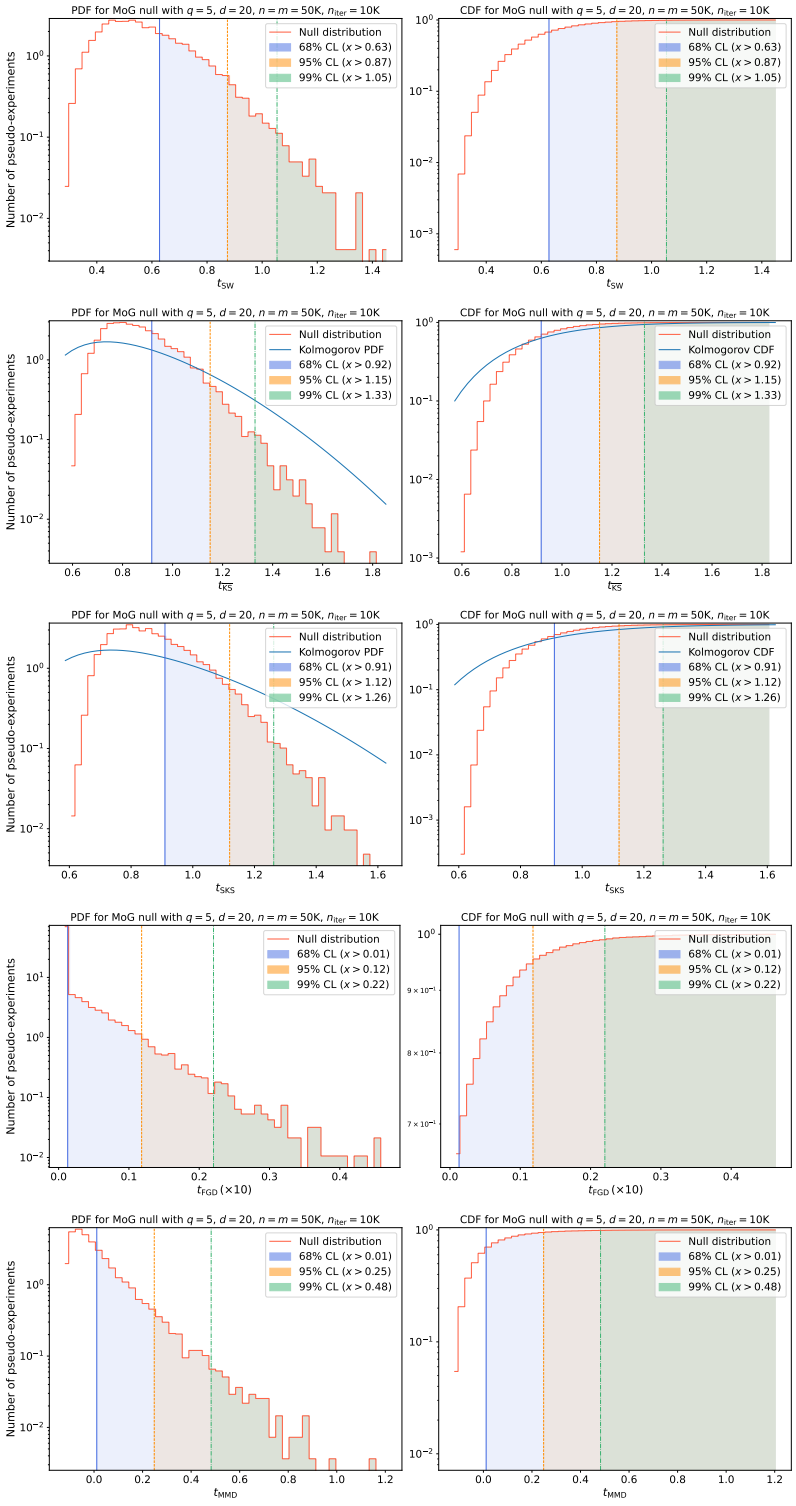


Figure 5.5: Each pair of plots represents the empirical PDF (left) and CDF (right) of the test statistic under the null hypothesis for the MoG model with $d = 20$ and $q = 5$, and $n = m = 5 \cdot 10^4$ samples. The histograms are built with 10^4 evaluations of the test statistics. From top to bottom, the five panels represent the SW, $\overline{\text{KS}}$, SKS, FGD, and MMD tests. See the main text for the full description.

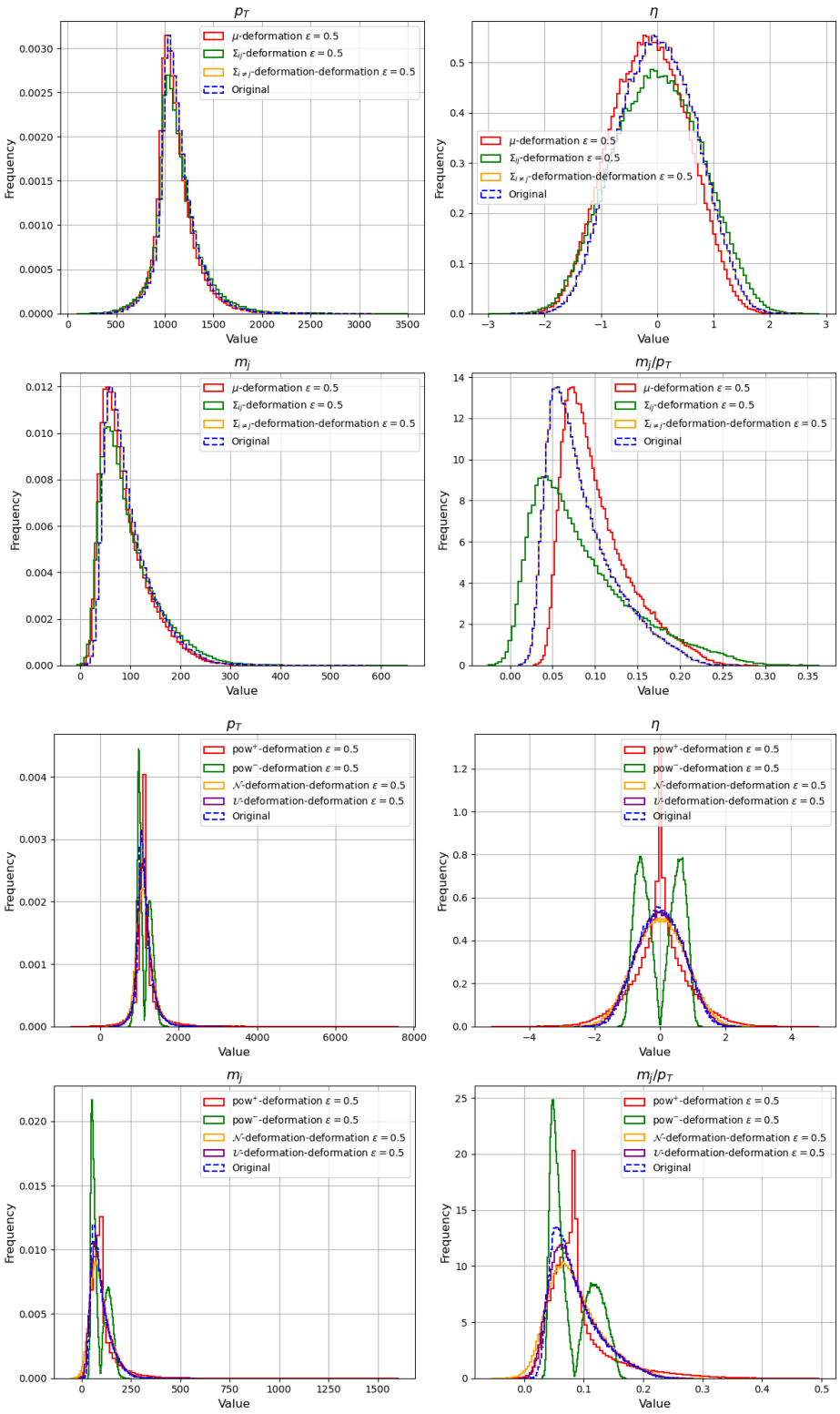


Figure 5.6: Original jet kinematic distributions compared with the μ, Σ_{ii}, and $\Sigma_{i\neq j}$ (left), and pow$^+$, pow$^-$, $\mathcal{N}$, and $\mathcal{U}$ (right) deformations with $\epsilon = 0.5$. The plots are made with 10^6 points per sample.

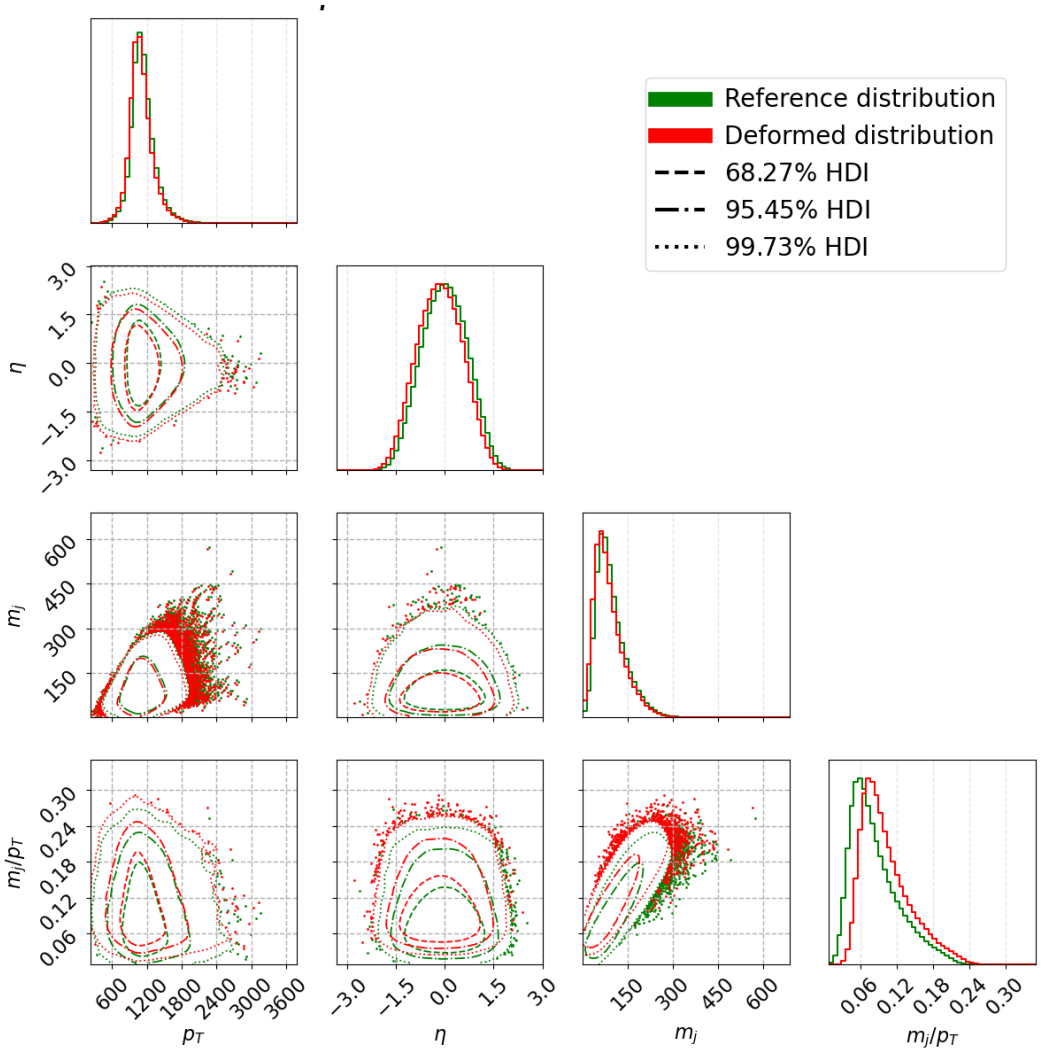


Figure 5.7: Corner plots of the original jet kinematic distributions compared with the μ deformation with $\epsilon = 0.5$. The plots are made with 10^6 points per sample.

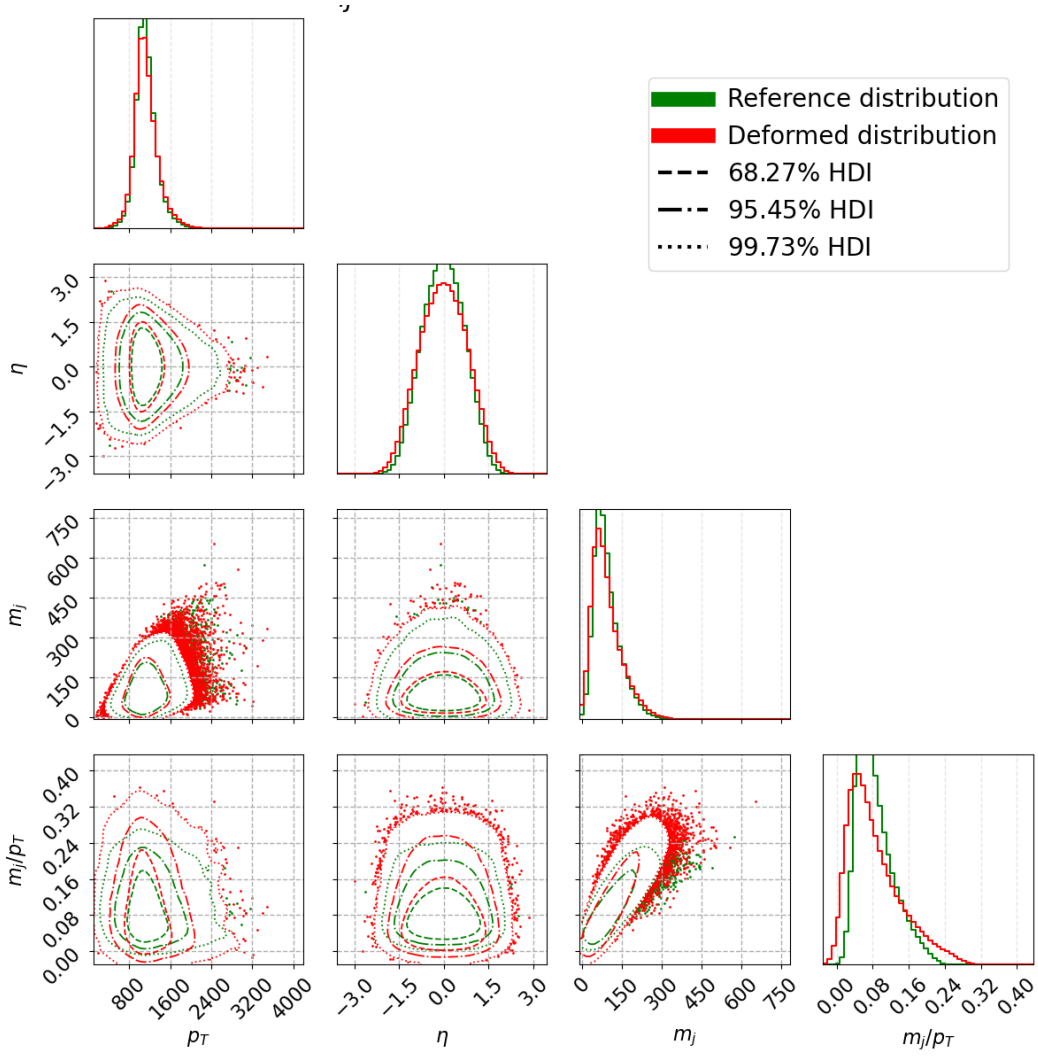


Figure 5.8: Corner plots of the original jet kinematic distributions compared with the Σ_{ii} deformations with $\epsilon = 0.5$. The plots are made with 10^6 points per sample.

Appendix

5.A List of Figures

5.A.1 List of Figures for the CG models

List of Figures for CG results in $5D$

The following Figures show the 1D and 2D marginal probability distributions for two samples $\mathbf{X}_1$, drawn from the reference distribution of the CG model with $d = 5$, and $\mathbf{X}_2$, corresponding to the specified deformation. The plots are made with 10^6 points per sample.

- CG_5D_corner_def_1: μ-deformation with $\epsilon = 0.5$.
- CG_5D_corner_def_2: Σ_{ii}-deformation with $\epsilon = 0.5$.
- CG_5D_corner_def_3: $\Sigma_{i\neq j}$-deformation with $\epsilon = 0.5$.
- CG_5D_corner_def_4: pow$^+$-deformation with $\epsilon = 0.1$.
- CG_5D_corner_def_5: pow$^-$-deformation with $\epsilon = 0.1$.
- CG_5D_corner_def_6: $\mathcal{N}$-deformation with $\epsilon = 0.5$.
- CG_5D_corner_def_7: $\mathcal{U}$-deformation with $\epsilon = 0.5$.

The following Figures report a color plot representation of the correlation matrices for two samples $\mathbf{X}_1$, drawn from the reference distribution of the CG model with $d = 5$, and $\mathbf{X}_2$, corresponding to the specified deformation. The plots are made with 10^6 points per sample.

- CG_5D_corr_matrix_def_1: μ-deformation with $\epsilon = 0.5$.
- CG_5D_corr_matrix_def_2: Σ_{ii}-deformation with $\epsilon = 0.5$.
- CG_5D_corr_matrix_def_3: $\Sigma_{i\neq j}$-deformation with $\epsilon = 0.5$.
- CG_5D_corr_matrix_def_4: pow$^+$-deformation with $\epsilon = 0.1$.
- CG_5D_corr_matrix_def_5: pow$^-$-deformation with $\epsilon = 0.1$.

- CG_5D_corr_matrix_def_6: $\mathcal{N}$-deformation with $\epsilon = 0.5$.
- CG_5D_corr_matrix_def_7: $\mathcal{U}$-deformation with $\epsilon = 0.5$.

The following Figures show the empirical PDF of the specified test statistic under the null hypothesis (both samples come from the same reference distribution) for the CG model with $d = 5$ and $n = m = (1, 2, 5, 10) \cdot 10^4$. The test statistics distributions are obtained with 10^4 iterations.

- CG_5D_10K_null_SW
  CG_5D_20K_null_SW
  CG_5D_50K_null_SW
  CG_5D_100K_null_SW
- CG_5D_10K_null_$\overline{\text{KS}}$
  CG_5D_20K_null_$\overline{\text{KS}}$
  CG_5D_50K_null_$\overline{\text{KS}}$
  CG_5D_100K_null_$\overline{\text{KS}}$
- CG_5D_10K_null_SKS
  CG_5D_20K_null_SKS
  CG_5D_50K_null_SKS
  CG_5D_100K_null_SKS
- CG_5D_10K_null_FGD
  CG_5D_20K_null_FGD
  CG_5D_50K_null_FGD
  CG_5D_100K_null_FGD
- CG_5D_10K_null_MMD
  CG_5D_20K_null_MMD
  CG_5D_50K_null_MMD
  CG_5D_100K_null_MMD

List of Figures for CG results in $20D$

The following Figures show the 1D and 2D marginal probability distributions for two samples $\mathbf{X}_1$, drawn from the reference distribution of the CG model with $d = 20$, and $\mathbf{X}_2$, corresponding to the specified deformation. The plots are made with 10^6 points per sample.

- CG_20D_corner_def_1: μ-deformation with $\epsilon = 0.5$.
- CG_20D_corner_def_2: Σ_{ii}-deformation with $\epsilon = 0.5$.
- CG_20D_corner_def_3: $\Sigma_{i\neq j}$-deformation with $\epsilon = 0.5$.
- CG_20D_corner_def_4: pow$^+$-deformation with $\epsilon = 0.1$.
- CG_20D_corner_def_5: pow$^-$-deformation with $\epsilon = 0.1$.
- CG_20D_corner_def_6: $\mathcal{N}$-deformation with $\epsilon = 0.5$.
- CG_20D_corner_def_7: $\mathcal{U}$-deformation with $\epsilon = 0.5$.

The following Figures report a color plot representation of the correlation matrices for two samples $\mathbf{X}_1$, drawn from the reference distribution of the CG model with $d = 20$, and $\mathbf{X}_2$, corresponding to the specified deformation. The plots are made with 10^6 points per sample.

- CG_20D_corr_matrix_def_1: μ-deformation with $\epsilon = 0.5$.
- CG_20D_corr_matrix_def_2: Σ_{ii}-deformation with $\epsilon = 0.5$.
- CG_20D_corr_matrix_def_3: $\Sigma_{i\neq j}$-deformation with $\epsilon = 0.5$.

- CG_20D_corr_matrix_def_4: pow$^+$-deformation with $\epsilon = 0.1$.
- CG_20D_corr_matrix_def_5: pow$^-$-deformation with $\epsilon = 0.1$.
- CG_20D_corr_matrix_def_6: $\mathcal{N}$-deformation with $\epsilon = 0.5$.
- CG_20D_corr_matrix_def_7: $\mathcal{U}$-deformation with $\epsilon = 0.5$.

The following Figures show the empirical PDF of the specified test statistic under the null hypothesis (both samples come from the same reference distribution) for the CG model with $d = 20$, and $n = m = (1, 2, 5, 10) \cdot 10^4$. The test statistics distributions are obtained with 10^4 iterations.

- CG_20D_10K_null_SW
  CG_20D_20K_null_SW
  CG_20D_50K_null_SW
  CG_20D_100K_null_SW
- CG_20D_10K_null_SKS
  CG_20D_20K_null_SKS
  CG_20D_50K_null_SKS
  CG_20D_100K_null_SKS
- CG_20D_10K_null_MMD
  CG_20D_20K_null_MMD
  CG_20D_50K_null_MMD
  CG_20D_100K_null_MMD
- CG_20D_10K_null_$\overline{\text{KS}}$
  CG_20D_20K_null_$\overline{\text{KS}}$
  CG_20D_50K_null_$\overline{\text{KS}}$
  CG_20D_100K_null_$\overline{\text{KS}}$
- CG_20D_10K_null_FGD
  CG_20D_20K_null_FGD
  CG_20D_50K_null_FGD
  CG_20D_100K_null_FGD

List of Figures for CG results in $100D$

The following Figures show the 1D and 2D marginal probability distributions for two samples $\mathbf{X}_1$, drawn from the reference distribution of the CG model with $d = 100$, and $\mathbf{X}_2$, corresponding to the specified deformation. The plots are made with 10^6 points per sample.

- CG_100D_corner_def_1: μ-deformation with $\epsilon = 0.5$.
- CG_100D_corner_def_2: Σ_{ii}-deformation with $\epsilon = 0.5$.
- CG_100D_corner_def_3: $\Sigma_{i\neq j}$-deformation with $\epsilon = 0.5$.
- CG_100D_corner_def_4: pow$^+$-deformation with $\epsilon = 0.1$.
- CG_100D_corner_def_5: pow$^-$-deformation with $\epsilon = 0.1$.
- CG_100D_corner_def_6: $\mathcal{N}$-deformation with $\epsilon = 0.5$.
- CG_100D_corner_def_7: $\mathcal{U}$-deformation with $\epsilon = 0.5$.

The following Figures report a color plot representation of the correlation matrices for two samples $\mathbf{X}_1$, drawn from the reference distribution of the CG model with $d = 100$, and $\mathbf{X}_2$, corresponding to the specified deformation. The plots are made with 10^6 points per sample.

- CG_100D_corr_matrix_def_1: μ-deformation with $\epsilon = 0.5$.

- CG_100D_corr_matrix_def_2: Σ_{ii}-deformation with $\epsilon = 0.5$.
- CG_100D_corr_matrix_def_3: $\Sigma_{i\neq j}$-deformation with $\epsilon = 0.5$.
- CG_100D_corr_matrix_def_4: pow$^+$-deformation with $\epsilon = 0.1$.
- CG_100D_corr_matrix_def_5: pow$^-$-deformation with $\epsilon = 0.1$.
- CG_100D_corr_matrix_def_6: $\mathcal{N}$-deformation with $\epsilon = 0.5$.
- CG_100D_corr_matrix_def_7: $\mathcal{U}$-deformation with $\epsilon = 0.5$.

The following Figures show the empirical PDF of the specified test statistic under the null hypothesis (both samples come from the same reference distribution) for the CG model with $d = 100$, and $n = m = (1, 2, 5, 10) \cdot 10^4$. The test statistics distributions are obtained with 10^4 iterations.

- CG_100D_10K_null_SW
  CG_100D_20K_null_SW
  CG_100D_50K_null_SW
  CG_100D_100K_null_SW
- CG_100D_10K_null_SKS
  CG_100D_20K_null_SKS
  CG_100D_50K_null_SKS
  CG_100D_100K_null_SKS
- CG_100D_10K_null_MMD
  CG_100D_20K_null_MMD
  CG_100D_50K_null_MMD
  CG_100D_100K_null_MMD
- CG_100D_10K_null_$\overline{\text{KS}}$
  CG_100D_20K_null_$\overline{\text{KS}}$
  CG_100D_50K_null_$\overline{\text{KS}}$
  CG_100D_100K_null_$\overline{\text{KS}}$
- CG_100D_10K_null_FGD
  CG_100D_20K_null_FGD
  CG_100D_50K_null_FGD
  CG_100D_100K_null_FGD

5.A.2 List of Figures for the MoG models

List of Figures for MoG results in $5D$

The following Figures show the 1D and 2D marginal probability distributions for two samples $\mathbf{X}_1$, drawn from the reference distribution of the MoG model with $d = 5$ and $q = 3$, and $\mathbf{X}_2$, corresponding to the specified deformation. The plots are made with 10^6 points per sample.

- MoG_5D_corner_def_1: μ-deformation with $\epsilon = 0.5$.
- MoG_5D_corner_def_2: Σ_{ii}-deformation with $\epsilon = 0.5$.
- MoG_5D_corner_def_3: $\Sigma_{i\neq j}$-deformation with $\epsilon = 0.5$.
- MoG_5D_corner_def_4: pow$^+$-deformation with $\epsilon = 0.1$.
- MoG_5D_corner_def_5: pow$^-$-deformation with $\epsilon = 0.1$.
- MoG_5D_corner_def_6: $\mathcal{N}$-deformation with $\epsilon = 0.5$.
- MoG_5D_corner_def_7: $\mathcal{U}$-deformation with $\epsilon = 0.5$.

The following Figures report a color plot representation of the correlation matrices for two samples $\mathbf{X}_1$, drawn from the reference distribution of the MoG model with $d = 5$ and $q = 3$, and $\mathbf{X}_2$, corresponding to the specified deformation. The plots are made with 10^6 points per sample.

- MoG_5D_corr_matrix_def_1: μ-deformation with $\epsilon = 0.5$.
- MoG_5D_corr_matrix_def_2: Σ_{ii}-deformation with $\epsilon = 0.5$.
- MoG_5D_corr_matrix_def_3: $\Sigma_{i\neq j}$-deformation with $\epsilon = 0.5$.
- MoG_5D_corr_matrix_def_4: pow$^+$-deformation with $\epsilon = 0.1$.
- MoG_5D_corr_matrix_def_5: pow$^-$-deformation with $\epsilon = 0.1$.
- MoG_5D_corr_matrix_def_6: $\mathcal{N}$-deformation with $\epsilon = 0.5$.
- MoG_5D_corr_matrix_def_7: $\mathcal{U}$-deformation with $\epsilon = 0.5$.

The following Figures show the empirical PDF of the specified test statistic under the null hypothesis (both samples come from the same reference distribution) for the MoG model with $d = 5$, $q = 3$, and $n = m = (1, 2, 5, 10) \cdot 10^4$. The test statistics distributions are obtained with 10^4 iterations.

- MoG_5D_10K_null_SW
  MoG_5D_20K_null_SW
  MoG_5D_50K_null_SW
  MoG_5D_100K_null_SW
- MoG_5D_10K_null_SKS
  MoG_5D_20K_null_SKS
  MoG_5D_50K_null_SKS
  MoG_5D_100K_null_SKS
- MoG_5D_10K_null_MMD
  MoG_5D_20K_null_MMD
  MoG_5D_50K_null_MMD
  MoG_5D_100K_null_MMD
- MoG_5D_10K_null_$\overline{\text{KS}}$
  MoG_5D_20K_null_$\overline{\text{KS}}$
  MoG_5D_50K_null_$\overline{\text{KS}}$
  MoG_5D_100K_null_$\overline{\text{KS}}$
- MoG_5D_10K_null_FGD
  MoG_5D_20K_null_FGD
  MoG_5D_50K_null_FGD
  MoG_5D_100K_null_FGD

List of Figures for MoG results in $20D$

The following Figures show the 1D and 2D marginal probability distributions for two samples $\mathbf{X}_1$, drawn from the reference distribution of the MoG model with $d = 20$ and $q = 5$, and $\mathbf{X}_2$, corresponding to the specified deformation. The plots are made with 10^6 points per sample.

- MoG_20D_corner_def_1: μ-deformation with $\epsilon = 0.5$.
- MoG_20D_corner_def_2: Σ_{ii}-deformation with $\epsilon = 0.5$.
- MoG_20D_corner_def_3: $\Sigma_{i\neq j}$-deformation with $\epsilon = 0.5$.
- MoG_20D_corner_def_4: pow$^+$-deformation with $\epsilon = 0.1$.
- MoG_20D_corner_def_5: pow$^-$-deformation with $\epsilon = 0.1$.

- MoG_20D_corner_def_6: $\mathcal{N}$-deformation with $\epsilon = 0.5$.
- MoG_20D_corner_def_7: $\mathcal{U}$-deformation with $\epsilon = 0.5$.

The following Figures report a color plot representation of the correlation matrices for two samples $\mathbf{X}_1$, drawn from the reference distribution of the MoG model with $d = 20$ and $q = 5$, and $\mathbf{X}_2$, corresponding to the specified deformation. The plots are made with 10^6 points per sample.

- MoG_20D_corr_matrix_def_1: μ-deformation with $\epsilon = 0.5$.
- MoG_20D_corr_matrix_def_2: Σ_{ii}-deformation with $\epsilon = 0.5$.
- MoG_20D_corr_matrix_def_3: $\Sigma_{i\neq j}$-deformation with $\epsilon = 0.5$.
- MoG_20D_corr_matrix_def_4: pow$^+$-deformation with $\epsilon = 0.1$.
- MoG_20D_corr_matrix_def_5: pow$^-$-deformation with $\epsilon = 0.1$.
- MoG_20D_corr_matrix_def_6: $\mathcal{N}$-deformation with $\epsilon = 0.5$.
- MoG_20D_corr_matrix_def_7: $\mathcal{U}$-deformation with $\epsilon = 0.5$.

The following Figures show the empirical PDF of the specified test statistic under the null hypothesis (both samples come from the same reference distribution) for the MoG model with $d = 20$, $q = 5$, and $n = m = (1, 2, 5, 10) \cdot 10^4$. The test statistics distributions are obtained with 10^4 iterations.

- MoG_20D_10K_null_SW
  MoG_20D_20K_null_SW
  MoG_20D_50K_null_SW
  MoG_20D_100K_null_SW
- MoG_20D_10K_null_SKS
  MoG_20D_20K_null_SKS
  MoG_20D_50K_null_SKS
  MoG_20D_100K_null_SKS
- MoG_20D_10K_null_MMD
  MoG_20D_20K_null_MMD
  MoG_20D_50K_null_MMD
  MoG_20D_100K_null_MMD
- MoG_20D_10K_null_$\overline{\text{KS}}$
  MoG_20D_20K_null_$\overline{\text{KS}}$
  MoG_20D_50K_null_$\overline{\text{KS}}$
  MoG_20D_100K_null_$\overline{\text{KS}}$
- MoG_20D_10K_null_FGD
  MoG_20D_20K_null_FGD
  MoG_20D_50K_null_FGD
  MoG_20D_100K_null_FGD

List of Figures for MoG results in $100D$

The following Figures show the 1D and 2D marginal probability distributions for two samples $\mathbf{X}_1$, drawn from the reference distribution of the MoG model with $d = 100$ and $q = 10$, and $\mathbf{X}_2$, corresponding to the specified deformation. The plots are made with 10^6 points per sample.

- MoG_100D_corner_def_1: μ-deformation with $\epsilon = 0.5$.
- MoG_100D_corner_def_2: Σ_{ii}-deformation with $\epsilon = 0.5$.
- MoG_100D_corner_def_3: $\Sigma_{i\neq j}$-deformation with $\epsilon = 0.5$.

- MoG_100D_corner_def_4: pow$^+$-deformation with $\epsilon = 0.1$.
- MoG_100D_corner_def_5: pow$^-$-deformation with $\epsilon = 0.1$.
- MoG_100D_corner_def_6: $\mathcal{N}$-deformation with $\epsilon = 0.5$.
- MoG_100D_corner_def_7: $\mathcal{U}$-deformation with $\epsilon = 0.5$.

The following Figures report a color plot representation of the correlation matrices for two samples $\mathbf{X}_1$, drawn from the reference distribution of the MoG model with $d = 100$ and $q = 10$, and $\mathbf{X}_2$, corresponding to the specified deformation. The plots are made with 10^6 points per sample.

- MoG_100D_corr_matrix_def_1: μ-deformation with $\epsilon = 0.5$.
- MoG_100D_corr_matrix_def_2: Σ_{ii}-deformation with $\epsilon = 0.5$.
- MoG_100D_corr_matrix_def_3: $\Sigma_{i\neq j}$-deformation with $\epsilon = 0.5$.
- MoG_100D_corr_matrix_def_4: pow$^+$-deformation with $\epsilon = 0.1$.
- MoG_100D_corr_matrix_def_5: pow$^-$-deformation with $\epsilon = 0.1$.
- MoG_100D_corr_matrix_def_6: $\mathcal{N}$-deformation with $\epsilon = 0.5$.
- MoG_100D_corr_matrix_def_7: $\mathcal{U}$-deformation with $\epsilon = 0.5$.

The following Figures show the empirical PDF of the specified test statistic under the null hypothesis (both samples come from the same reference distribution) for the MoG model with $d = 100$, $q = 10$, and $n = m = (1, 2, 5, 10) \cdot 10^4$. The test statistics distributions are obtained with 10^4 iterations.

- MoG_100D_10K_null_SW
  MoG_100D_20K_null_SW
  MoG_100D_50K_null_SW
  MoG_100D_100K_null_SW
- MoG_100D_10K_null_SKS
  MoG_100D_20K_null_SKS
  MoG_100D_50K_null_SKS
  MoG_100D_100K_null_SKS
- MoG_100D_10K_null_MMD
  MoG_100D_20K_null_MMD
  MoG_100D_50K_null_MMD
  MoG_100D_100K_null_MMD
- MoG_100D_10K_null_$\overline{\text{KS}}$
  MoG_100D_20K_null_$\overline{\text{KS}}$
  MoG_100D_50K_null_$\overline{\text{KS}}$
  MoG_100D_100K_null_$\overline{\text{KS}}$
- MoG_100D_10K_null_FGD
  MoG_100D_20K_null_FGD
  MoG_100D_50K_null_FGD
  MoG_100D_100K_null_FGD

5.A.3 List of Figures for the JetNet at jet level

The following Figures show the 1D and 2D marginal probability distributions for two samples $\mathbf{X}_1$, extracted from the JetNet reference distribution for jet level features in the gluon dataset, and $\mathbf{X}_2$, corresponding to the specified deformation. The plots are made with 10^6 points per sample.

- jet_corner_def_1: μ-deformation with $\epsilon = 0.5$.

- jet_corner_def_2: Σ_{ii}-deformation with $\epsilon = 0.2$.
- jet_corner_def_3: $\Sigma_{i\neq j}$-deformation with $\epsilon = 0.2$.
- jet_corner_def_4: pow^{+}-deformation with $\epsilon = 0.1$.
- jet_corner_def_5: pow^{-}-deformation with $\epsilon = 0.1$.
- jet_corner_def_6: $\mathcal{N}$-deformation with $\epsilon = 0.2$.
- jet_corner_def_7: $\mathcal{U}$-deformation with $\epsilon = 0.2$.

The following Figures report a color plot representation of the correlation matrices for two samples $\mathbf{X}_1$, extracted from the JetNet reference distribution for jet level features in the gluon dataset, and $\mathbf{X}_2$, corresponding to the specified deformation. The plots are made with 10^6 points per sample.

- jet_corr_matrix_def_1: μ-deformation with $\epsilon = 0.5$.
- jet_corr_matrix_def_2: Σ_{ii}-deformation with $\epsilon = 0.2$.
- jet_corr_matrix_def_3: $\Sigma_{i\neq j}$-deformation with $\epsilon = 0.2$.
- jet_corr_matrix_def_4: pow^{+}-deformation with $\epsilon = 0.1$.
- jet_corr_matrix_def_5: pow^{-}-deformation with $\epsilon = 0.1$.
- jet_corr_matrix_def_6: $\mathcal{N}$-deformation with $\epsilon = 0.2$.
- jet_corr_matrix_def_7: $\mathcal{U}$-deformation with $\epsilon = 0.2$.

The following Figures report a summary of the marginal distributions for two samples $\mathbf{X}_1$, extracted from the JetNet reference distribution for jet level features in the gluon dataset, and $\mathbf{X}_2$, corresponding to the specified deformation. The plots are made with 10^6 points per sample.

- jet_deformation_summary_1
- jet_deformation_summary_2

The following Figures show the empirical PDF of the specified test statistic under the null hypothesis (both samples come from the same reference distribution) for the JetNet distribution for jet level features in the gluon dataset, considering $n = m = (1, 2, 5) \cdot 10^4$. The test statistics distributions are obtained with 10^3 iterations.

- jet_10K_null_FGD
  jet_20K_null_FGD
  jet_50K_null_FGD
- jet_10K_null_KS
  jet_20K_null_KS
  jet_50K_null_KS
- jet_10K_null_MMD
  jet_20K_null_MMD
  jet_50K_null_MMD

- jet_10K_null_SKS
  jet_20K_null_SKS
  jet_50K_null_SKS
- jet_10K_null_SWD
  jet_20K_null_SWD
  jet_50K_null_SWD

5.A.4 List of Figures for the JetNet at particle level

The following Figures report a color plot representation of the correlation matrices for two samples $\mathbf{X}_1$, extracted from the JetNet reference distribution for particle level features including the first 30 particles (ordered in p_T) in the gluon dataset, and $\mathbf{X}_2$, corresponding to the specified deformation. The plots are made with 10^6 points per sample.

- part_30_corr_matrix_def_1: μ-deformation with $\epsilon = 0.5$.
- part_30_corr_matrix_def_2: Σ_{ii}-deformation with $\epsilon = 0.2$.
- part_30_corr_matrix_def_3: $\Sigma_{i\neq j}$-deformation with $\epsilon = 0.2$.
- part_30_corr_matrix_def_4: pow$^+$-deformation with $\epsilon = 0.1$.
- part_30_corr_matrix_def_5: pow$^-$-deformation with $\epsilon = 0.1$.
- part_30_corr_matrix_def_6: $\mathcal{N}$-deformation with $\epsilon = 0.2$.
- part_30_corr_matrix_def_7: $\mathcal{U}$-deformation with $\epsilon = 0.2$.

The following Figures report a summary of the marginal distributions for two samples $\mathbf{X}_1$, extracted from the JetNet reference distribution for particle level features including the first 30 particles (ordered in p_T) in the gluon dataset, and $\mathbf{X}_2$, corresponding to the specified deformation. The plots are made with 10^6 points per sample.

- part_30_deformation_summary_1
- part_30_deformation_summary_2

The following Figures show the empirical PDF of the specified test statistic under the null hypothesis (both samples come from the same reference distribution) for the JetNet distribution for particle level features with 30 particles in the gluon dataset, considering $n = m = (1, 2, 5) \cdot 10^4$. The test statistics distributions are obtained with 10^3 iterations.

- part_30_10K_null_FGD
  part_30_20K_null_FGD
  part_30_50K_null_FGD
- part_30_10K_null_MMD
  part_30_20K_null_MMD
  part_30_50K_null_MMD
- part_30_10K_null_SWD
  part_30_20K_null_SWD
  part_30_50K_null_SWD
- part_30_10K_null_KS
  part_30_20K_null_KS
  part_30_50K_null_KS
- part_30_10K_null_SKS
  part_30_20K_null_SKS
  part_30_50K_null_SKS

5.B List of Tables

5.B.1 List of Tables for the CG models

The following Tables summarize the results for the CG model with $d = 5$, 20, and 100, for different $n = m$ number of samples. The tables show, for the different deformations considered in the CG models, the values of ϵ for which the null hypothesis is rejected at the 95% and 99% confidence levels, and the time in seconds needed to compute them with percent-level error given the null hypothesis distributions. The null hypothesis distribution was evaluated with $n_{\text{iter}} = 10^4$ iterations, while the value of the test statistics under the alternative hypothesis was estimated from the mean of $n_{\text{iter}} = 100$ iterations. The last column shows the computing time of the test statistics distributions under the null hypothesis, which is independent of ϵ for all the test statistics but t_{LLR}.

- CG_5D_10K
  CG_5D_20K
  CG_5D_50K
  CG_5D_100K
- CG_20D_10K
  CG_20D_20K
  CG_20D_50K
  CG_20D_100K
- CG_100D_10K
  CG_100D_20K
  CG_100D_50K
  CG_100D_100K

5.B.2 List of Tables for the MoG models

The following Tables summarize the results for the MoG model with $d = 5$, $q = 3$, $d = 20$, $q = 5$, and $d = 100$, $q = 10$, for different $n = m$ number of samples. The tables show, for the different deformations considered in the MoG models, the values of ϵ for which the null hypothesis is rejected at the 95% and 99% confidence levels, and the time in seconds needed to compute them with percent-level error given the null hypothesis distributions. The null hypothesis distribution was evaluated with $n_{\text{iter}} = 10^4$ iterations, while the value of the test statistics under the alternative hypothesis was estimated from the mean of $n_{\text{iter}} = 100$ iterations. The last column shows the computing time of the test statistics distributions under the null hypothesis, which is independent of ϵ for all the test statistics but t_{LLR}.

- MoG_5D_10K
  MoG_5D_20K
  MoG_5D_50K
  MoG_5D_100K
- MoG_20D_10K
  MoG_20D_20K
  MoG_20D_50K
  MoG_20D_100K
- MoG_100D_10K
  MoG_100D_20K
  MoG_100D_50K
  MoG_100D_100K

5.B.3 List of Tables for the JetNet at jet level

The following Tables summarize the results for the JetNet gluon dataset at the jet level for different $n = m$ number of samples. The tables show, for the different deformations of the JetNet reference distribution for jet level features in the gluon dataset, the values of ϵ for which the null hypothesis is rejected at the 95% and 99% confidence levels, and the time in seconds needed to compute them with percent-level error given the null hypothesis distributions. The null hypothesis distribution

was evaluated with $n_{\text{iter}} = 10^3$ iterations, while the value of the test statistics under the alternative hypothesis was estimated from the mean of $n_{\text{iter}} = 100$ iterations. The last column shows the computing time of the test statistics distributions under the null hypothesis, which is independent of ϵ for all the test statistics but t_{LLR}. To better understand the effect of scaling features (to zero mean and unit standard deviation) on the performance of the tests, we also show the results for the same deformations but with the test performed on scaled features. The latter results are denoted with a suffix 'scaled'.

- jet_10K
  jet_20K
  jet_50K
- jet_10K_scaled
  jet_20K_scaled
  jet_50K_scaled

5.B.4 List of Tables for the JetNet at particle level

The following Tables summarize the results for the JetNet gluon dataset at the particle level, including the first 30 particles (ordered in p_T), for different $n = m$ number of samples. The tables show, for the different deformations of the JetNet reference distribution for particle level features in the gluon dataset, the values of ϵ for which the null hypothesis is rejected at the 95% and 99% confidence levels, and the time in seconds needed to compute them with percent-level error given the null hypothesis distributions. The null hypothesis distribution was evaluated with $n_{\text{iter}} = 10^3$ iterations, while the value of the test statistics under the alternative hypothesis was estimated from the mean of $n_{\text{iter}} = 100$ iterations. The last column shows the computing time of the test statistics distributions under the null hypothesis, which is independent of ϵ for all the test statistics but t_{LLR}. To better understand the effect of scaling features (to zero mean and unit standard deviation) on the performance of the tests, we also show the results for the same deformations but with the test performed on scaled features. The latter results are denoted with a suffix 'scaled'.

- part_30_10K
  part_30_20K
  part_30_50K
- part_30_10K_scaled
  part_30_20K_scaled
  part_30_50K_scaled

Chapter 6

Comparing Generative Models with the New Physics Learning Machine

In the previous Chapter, we analysed the performances of various test-statistics based on 1D-distances in discriminating samples drawn from a reference distribution from samples drawn from a deformed distribution. In this Chapter, we aim to deploy the same framework to assess the performance of the New Physics Learning Machine (NPLM), a classifier-based methodology designed to perform a data-driven likelihood-ratio (LR) test, primarily for HEP data. A potential drawback of learning-based methods is the introduction of a training step that inevitably impacts the efficiency of the test in terms of computational time. It then becomes relevant to assess the tradeoff between sensitivity and efficiency and determine whether non-learning methodologies retain an advantage, especially when generative models have typically not yet reached a high degree of fidelity by the standards of precision sciences. Moreover, machine learning methods generally require a model selection phase that could further affect their overall efficiency.

For these reasons, we aim to test the NPLM performances against the 1D-based metrics, the unbiased Fréchet Gaussian Distance and the unbiased quadratic Maximum Mean Discrepancy studied in the previous Chapter. While NPLM has been primarily tested on relatively low-dimensional problems (around ten features), as is common in many HEP applications, we also evaluate its performance in more challenging regimes with dimensionalities up to $d = 100$. We adopt here balanced samples from the reference and generative distributions to enable a clean and controlled comparison across evaluation methods.

This Chapter is organized as follows. In Section 6.1, we briefly review the framework and methodology introduced in Chapter 5, which are adopted here without modification. In Section 6.2 we provide a technical description of the NPLM, focusing on its construction and on how it approximates the LR. Section 6.3 is structured in several subsections. Firstly, we recall what dataset is used to perform the analysis; next, we motivate the choice of models considered in this study; we then outline the construction of the null hypothesis and summarize the results; finally, we detail the computational resources employed. In Section 6.4 we draw our conclusions.

Code and full results are available on GitHub in [190–192].

6.1 Comparing generators through two-sample tests

In this Section, we briefly summarize the methodology for comparing generators via two-sample testing, following the framework introduced in [3].

Let $\mathcal{G}_p$ denote a reference generator producing samples from a known probability density function (pdf) p. To test whether another generator $\mathcal{G}'_q$ produces samples consistent with $\mathcal{G}_p$, we compare two datasets $\mathcal{X} = \{x_i\}_{i=1}^n$ and $\mathcal{Y} = \{y_j\}_{j=1}^m$, produced by the two generators, using a two-sample test. The null hypothesis H_0 states that both datasets are generated by the reference generator $\mathcal{G}_p$.

A test statistic t for a two-sample test is defined as a scalar-valued function that maps two samples of sizes n and m in d dimensions into a real number:

$$t : \mathbb{R}^{n\times d} \times \mathbb{R}^{m\times d} \to \mathbb{R}. \tag{6.1}$$

Its distribution under H_0 is estimated empirically by computing t over multiple pairs of samples independently generated by $\mathcal{G}_p$. This results in a set of values of t_0 from which we construct the empirical pdf $f(t_0)$ and cumulative distribution function (cdf) $F(t_0)$. These define the significance threshold t_0^α corresponding to a fixed significance $1-\alpha$:

$$\alpha = \int_{t_0^\alpha}^{\infty} dF(t_0) = \int_{t_0^\alpha}^{\infty} f(t_0)dt_0, \tag{6.2}$$

which we estimate empirically as:

$$\begin{aligned} 1-\alpha = F(t_0^\alpha) &\approx 1 - \frac{\#\text{ of tests with } t_0 \geq t_0^\alpha}{\text{total }\#\text{ of tests}} \\ &= \frac{\#\text{ of tests with } t_0 < t_0^\alpha}{\text{total }\#\text{ of tests}}. \end{aligned} \tag{6.3}$$

In this work, we consider $\alpha = 0.05$ and 0.01.

To assess the sensitivity of a test statistic to deviations from p, we define alternative generators $\mathcal{G}'_{q_\epsilon}$, obtained by deforming the pdf p through a scalar parameter ϵ, namely

$$q_\epsilon \xrightarrow{\epsilon\to 0} p.$$

These ϵ-deformations are designed to systematically test the sensitivity of each test statistic, and are listed in 6.A. The corresponding alternative hypothesis H_1 states that $\mathcal{X} \sim p^n$ and $\mathcal{Y} \sim q_\epsilon^m$ are generated by different generators ($\epsilon \neq 0$).

The critical deformation ϵ_α is defined as the smallest $\epsilon \geq 0$ that can be "detected" (leading to rejection of the null-hypothesis) by a given test with a significance level $1-\alpha$. This corresponds to the smallest ϵ-deformation for which the test statistic exceeds the threshold t_0^α and can be formulated

as the following optimization problem:

$$\epsilon_\alpha = \arg\min_\epsilon |t(\epsilon) - t_0^\alpha|. \tag{6.4}$$

We solve this numerically using a simple bisection method. At each step, the test is repeated 100 times to estimate the mean $\mu_{t(\epsilon)}$ and standard deviation $\sigma_{t(\epsilon)}$ of the test statistic. The interval for ϵ is iteratively refined by comparing $\mu_{t(\epsilon)} \pm \sigma_{t(\epsilon)}$ to t_0^α, until convergence is reached within a fixed tolerance of 10^{-2}. The central value defines ϵ_α, with bounds $\epsilon_{\alpha\text{-low}}$ and $\epsilon_{\alpha\text{-up}}$ determined by where the uncertainty bands meet the threshold. In formulae, this is expressed as:

$$\begin{aligned}
\epsilon_\alpha &= \arg\min_\epsilon |\mu_{t(\epsilon)} - t_0^\alpha|, \\
\epsilon_{\alpha\text{-low}} &= \arg\min_\epsilon |(\mu_{t(\epsilon)} + \sigma_{t(\epsilon)}) - t_0^\alpha|, \\
\epsilon_{\alpha\text{-up}} &= \arg\min_\epsilon |(\mu_{t(\epsilon)} - \sigma_{t(\epsilon)}) - t_0^\alpha|.
\end{aligned} \tag{6.5}$$

This framework applies to non-parametric test statistics, which do not require explicit knowledge of p or q_ϵ. However, we also include the log-likelihood ratio (LLR) test, which requires both pdfs to be known. Its test statistic is defined as:

$$t(\epsilon) = -2 \sum_{y \in \mathcal{Y}} \log \frac{p(y)}{q_\epsilon(y)}. \tag{6.6}$$

This, by construction, only depends on the samples $\mathcal{Y}$ which are generated by $\mathcal{G}_p$ in the case of the null hypothesis H_0, and by $\mathcal{G}'_{q_\epsilon}$ in the case of the alternative hypothesis H_1. Moreover, since this test statistic depends explicitly on q_ϵ, the null distribution $f(t_0)$ becomes ϵ-dependent. Accordingly, the optimization problem becomes:

$$\epsilon_\alpha = \arg\min_\epsilon |t(\epsilon) - t_0^\alpha(\epsilon)|, \tag{6.7}$$

where $t_0^\alpha(\epsilon)$ must be computed for each value of ϵ. When applicable, the LLR provides the most powerful test according to the Neyman–Pearson lemma ([193]).

Finally, in many practical applications the generators $\mathcal{G}$ may not provide a closed analytical form for the pdf. In such cases, only a finite number of samples is available, and the LLR test is not applicable. The distribution $f(t_0)$ must then be estimated empirically by using a bootstrap approach (namely, sampling with replacement) to mimic the behavior of $\mathcal{G}_p$ and $\mathcal{G}'_{q_\epsilon}$.

6.2 The NPLM method

NPLM is a machine learning-based, signal-agnostic hypothesis testing approach designed on the basis of the maximum likelihood-ratio test as formulated by [193]. Originally developed for the discovery of new physics in high-energy collider experiments such as the LHC ([153]), we consider it here as a general-purpose testing methodology for comparing data generators.

At its core, the NPLM method leverages the ability of classifiers to estimate the ratio of data-generating pdfs (see e.g. [194]). Adopting the notation from the previous Section, a classifier is trained to approximate the following function[1]

$$f_{\hat{w}}(z) \approx \log \frac{q(z)}{p(z)}, \tag{6.8}$$

where $\hat{w}$ are the model parameters obtained at the end of training. The model is then evaluated in-sample on the full dataset using the metric

$$t_{\text{NPLM}}(\mathcal{X}, \mathcal{Y}) = -2 \left[\frac{m}{n} \sum_{z \in \mathcal{X}} \left(e^{f_{\hat{w}}(z)} - 1 \right) - \sum_{z \in \mathcal{Y}} f_{\hat{w}}(z) \right], \tag{6.9}$$

which represents a Monte Carlo-based formulation of the extended likelihood ratio (see [153, 156, 195] and 6.B). The NPLM method operates as a two-sample test, taking $\mathcal{X}$ and $\mathcal{Y}$ as inputs and returning a scalar value in accordance with Eq. (6.1). It then integrates naturally within the framework introduced in [3] to assess its performance against other approaches. Due to its training efficiency, we employ the implementation presented in [156], in which the learning model is based on kernel methods and the classifier spans a parameterized function space $\mathcal{F} = \{f_w\}$, defined as a weighted sum of Gaussian kernels:

$$f_w(z) = \sum_{i=1}^{n+m} w_i k_\sigma(z, z_i), \quad k_\sigma(z, z') = \exp\left(-\frac{\|z - z'\|^2}{2\sigma^2} \right), \tag{6.10}$$

where the kernel width σ is treated as a hyperparameter. The loss function is the weighted binary cross-entropy loss:

$$\ell(c, f_w(z)) = (1 - c)\frac{m}{n} \log\left(1 + e^{f_w(z)}\right) + c \log\left(1 + e^{-f_w(z)}\right), \tag{6.11}$$

where $\mathcal{Z} = \{z_i\}_{i=1}^{n+m} = \{x_1, \ldots, x_n, y_1, \ldots, y_m\}$ and class labels are defined as $c = 0$ for $z \in \mathcal{X}$ and $c = 1$ for $z \in \mathcal{Y}$. The model is trained to minimize the empirical risk:

$$L(f_w) = \frac{1}{n+m} \sum_{i=1}^{n+m} \ell(c_i, f_w(z_i)) + \lambda R(f_w), \tag{6.12}$$

where $R(f_w)$ is the analogue of L^2 regularization in the context of kernel methods ([196]).

Despite its effectiveness, this approach can have high computational costs when the sample size is large. To mitigate this issue, the authors of [156] rely on Falkon ([197]), a modern solver for

[1] Because the original NPLM framework was developed for LHC analyses, these densities are typically normalized to different expected event counts. This distinction is not relevant for our discussion.

large-scale kernel methods. Falkon replaces Eq. (6.10) with:

$$f_w(z) = \sum_{i=1}^{M} w_i k_\sigma(z, \tilde{z}_i), \tag{6.13}$$

where $\{\tilde{z}_1, \ldots, \tilde{z}_M\}$, known as Nyström centers, are sampled uniformly at random from the full dataset. The parameter M is a tunable hyperparameter.

6.2.1 Hyperparameter Tuning

The first step in the NPLM methodology is hyperparameter tuning. The kernel-based implementation of NPLM involves three primary hyperparameters: the kernel width σ, the regularization parameter λ, and the number of centers M. These are tuned exclusively using reference data, ensuring an unbiased estimate of the test statistic's distribution under the null hypothesis. Once this distribution is estimated, it can be used to evaluate samples generated by any model trained on data from the same reference distribution.[2]Following [156], the selection criteria are:

- The Gaussian kernel width σ is set to the 90th percentile of the pairwise distances among reference-distributed data points. Heuristics of this type are commonly used in kernel methods ([172]).
- The regularization parameter λ is chosen to be as small as possible, subject to computational constraints and while ensuring stable training dynamics ([198]).
- The number of Nyström centers M should be at least of order $\sqrt{n+m}$ ([198]). Larger values of M improve performance but increase computational costs, both in terms of training time and memory. For small M, the test statistic increases with M until it reaches a plateau. We therefore require that the average value of the test statistic — computed over a small number of tests on reference data (i.e., under H_0) — be approximately stable as a function of M.

We will show the results of this pipeline on our data in the next Section.

6.3 Numerical analysis

In this Section, we present the results of the analysis based on [3], and summarized in Section 6.1, when applied to the NPLM method. We start by summarizing the properties of the datasets we consider (see [3] for a in-depth description) and we then detail the hyperparameter tuning of the NPLM method. This is an important point of deviation with respect to standard testing methodologies.

[2]This procedure assumes a certain availability of reference data. If data is scarce, it is possible to rely on bootstrapping methods.

	Mixture of Gaussians		
n \ d	5	20	100
$10K$	$(3.88, 10000, 10^{-8})$	$(5.98, 13500, 10^{-7})$	$(10.31, 16000, 10^{-5})$
$20K$	$(3.88, 7000, 10^{-8})$	$(5.98, 12500, 10^{-7})$	$(10.31, 16000, 10^{-5})$
$50K$	$(3.88, 7000, 10^{-8})$	$(5.98, 12500, 10^{-7})$	$(10.31, 16000, 10^{-5})$
$100K$	$(3.88, 5000, 10^{-8})$	$(5.98, 11000, 10^{-6})$	$(10.31, 11000, 10^{-5})$

Table 6.1: Values of the hyperparameters (σ, M, λ) for the MoG distributions for the different dimensionalities d and sample sizes n.

6.3.1 Data

Mixtures of Gaussians (MoG) We consider mixtures of q multivariate Gaussians (q components) in d dimensions, each with diagonal covariance matrices. This framework enables the study of probability density functions with multiple local maxima, which manifest as multiple peaks in the marginal distributions. For our analysis, we examine three MoG configurations: $q = 3$ components in $d = 5$ dimensions, $q = 5$ components in $d = 20$ dimensions, and $q = 10$ components in $d = 100$ dimensions.

Correlated Gaussians These are correlated d-dimensional unimodal Gaussian distributions. In our analysis, we consider $d = 5, 20, 100$.

JetNet To explore a scenario relevant for HEP, we also consider a dataset of simulated gluon jets from the JetNet dataset. We examine two complementary data representations: a particle-level dataset (90 dimensions) that includes features of individual particles within each jet, and a jet-level dataset (3 dimensions) that captures only high-level jet characteristics.

6.3.2 Hyperparameter tuning

The choice of hyperparameters determines the complexity of the learning model. In this work, we perform model selection aiming for a reasonable trade-off between model complexity and computational efficiency, and we do not target the most complex model that can be computed given available hardware resources.

We start by considering the kernel bandwidth σ. As discussed in the NPLM literature, the distribution of pairwise Euclidean distances between reference-distributed points is considered as a method to estimate the most relevant scales in the problem, and it was observed that selecting σ as the 90th percentile of this distribution allows one to obtain a statistical test that has reasonably homogeneous sensitivity over a wide range of potential anomalies. However, this distribution can in general be multimodal, signaling that the reference distribution is characterized by more than one scale. We then select this particular hyperparameter relying on a straightforward modification of the methodology outlined in the previous Section. If the distribution of pairwise distances has

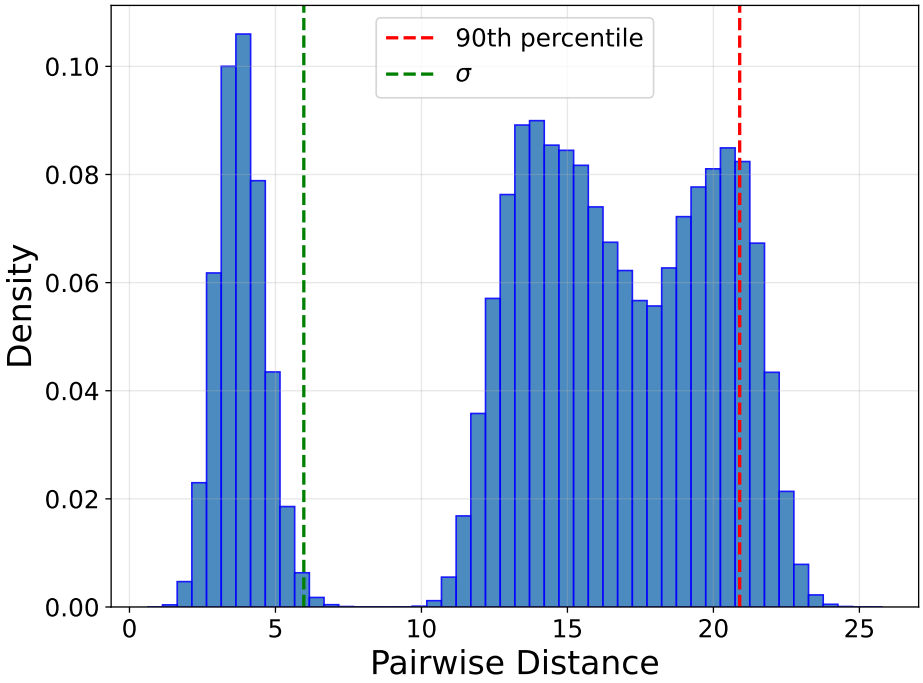


Figure 6.1: Example of a multimodal distribution of the pairwise distance from the MoG model in 20 dimensions.

multiple peaks, as in the case of data from the MoG models, we focus on the first peak (the one at the smallest value of pairwise distance). In particular, we take a value lying on the right tail of the peak. This choice preserves the rationale of the standard single-mode approach while avoiding excessive smoothing that could hide localized, sharp features by selecting an overly large bandwidth. We do not fine-tune this hyperparameter further, as its exact value is not expected to meaningfully affect the sensitivity of the test. This approach is illustrated in Figure 6.1 for the MoG in 20 dimensions. [3]

To explore the space of M and λ values, we compute the average NPLM test statistic over 20 randomly selected pairs of samples drawn from the reference distribution. We first estimate the impact of λ on the training time. The parameter tuning for this analysis has been performed selecting $M = 1000, 2000, 3000, 5000$ for $n = 10K, 20K, 50K, 100K$. We report in Figure 6.2 an example of the computing time as a function of λ for the Mixture of Gaussians (MoG) model with $d = 20$ and $n = 50K$, an intermediate representative case. The plot shows an exponential behavior, with a significant increase starting at $\lambda = 10^{-9}$. Based on this trend, reasonable values for λ lie between 10^{-6} and 10^{-8}. Once an appropriate range for λ is chosen, we vary M. In Figure 6.3 we show, on the same data, the dependence of the computing time and the average test statistic on M using $\lambda = (10^{-6}, 10^{-7}, 10^{-8})$. As expected, computation time increases with M and decreases with λ. Moreover, the average test statistic stabilizes with a large enough M. However, the plateau

[3] See also [199] for a more refined, but with higher computational costs, approach to this problem that can also be applied to other statistical tests characterized by tunable hyperparameters.

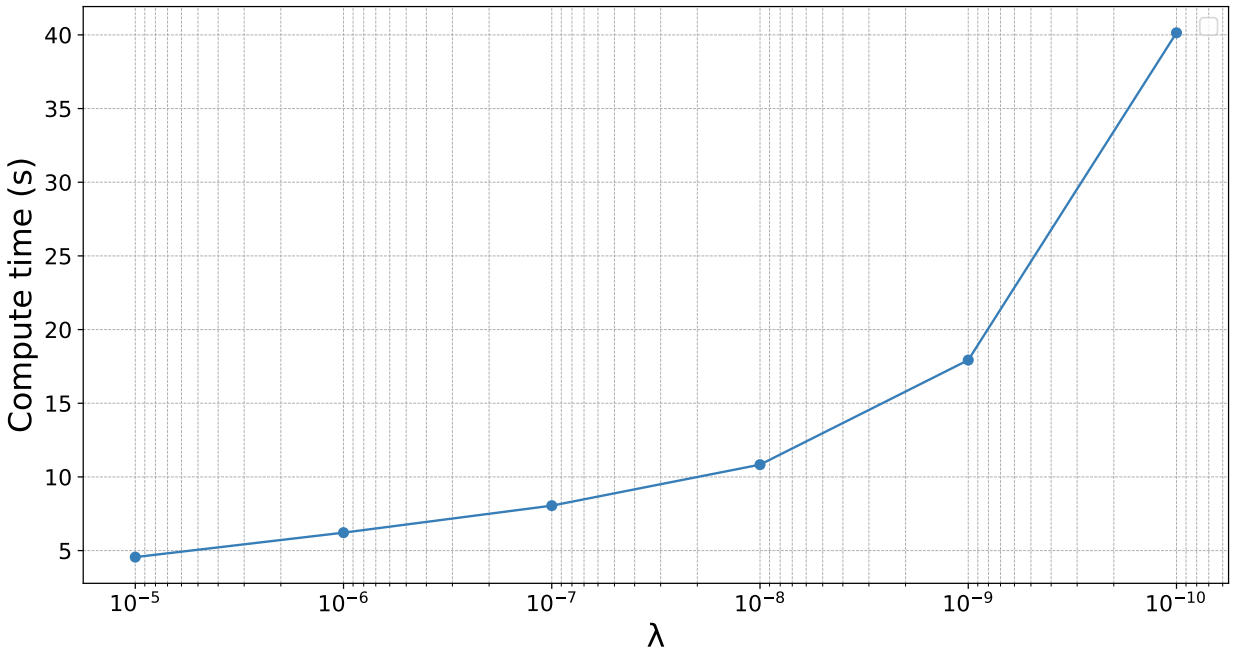


Figure 6.2: Mean compute time over 20 tests as a function of λ for the MoG model with $d = 20$ and $n = 50K$.

	Correlated Gaussians		
n \ d	5	20	100
$10K$	$(9.8, 10000, 10^{-8})$	$(21.8, 13500, 10^{-7})$	$(52.6, 16000, 10^{-5})$
$20K$	$(9.8, 7000, 10^{-8})$	$(21.8, 12500, 10^{-7})$	$(52.6, 16000, 10^{-5})$
$50K$	$(9.8, 7000, 10^{-8})$	$(21.8, 12000, 10^{-7})$	$(52.6, 16000, 10^{-5})$
$100K$	$(9.8, 5000, 10^{-8})$	$(21.8, 11000, 10^{-6})$	$(52.6, 14000, 10^{-5})$

Table 6.2: Values of the hyperparameters (σ, M, λ) for the CG distributions for the different dimensionalities d and sample sizes n.

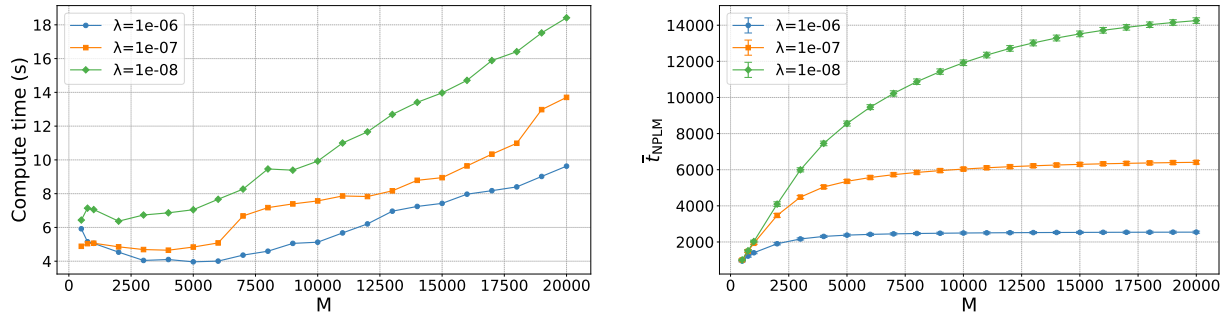


Figure 6.3: Compute time and mean test statistic as functions of M at varying λ for the MoG model with $d = 20$ and $n = 50K$.

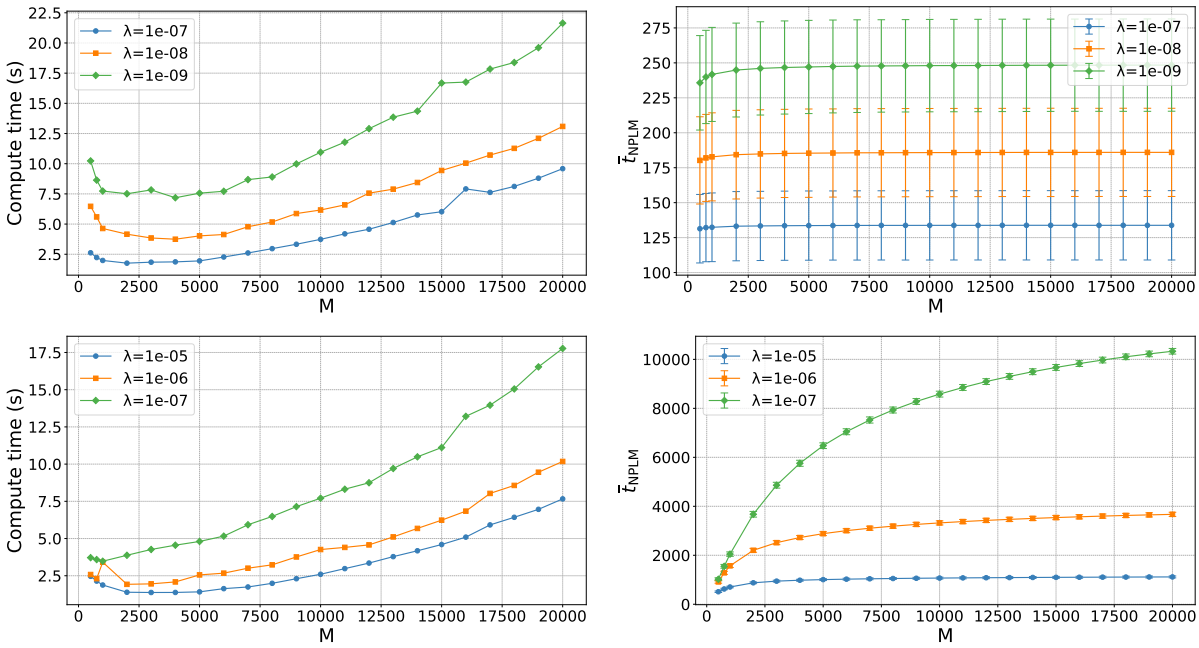


Figure 6.4: Compute time and mean test statistic as functions of M at varying λ for the JetNet dataset. Top row: jet-level features. Bottom row: particle-level features.

n	Jet features	Particle features
$10K$	$(4.2, 10000, 10^{-8})$	$(18.2, 20000, 10^{-7})$
$20K$	$(4.2, 7000, 10^{-8})$	$(18.2, 20000, 10^{-7})$
$50K$	$(4.2, 7000, 10^{-8})$	$(18.2, 20000, 10^{-7})$

Table 6.3: Values of the hyperparameters (σ, M, λ) for the jet-level and particle-level datasets from the JetNet dataset, for the different sample sizes n.

is reached at higher values as λ decreases. In this specific case, we select $(M, \lambda) = (12000, 10^{-7})$, a choice that balances computational efficiency and stability. Note how, by choosing $\lambda = 10^{-6}$, the plateau is reached earlier in M but the resulting learning model would be more regularized, hence less expressive. This procedure has been applied for all the datasets, namely all the other MoG distributions with different dimensionalities and samples sizes, the CG distributions and the particle-level and jet-level datasets from the JetNet dataset. Ultimately, we find that in most cases the hyperparameters selected for the MoG data were suitable for the CG data, at fixed d and n, as shown in Tables 6.1 and 6.2. The number of Nyström centers decreases with n to maintain a reasonable average computation time. To ensure to be in the plateau of the test statistic as a function of M, it is at times needed to increase λ, given our computational constraints. In table 6.3 we report the hyper-parameters selected for the jet-level and particle-level datasets from the JetNet dataset. For completeness we report in Figure 6.4 the plots which validates the chosen values for the JetNet datasets for the case with $n = 20K$. The whole set of plots for each distribution and dataset can be found in [192]. Overall, we find that model selection is not computationally cheap as

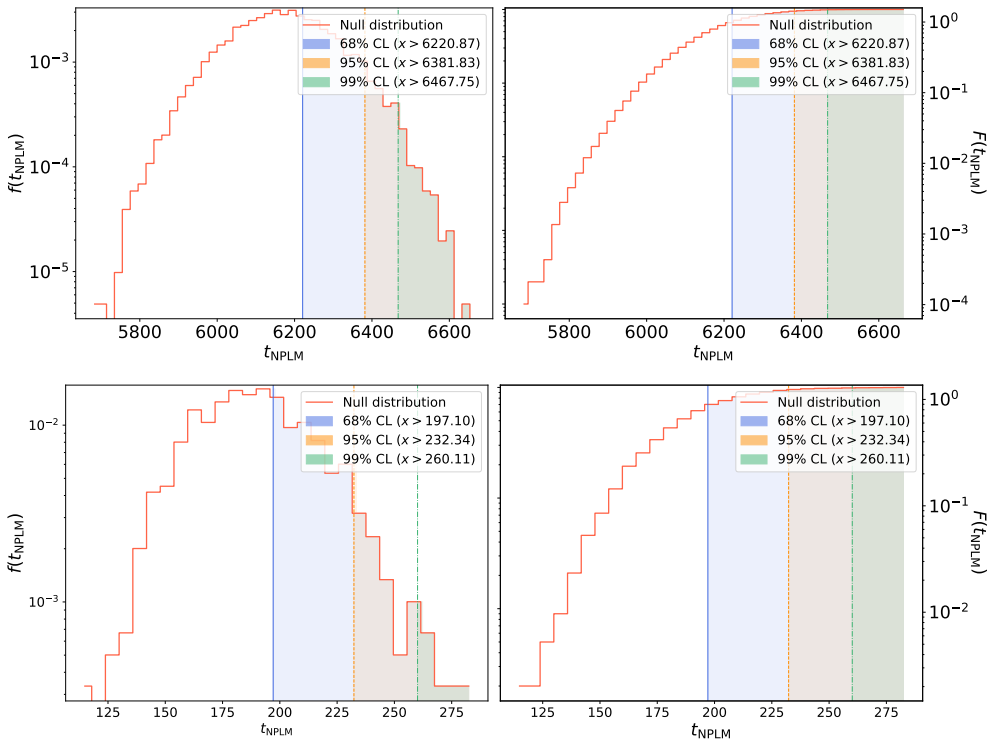


Figure 6.5: Estimated pdf and cdf of the NPLM test statistic under the null hypothesis for the MoG model with $d = 20$ and $n = 50K$ ($10k$ points) and the JetNet (jet features) dataset with $n = 20K$ ($1k$ points).

Data	t_{SW}	$t_{\overline{\text{KS}}}$	t_{SKS}	t_{FGD}	t_{MMD}	t_{NPLM}
CG ($d = 20, n = 20k$)	0.028	**0.011**	0.040	0.12	0.23	4.4
MoG ($d = 20, n = 50k$)	0.034	**0.016**	0.051	0.28	1.4	6.2
JetNet (jet featues, $n = 20k$)	**0.27**	0.32	0.87	0.50	0.95	4.5
JetNet (particle featues, $n = 20k$)	**0.27**	0.35	0.86	1.3	0.86	18.2

Table 6.4: Mean evaluation time in seconds for a single test (null hypothesis).

it requires multiple evaluations of the test. This needs to be taken into account when deciding which testing method is most suitable for the specific use case. However, datasets characterized by specific dimensionalities and sizes yield models with similar hyperparameters. This suggests that the search for optimal hyperparameters can leverage prior studies, hence mitigating its computational impact.

6.3.3 Null hypothesis

Once model selection is performed, we construct the distribution of the test statistic under the null hypothesis H_0, denoted by $f(t_0)$, by evaluating the test on pairs of samples randomly drawn from the reference distribution. Specifically, we perform 10,000 tests for the Mixture of Gaussians and Correlated Gaussians distributions, and 1,000 for the jet and particle-level features from the JetNet

dataset. Once $f(t_0)$ is estimated, we compute the values of the test statistic corresponding to 5% and 1% of false positive rates. Figure 6.5 shows, as an example, both the pdf $f(t_0)$ and the cdf $F(t_0)$ for the MoG model with dimensionality $d = 20$ and sample size $n = 50K$. The figure also highlights the 32%, 5%, and 1% thresholds.

6.3.4 Results

The performance of the NPLM test is compared against the metrics considered in [3], namely the sliced Wasserstein distance (SW), the Kolmogorov-Smirnov test averaged over marginals ($\overline{KS}$), the sliced Kolmogorov-Smirnov test (SKS), the Fréchet Gaussian Distance (FGD), and the Maximum Mean Discrepancy (MMD). Figure 6.6 shows the behavior of the test statistics as functions of the deformations on selected cases for all the datasets. The corresponding numerical values and the complete set of results can be found in 6.C and in [190, 191] respectively. When available, the exact likelihood ratio test statistic is also reported as a reference, being the most powerful test according to the Neyman-Pearson lemma. We report in Table 6.4 the average time per evaluation for the null hypothesis.

Overall, NPLM is quite robust, with no specific failure cases. It typically ranks as the best or second-best performing metric, within uncertainties. It should be noted that, in this setup, no test is guaranteed to be the most powerful against all alternative hypotheses, and it is always possible for certain tests to outperform others in specific scenarios. In general, it is observed that the performance of the NPLM test is higher at low to intermediate dimensionalities ($d \leq 20$) and improves with increasing sample size. This behaviour is expected, as the performance of kernel methods tends to degrade in high-dimensional settings due to the curse of dimensionality and their limited ability, in their simplest forms, to perform feature learning and extract relevant high-level features for distinguishing between two samples (see however [200] for recent advances using pretrained deep networks). The advantages of NPLM are also evident when discrepancies involve the correlation structure, this is natural as NPLM is a natively multivariate ML-based approach. This is in contrast to the $\overline{KS}$ test, which is fast and performs well in most cases, but is totally insensitive to the $\Sigma_{i\neq j}$ deformation, as discussed in [3]. All other metrics exhibit similar levels of performance, with FGD being slightly more consistent. However, both MMD and FGD are more computationally expensive to evaluate.

The relatively high computational cost of NPLM makes it better suited to applications that do not require ultra-fast evaluations, such as offline data analyses, where increased model complexity can be traded for longer training times. Nevertheless, model selection within the NPLM framework can be leveraged to prioritize computational efficiency, as demonstrated in Ref. [201]. Figure 6.7 illustrates the performance of NPLM on the JetNet dataset using a strongly regularized model designed for fast training. Specifically, we consider ($M = 2000, \lambda = 10^{-4}$) for jet-level features and ($M = 5000, \lambda = 10^{-5}$) for particle-level features, resulting in training and evaluation times of 0.2 seconds and 0.7 seconds, respectively, to be compared with Table 6.4. Despite the reduced

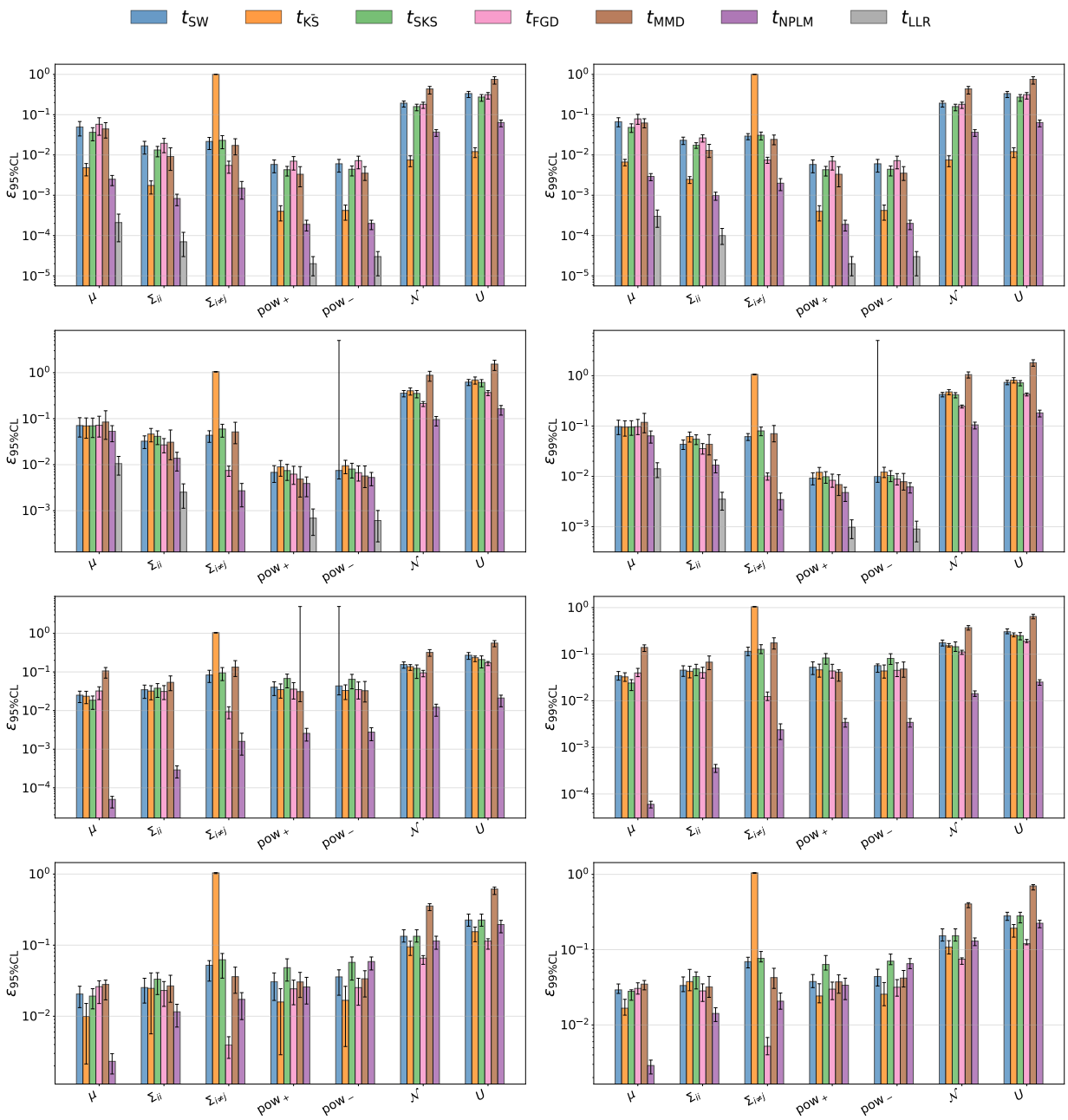


Figure 6.6: Upper bounds and associated uncertainties on ϵ at 95% and 99% confidence levels, computed for different metrics and deformations.
First row: MoG ($d = 20$, $n = 50k$). Second row: CG ($d = 20$, $n = 20k$). Third row: JetNet (jet features, $n = 20k$. Fourth row: JetNet (particle features, $n = 20k$).

computational cost, the performance remains comparable to, or better than, that of alternative methods. Finally, we note that hyperparameter tuning can influence not only the overall sensitivity of the test but also its sensitivity to specific alternatives, as also illustrated in Figure 6.8.[4] In 6.D we report a small study on the performance of the NPLM test in terms of the parameter ϵ as a

[4]See Ref. [199] for a detailed discussion of this effect and possible mitigation strategies.

function of the evaluation time for the JetNet data with $n = m = 2 \cdot 10^4$.

6.3.5 Computational resources

All methods considered in this work are publicly available and have been developed to run on both CPUs and GPUs, with implementations hosted on GitHub ([174]). The NPLM method is built on Falkon [197], a modern solver for large-scale kernel methods. All numerical experiments were performed on NVIDIA A100 GPUs within the Leonardo supercomputer (Cineca).[5]

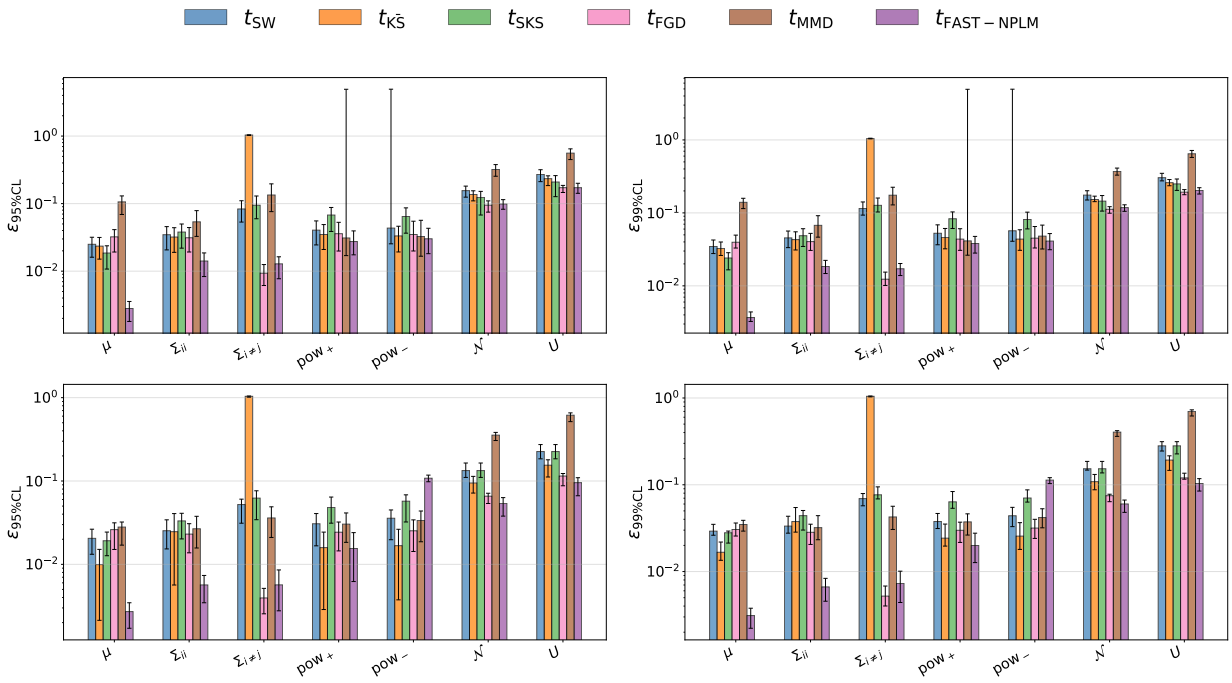


Figure 6.7: Upper bounds and associated uncertainties on ϵ at 95% and 99% confidence levels, computed for different metrics and deformations. Upper row: JetNet (jet features, $n = 20k$); the NPLM method is trained with the following hyperparameters to prioritize fast training and evaluation: $M = 2000$, $\lambda = 10^{-4}$. Bottom row: JetNet (particle features, $n = 20k$); NPLM hyperparameters: $M = 5000$, $\lambda = 10^{-5}$.

6.4 Conclusions

In this work, we tested the NPLM method as a general-purpose two-sample test against several alternatives from the literature, following the framework introduced in [3]. We find that NPLM is overall the most robust approach. The method consistently ranks as either the best- or second-best-performing metric across all tested scenarios, and our study does not identify specific failure cases. This is particularly relevant, as we expect this advantage to become more pronounced in real-world applications, where potential discrepancies may manifest in multiple simultaneous ways. We find

[5] `https://leonardo-supercomputer.cineca.eu/`

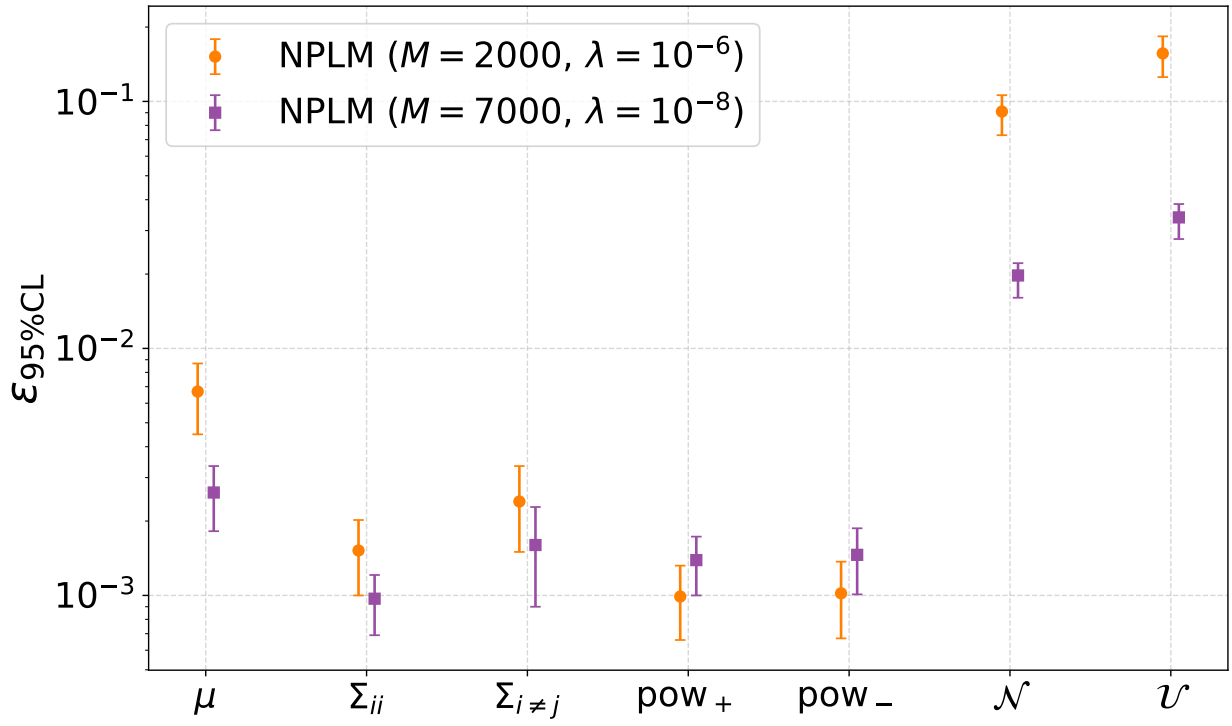


Figure 6.8: Two NPLM models with different hyperparameters on the same data (MoG, $d = 5$, $n = 20K$). Lower is better.

that NPLM is often the most effective solution for identifying discrepancies in the correlation structure, where it significantly outperforms methods based on marginal comparisons. Its performance on the JetNet dataset (jet features) further demonstrates its effectiveness on real-world data, achieving sensitivity improvements of up to two orders of magnitude over competing methods. However, in higher-dimensional problems, the differences with other approaches are less pronounced, likely due to the specific kernel-based implementation and the impact of the curse of dimensionality (see, however, [200] for recent advances using pretrained networks).

This performance, however, comes at the expense of higher computational costs. This cost can be separated into two components: the cost of performing model selection and the cost of executing the test itself. The first involves running a non-negligible number of studies on reference data and is necessary to allow the user to select the optimal working point in the hyperparameter space. This search should be guided by the specific use case. For example, if the goal is to conduct a final comparison of several pre-trained generative models, for which fast validation is not critical, then a model that prioritizes performance over efficiency is preferable (see for example the case of the Challenge for Fast Calorimeter Simulation [135]). However, hyperparameter tuning makes it possible to select a different working point that trades some performance for improved efficiency. As noted in our study, the hyperparameters selected for datasets with similar dimensionalities and sample sizes tend to be close. This suggests that prior studies can be leveraged to reduce the cost

of this step. Conversely, if fast evaluation is essential, such as during exploration of the architecture and hyperparameters of the generative model itself, then other approaches such as the $\overline{KS}$ test, the SW test, or the FGD are more appropriate options. Similar considerations apply to the evaluation time of the test, for which NPLM can be between one and three orders of magnitude slower than the fastest approach, depending on the dataset. It is important to note that hyperparameter tuning needs to be performed only once, using the reference data. This step is completely independent of the training of the generative model. Once the optimal hyperparameters have been determined, and the distribution of the test under the null hypothesis has been estimated, only a single evaluation of the test is required for each new generative model. Although the NPLM method is computationally more demanding than alternative approaches, in practical scenarios this trade-off may remain acceptable. In fact, executing a single model evaluation typically requires only a few seconds, representing an efficiency level comparable to that of relatively simple classifiers.

Another important aspect that we only briefly mentioned in this study is the impact of hyperparameter tuning of ML-powered tests on their sensitivity to specific sources of discrepancy between two generators. While some ML models allow one to approximately infer how model selection affects sensitivity, modern black-box approaches based on neural networks pose significant challenges in this regard (see [199] for a recent contribution on this topic).

Appendix

6.A Deformations

Various deformations of the reference model are considered, each parametrized by ϵ. The deformations are listed below, using design matrix notation, with x_{iI} and y_{iI} denoting the I-th component of the i-th point drawn by the reference and deformed model, respectively:

1. $\boldsymbol{\mu}$**-deformation** $\longrightarrow$ a shift in the mean:

$$y_{iI} = x_{iI} + \delta_{\mu\, I}\,, \quad \delta_{\mu\, I} \sim \mathcal{U}_{[-\epsilon,\epsilon]}$$

2. $\boldsymbol{\Sigma_{II}}$**-deformation** $\longrightarrow$ a standard deviation increasing:

$$y_{iI} = \mu_I + c_{\Sigma\, I}(x_{iI} - \mu_I)\,, \quad c_{\Sigma\, I} \sim \mathcal{U}_{[1,1+\epsilon]}$$

3. $\boldsymbol{\Sigma_{I\neq J}}$**-deformation** $\longrightarrow$ a shrinking of the off-diagonal terms of the correlation matrix:

$$y_{iI} = \sum_j P_{ij}^{(I)} x_{jI}\,, \quad P_{ij}^{(I)} = P_{ij}^{(I)}(\epsilon)\,, \quad P \text{ permutation matrix}$$

4. $\mathbf{pow}^{+}$**-deformation** $\longrightarrow$ a smearing of each point through a power bigger than one:

$$y_{iI} = \mathrm{sign}(x_{iI})|x_{iI}|^{1+\epsilon}\,, \quad \epsilon \geq 0$$

5. $\mathbf{pow}^{-}$**-deformation** $\longrightarrow$ a smearing of each point through a power smaller than one:

$$y_{iI} = \mathrm{sign}(x_{iI})|x_{iI}|^{1-\epsilon}\,, \quad \epsilon \geq 0$$

6. $\boldsymbol{\mathcal{N}}$**-deformation** $\longrightarrow$ a smearing of each point through a random shift obtained sampling from a normal distribution:

$$y_{iI} = x_{iI} + \delta_{iI}\,, \quad \delta_{iI} \sim \mathcal{N}_{0,\epsilon}$$

7. $\boldsymbol{\mathcal{U}}$**-deformation** $\longrightarrow$ a smearing of each point through a random shift obtained sampling from

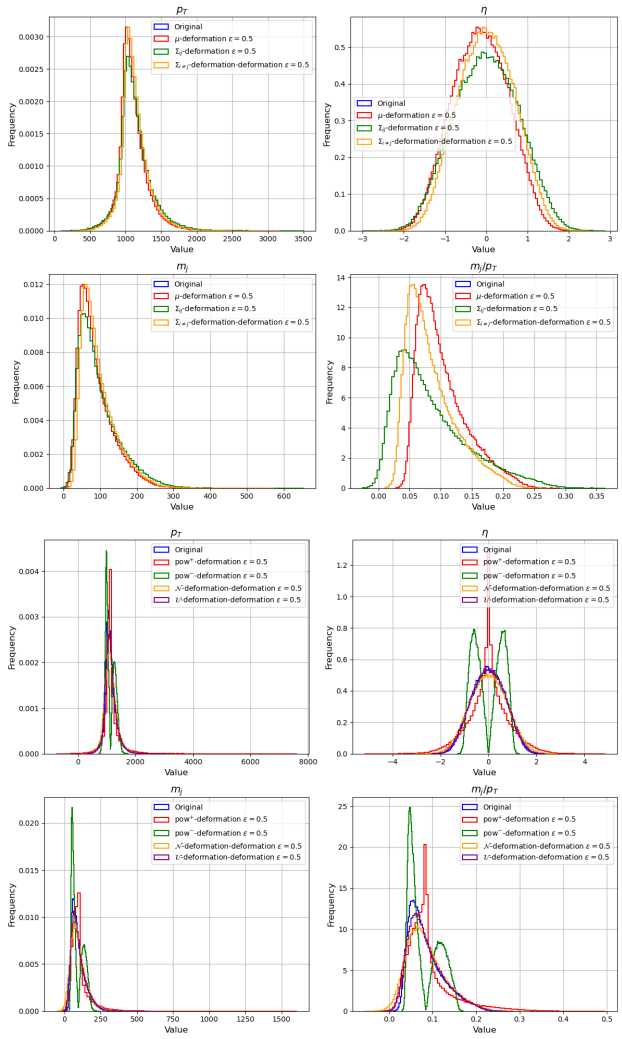


Figure 6.9: Original jet kinematic distributions compared with the μ, Σ_{ii}, and $\Sigma_{i\neq j}$ (top two rows), and pow$^+$, pow$^-$, $\mathcal{N}$, and $\mathcal{U}$ (bottom two rows) deformations with $\epsilon = 0.5$. The plots are made with 10^6 points per sample.

a uniform distribution:

$$y_{iI} = x_{iI} + \delta_{iI}\,, \quad \delta_{iI} \sim \mathcal{U}_{[-\epsilon,\epsilon]}$$

Each of these deformations is used to define an alternative hypothesis H_1, which is tested against H_0. See Figure 6.9 for an example of the effect of the deformations on the distribution of the particle features in the JetNet dataset .

6.B The NPLM test statistic

In this section we review the derivation of the NPLM test statistic starting from the extended likelihood formalism [195]. Given a model characterized by a pdf p and an expected number of

events n_p, we can write the likelihood function of a set of observations $\mathcal{Y} = \{y_i\}_{i=1}^m$ as

$$\mathcal{L}_p(\mathcal{Y}) = \frac{n_p^m e^{-n_p}}{m!} \prod_{y \in \mathcal{Y}} p(y), \tag{6.14}$$

where the number of observations m is modeled as a Poisson random variable with mean n_p. Given a second model with pdf q and expected number of events n_q, the log-likelihood ratio test statistic takes the following form

$$\begin{aligned} t(\mathcal{Y}) &= -2 \log \frac{\mathcal{L}_p(\mathcal{Y})}{\mathcal{L}_q(\mathcal{Y})} = -2 \log \left[e^{n_q - n_p} \prod_{y \in \mathcal{Y}} \frac{n_q\, q(y)}{n_p\, p(y)} \right] \\ &= -2 \left[n_q - n_p + \sum_{y \in \mathcal{Y}} \log \frac{n_q\, q(y)}{n_p\, p(y)} \right]. \end{aligned} \tag{6.15}$$

By defining $f(y) = \log \frac{\tilde{q}(y)}{\tilde{p}(y)}$, where $\tilde{p} = (n_q\, q)$ and $\tilde{p} = (n_p\, p)$ are the pdfs normalized to the respective number of events, one obtains

$$t(\mathcal{Y}) = -2 \left[n_q - n_p - \sum_{y \in \mathcal{Y}} f(y) \right]. \tag{6.16}$$

If the density p is not known in closed analytical form but it can be sampled via a generator, we can introduce a reference sample $\mathcal{X} = \{x_i\}_{i=1}^n$ drawn from $p(x)$. Using the definition of f, this can be used to estimate n_q as a Monte Carlo sum

$$\begin{aligned} n_q &= \int \tilde{q}(x)\, dx = \int \tilde{p}(x)\, e^{f(x)}\, dx \\ &= n_p \int p(x)\, e^{f(x)}\, dx \approx \frac{n_p}{n} \sum_{x \in \mathcal{X}} e^{f(x)}. \end{aligned} \tag{6.17}$$

Assuming that n_p is known, the test statistic becomes

$$t(\mathcal{X}, \mathcal{Y}) = -2 \left[\frac{n_p}{n} \sum_{x \in \mathcal{X}} \left(e^{f(x)} - 1 \right) - \sum_{y \in \mathcal{Y}} f(y) \right]. \tag{6.18}$$

If the number of expected events is not a relevant variable, we take $n_p = m$, namely the actual number of data points in the set of observations.

6.C Tables

CG model with $d = 20$ and $n = m = 2 \cdot 10^4$						
	μ-deformation			Σ_{ii}-deformation		
Statistic	$\epsilon_{95\%CL}$	$\epsilon_{99\%CL}$	t (s)	$\epsilon_{95\%CL}$	$\epsilon_{99\%CL}$	t (s)
t_{SW}	$0.07086^{+0.034}_{-0.031}$	$0.09763^{+0.033}_{-0.03}$	496	$0.03243^{+0.0098}_{-0.01}$	$0.04336^{+0.0093}_{-0.0091}$	525
$t_{\overline{KS}}$	$0.06957^{+0.034}_{-0.032}$	$0.09504^{+0.032}_{-0.032}$	**366**	$0.04632^{+0.015}_{-0.015}$	$0.06199^{+0.014}_{-0.014}$	**387**
t_{SKS}	$0.0699^{+0.033}_{-0.031}$	$0.09548^{+0.032}_{-0.03}$	579	$0.04131^{+0.013}_{-0.014}$	$0.05484^{+0.012}_{-0.012}$	620
t_{FGD}	$0.07185^{+0.041}_{-0.032}$	$0.09756^{+0.038}_{-0.03}$	1094	$0.0269^{+0.01}_{-0.0088}$	$0.03574^{+0.0091}_{-0.0081}$	1179
t_{MMD}	$0.08449^{+0.064}_{-0.049}$	$0.11846^{+0.06}_{-0.045}$	1574	$0.03081^{+0.026}_{-0.018}$	$0.04364^{+0.024}_{-0.017}$	1679
t_{NPLM}	$\mathbf{0.05351^{+0.017}_{-0.022}}$	$\mathbf{0.06383^{+0.016}_{-0.018}}$	23774	$\mathbf{0.01378^{+0.0048}_{-0.0065}}$	$\mathbf{0.01685^{+0.0043}_{-0.0051}}$	26822
t_{LLR}	$0.01053^{+0.0046}_{-0.0046}$	$0.01415^{+0.0045}_{-0.0047}$	1306	$0.00253^{+0.0013}_{-0.0014}$	$0.00353^{+0.0013}_{-0.0014}$	1504
	$\Sigma_{i\neq j}$-deformation			pow_+-deformation		
Statistic	$\epsilon_{95\%CL}$	$\epsilon_{99\%CL}$	t (s)	$\epsilon_{95\%CL}$	$\epsilon_{99\%CL}$	t (s)
t_{SW}	$0.04358^{+0.011}_{-0.013}$	$0.06168^{+0.0092}_{-0.01}$	**1024**	$0.00681^{+0.0027}_{-0.0027}$	$0.00928^{+0.0025}_{-0.0025}$	566
$t_{\overline{KS}}$	$1.04753^{+0.011}_{-0.016}$	$1.06042^{+0.017}_{-0.011}$	1376	$0.00901^{+0.0033}_{-0.0034}$	$0.01206^{+0.003}_{-0.0032}$	**422**
t_{SKS}	$0.05959^{+0.016}_{-0.02}$	$0.08015^{+0.015}_{-0.016}$	1128	$0.00745^{+0.0027}_{-0.0029}$	$0.0099^{+0.0025}_{-0.0026}$	632
t_{FGD}	$0.00749^{+0.0018}_{-0.0019}$	$0.01011^{+0.0016}_{-0.0017}$	2468	$0.00623^{+0.003}_{-0.0025}$	$0.00837^{+0.0027}_{-0.0023}$	1085
t_{MMD}	$0.05154^{+0.032}_{-0.023}$	$0.07052^{+0.032}_{-0.022}$	2534	$0.00488^{+0.0042}_{-0.0029}$	$0.00687^{+0.0039}_{-0.0027}$	1853
t_{NPLM}	$\mathbf{0.00271^{+0.0012}_{-0.0015}}$	$\mathbf{0.00347^{+0.0012}_{-0.0013}}$	36589	$\mathbf{0.00391^{+0.0015}_{-0.0019}}$	$\mathbf{0.00478^{+0.0013}_{-0.0016}}$	33561
t_{LLR}	-	-	-	$0.00069^{+0.0004}_{-0.0004}$	$0.00098^{+0.00039}_{-0.0004}$	1628
	pow_--deformation			$\mathcal{N}$-deformation		
Statistic	$\epsilon_{95\%CL}$	$\epsilon_{99\%CL}$	t (s)	$\epsilon_{95\%CL}$	$\epsilon_{99\%CL}$	t (s)
t_{SW}	$0.00748^{+5}_{-0.0026}$	$0.01003^{+5}_{-0.0024}$	513	$0.36054^{+0.048}_{-0.063}$	$0.42418^{+0.041}_{-0.045}$	472
$t_{\overline{KS}}$	$0.00946^{+0.003}_{-0.0031}$	$0.01227^{+0.0029}_{-0.0028}$	**418**	$0.40421^{+0.061}_{-0.079}$	$0.47653^{+0.053}_{-0.058}$	**353**
t_{SKS}	$0.00803^{+0.0027}_{-0.0029}$	$0.01046^{+0.0025}_{-0.0026}$	635	$0.35347^{+0.055}_{-0.07}$	$0.41869^{+0.042}_{-0.056}$	519
t_{FGD}	$0.0067^{+0.0027}_{-0.0023}$	$0.00886^{+0.0025}_{-0.0021}$	1090	$0.21348^{+0.022}_{-0.03}$	$0.24723^{+0.014}_{-0.019}$	850
t_{MMD}	$0.00568^{+0.0037}_{-0.0025}$	$0.00784^{+0.0036}_{-0.0025}$	1925	$0.87964^{+0.19}_{-0.23}$	$1.04846^{+0.14}_{-0.15}$	1413
t_{NPLM}	$\mathbf{0.00528^{+0.0015}_{-0.0018}}$	$\mathbf{0.00617^{+0.0013}_{-0.0015}}$	26885	$\mathbf{0.09544^{+0.016}_{-0.026}}$	$\mathbf{0.10546^{+0.014}_{-0.018}}$	21954
t_{LLR}	$0.00061^{+0.0004}_{-0.0004}$	$0.0009^{+0.00039}_{-0.0004}$	1652	-	-	-
	$\mathcal{U}$-deformation			Timing		
Statistic	$\epsilon_{95\%CL}$	$\epsilon_{99\%CL}$	t (s)	t^{null} (s)		
t_{SW}	$0.62405^{+0.086}_{-0.1}$	$0.73669^{+0.072}_{-0.076}$	458	276		
$t_{\overline{KS}}$	$0.69433^{+0.11}_{-0.13}$	$0.8144^{+0.099}_{-0.1}$	**337**	**108**		
t_{SKS}	$0.61401^{+0.092}_{-0.12}$	$0.72149^{+0.078}_{-0.092}$	505	398		
t_{FGD}	$0.37006^{+0.038}_{-0.053}$	$0.42858^{+0.025}_{-0.033}$	812	1150		
t_{MMD}	$1.53597^{+0.32}_{-0.42}$	$1.81389^{+0.26}_{-0.27}$	1386	2284		
t_{NPLM}	$\mathbf{0.16518^{+0.028}_{-0.046}}$	$\mathbf{0.18251^{+0.024}_{-0.031}}$	21414	44027		
t_{LLR}	-	-	-	-		

Table 6.5: Upper bounds and associated uncertainties on ϵ at 95% and 99% confidence levels, computed for different metrics and deformations. The table also reports the computation times required to estimate these values and to construct the $f(t_0)$ distribution. For each deformation, the best performing metric is indicated in bold.

MoG model with d = 20, q = 5, and n = m = $5 \cdot 10^4$

	μ-deformation			Σ_{ii}-deformation		
Statistic	$\epsilon_{95\%\text{CL}}$	$\epsilon_{99\%\text{CL}}$	t (s)	$\epsilon_{95\%\text{CL}}$	$\epsilon_{99\%\text{CL}}$	t (s)
t_{SW}	$0.04957^{+0.018}_{-0.02}$	$0.06694^{+0.017}_{-0.017}$	3023	$0.01679^{+0.005}_{-0.0063}$	$0.02315^{+0.0045}_{-0.005}$	3197
$t_{\overline{\text{KS}}}$	$0.00482^{+0.0013}_{-0.0018}$	$0.00667^{+0.0011}_{-0.0013}$	2966	$0.00175^{+0.00052}_{-0.00068}$	$0.00248^{+0.00042}_{-0.00052}$	3185
t_{SKS}	$0.03647^{+0.011}_{-0.014}$	$0.04821^{+0.011}_{-0.012}$	**2899**	$0.01329^{+0.003}_{-0.0043}$	$0.01759^{+0.0025}_{-0.003}$	**3022**
t_{FGD}	$0.05778^{+0.026}_{-0.027}$	$0.0787^{+0.023}_{-0.021}$	4047	$0.01945^{+0.0063}_{-0.0081}$	$0.02651^{+0.0053}_{-0.0056}$	4507
t_{MMD}	$0.04425^{+0.019}_{-0.018}$	$0.06215^{+0.017}_{-0.015}$	10204	$0.00923^{+0.0058}_{-0.0051}$	$0.01305^{+0.0053}_{-0.0044}$	11217
t_{NPLM}	$\mathbf{0.0025^{+0.0006}_{-0.00077}}$	$\mathbf{0.00294^{+0.00051}_{-0.00064}}$	41843	$\mathbf{0.00083^{+0.00023}_{-0.00028}}$	$\mathbf{0.00099^{+0.0002}_{-0.00024}}$	46776
t_{LLR}	$0.00021^{+0.00013}_{-0.00014}$	$0.0003^{+0.00013}_{-0.00014}$	5911	$0.00007^{+0.00005}_{-0.00004}$	$0.0001^{+0.00005}_{-0.00004}$	6304
	$\Sigma_{i\neq j}$-deformation			pow_+-deformation		
Statistic	$\epsilon_{95\%\text{CL}}$	$\epsilon_{99\%\text{CL}}$	t (s)	$\epsilon_{95\%\text{CL}}$	$\epsilon_{99\%\text{CL}}$	t (s)
t_{SW}	$0.02162^{+0.0056}_{-0.008}$	$0.02935^{+0.0045}_{-0.0055}$	**3410**	$0.00581^{+0.0017}_{-0.0022}$	$0.00798^{+0.0015}_{-0.0017}$	**3157**
$t_{\overline{\text{KS}}}$	$1.00146^{+0.00074}_{-0.00031}$	$1.00238^{+0.00055}_{-0.00031}$	3967	$0.0004^{+0.00015}_{-0.00017}$	$0.00059^{+0.00013}_{-0.00014}$	3363
t_{SKS}	$0.02306^{+0.0071}_{-0.0088}$	$0.03079^{+0.0062}_{-0.0072}$	3553	$0.0043^{+0.0009}_{-0.0013}$	$0.00565^{+0.00074}_{-0.0009}$	3193
t_{FGD}	$0.00551^{+0.0015}_{-0.002}$	$0.00748^{+0.0013}_{-0.0013}$	6327	$0.00702^{+0.0021}_{-0.0028}$	$0.00965^{+0.0016}_{-0.0019}$	4870
t_{MMD}	$0.01723^{+0.008}_{-0.0072}$	$0.02431^{+0.0069}_{-0.0064}$	11450	$0.00332^{+0.0018}_{-0.0017}$	$0.00467^{+0.0017}_{-0.0014}$	11801
t_{NPLM}	$\mathbf{0.0015^{+0.0007}_{-0.0007}}$	$\mathbf{0.00199^{+0.0006}_{-0.00069}}$	112295	$\mathbf{0.00019^{+0.00005}_{-0.00006}}$	$\mathbf{0.00022^{+0.00005}_{-0.00005}}$	52507
t_{LLR}	-	-	-	$0.00002^{+0.00001}_{-0.00001}$	$0.00002^{+0.00001}_{-0.00001}$	6877
	pow_--deformation			$\mathcal{N}$-deformation		
Statistic	$\epsilon_{95\%\text{CL}}$	$\epsilon_{99\%\text{CL}}$	t (s)	$\epsilon_{95\%\text{CL}}$	$\epsilon_{99\%\text{CL}}$	t (s)
t_{SW}	$0.00604^{+0.0017}_{-0.0023}$	$0.00825^{+0.0016}_{-0.0018}$	**3051**	$0.19318^{+0.025}_{-0.039}$	$0.22704^{+0.019}_{-0.026}$	**2403**
$t_{\overline{\text{KS}}}$	$0.00042^{+0.00015}_{-0.00018}$	$0.00061^{+0.00013}_{-0.00015}$	3372	$\mathbf{0.00751^{+0.002}_{-0.0024}}$	$\mathbf{0.00993^{+0.0018}_{-0.002}}$	2934
t_{SKS}	$0.00441^{+0.00092}_{-0.0014}$	$0.00574^{+0.00077}_{-0.00094}$	3324	$0.15874^{+0.023}_{-0.034}$	$0.18473^{+0.019}_{-0.023}$	2726
t_{FGD}	$0.00722^{+0.0021}_{-0.0027}$	$0.00987^{+0.0016}_{-0.0019}$	4892	$0.18095^{+0.023}_{-0.038}$	$0.21269^{+0.016}_{-0.02}$	3756
t_{MMD}	$0.00353^{+0.0016}_{-0.0015}$	$0.00494^{+0.0014}_{-0.0012}$	11418	$0.43531^{+0.066}_{-0.11}$	$0.51609^{+0.045}_{-0.054}$	8642
t_{NPLM}	$\mathbf{0.0002^{+0.00004}_{-0.00006}}$	$\mathbf{0.00023^{+0.00004}_{-0.00005}}$	48084	$0.03697^{+0.0054}_{-0.0081}$	$0.04073^{+0.0045}_{-0.0056}$	36180
t_{LLR}	$0.00002^{+0.00001}_{-0.00001}$	$0.00002^{+0.00001}_{-0.00001}$	6991	-	-	-
	$\mathcal{U}$-deformation			Timing		
Statistic	$\epsilon_{95\%\text{CL}}$	$\epsilon_{99\%\text{CL}}$	t (s)	t^{null} (s)		
t_{SW}	$0.33394^{+0.044}_{-0.068}$	$0.39248^{+0.033}_{-0.044}$	**2354**	338		
$t_{\overline{\text{KS}}}$	$\mathbf{0.01211^{+0.003}_{-0.0035}}$	$\mathbf{0.01575^{+0.0027}_{-0.003}}$	2835	**155**		
t_{SKS}	$0.27395^{+0.041}_{-0.059}$	$0.3188^{+0.033}_{-0.04}$	2601	509		
t_{FGD}	$0.31409^{+0.04}_{-0.07}$	$0.36919^{+0.027}_{-0.036}$	3643	2795		
t_{MMD}	$0.75353^{+0.12}_{-0.18}$	$0.89336^{+0.078}_{-0.098}$	7700	13860		
t_{NPLM}	$0.06387^{+0.0096}_{-0.014}$	$0.07083^{+0.0074}_{-0.01}$	34879	61789		
t_{LLR}	-	-	-	-		

Table 6.6: Upper bounds and associated uncertainties on ϵ at 95% and 99% confidence levels, computed for different metrics and deformations. The table also reports the computation times required to estimate these values and to construct the $f(t_0)$ distribution. For each deformation, the best performing metric is indicated in bold.

JetNet - Jet features with $\mathbf{n = m = 2 \cdot 10^4}$						
	μ-deformation			Σ_{ii}-deformation		
Statistic	$\epsilon_{95\%\text{CL}}$	$\epsilon_{99\%\text{CL}}$	t (s)	$\epsilon_{95\%\text{CL}}$	$\epsilon_{99\%\text{CL}}$	t (s)
t_{SW}	$0.02498^{+0.0068}_{-0.0089}$	$0.0347^{+0.0078}_{-0.007}$	**1856**	$0.03464^{+0.011}_{-0.014}$	$0.0454^{+0.011}_{-0.012}$	**1983**
$t_{\overline{\text{KS}}}$	$0.02347^{+0.0081}_{-0.0084}$	$0.0326^{+0.0073}_{-0.0065}$	2379	$0.03199^{+0.012}_{-0.013}$	$0.04302^{+0.012}_{-0.012}$	3749
t_{SKS}	$0.01854^{+0.0052}_{-0.0078}$	$0.02407^{+0.0044}_{-0.0075}$	5681	$0.03791^{+0.012}_{-0.016}$	$0.04866^{+0.012}_{-0.014}$	6885
t_{FGD}	$0.0322^{+0.009}_{-0.013}$	$0.03958^{+0.01}_{-0.0066}$	4014	$0.03122^{+0.013}_{-0.012}$	$0.04039^{+0.012}_{-0.0098}$	3399
t_{MMD}	$0.10604^{+0.024}_{-0.037}$	$0.13954^{+0.019}_{-0.025}$	6850	$0.05372^{+0.025}_{-0.021}$	$0.06748^{+0.024}_{-0.021}$	12626
t_{NPLM}	$\mathbf{5e-05^{+1e-05}_{-2e-05}}$	$\mathbf{6e-05^{+1e-05}_{-1e-05}}$	34148	$\mathbf{0.00029^{+8e-05}_{-0.00011}}$	$\mathbf{0.00036^{+7e-05}_{-7e-05}}$	32391
	$\Sigma_{i\neq j}$-deformation			pow_+-deformation		
Statistic	$\epsilon_{95\%\text{CL}}$	$\epsilon_{99\%\text{CL}}$	t (s)	$\epsilon_{95\%\text{CL}}$	$\epsilon_{99\%\text{CL}}$	t (s)
t_{SW}	$0.08331^{+0.027}_{-0.03}$	$0.11519^{+0.026}_{-0.022}$	1581	$0.04046^{+0.015}_{-0.016}$	$0.05256^{+0.016}_{-0.016}$	**1560**
$t_{\overline{\text{KS}}}$	$1.03549^{+0.0094}_{-0.015}$	$1.04064^{+0.015}_{-0.0051}$	**1173**	$0.03495^{+0.014}_{-0.014}$	$0.04617^{+0.015}_{-0.014}$	6217
t_{SKS}	$0.09468^{+0.035}_{-0.035}$	$0.12683^{+0.033}_{-0.024}$	5589	$0.06771^{+0.02}_{-0.029}$	$0.08308^{+0.02}_{-0.022}$	9295
t_{FGD}	$0.00933^{+0.0032}_{-0.0032}$	$0.01233^{+0.0031}_{-0.0022}$	6144	$0.03586^{+0.017}_{-0.016}$	$0.04367^{+0.017}_{-0.013}$	3613
t_{MMD}	$0.1341^{+0.062}_{-0.058}$	$0.17463^{+0.05}_{-0.046}$	79177	$0.03099^{+4.9}_{-0.014}$	$0.0413^{+4.9}_{-0.015}$	13970
t_{NPLM}	$\mathbf{0.00159^{+0.001}_{-0.00089}}$	$\mathbf{0.00239^{+0.00081}_{-0.00092}}$	37314	$\mathbf{0.00257^{+0.00089}_{-0.00095}}$	$\mathbf{0.00329^{+0.00074}_{-0.00069}}$	35528
	pow_--deformation			$\mathcal{N}$-deformation		
Statistic	$\epsilon_{95\%\text{CL}}$	$\epsilon_{99\%\text{CL}}$	t (s)	$\epsilon_{95\%\text{CL}}$	$\epsilon_{99\%\text{CL}}$	t (s)
t_{SW}	$0.0434^{+4.9}_{-0.018}$	$0.05684^{+4.9}_{-0.016}$	**1381**	$0.15561^{+0.026}_{-0.031}$	$0.17515^{+0.026}_{-0.025}$	**1501**
$t_{\overline{\text{KS}}}$	$0.03324^{+0.013}_{-0.014}$	$0.04369^{+0.015}_{-0.013}$	7555	$0.13682^{+0.018}_{-0.027}$	$0.15251^{+0.017}_{-0.01}$	7585
t_{SKS}	$0.06451^{+0.022}_{-0.028}$	$0.08132^{+0.021}_{-0.021}$	10657	$0.12378^{+0.028}_{-0.056}$	$0.1448^{+0.028}_{-0.039}$	11391
t_{FGD}	$0.03489^{+0.02}_{-0.015}$	$0.04502^{+0.02}_{-0.012}$	4270	$\mathbf{0.0948^{+0.015}_{-0.02}}$	$\mathbf{0.11199^{+0.0096}_{-0.013}}$	4014
t_{MMD}	$0.03264^{+0.024}_{-0.016}$	$0.048^{+0.02}_{-0.016}$	15086	$0.32021^{+0.058}_{-0.066}$	$0.37016^{+0.041}_{-0.041}$	11399
t_{NPLM}	$\mathbf{0.00276^{+0.00083}_{-0.0011}}$	$\mathbf{0.00344^{+0.00071}_{-0.00072}}$	28707	$0.01222^{+0.0024}_{-0.0051}$	$0.01421^{+0.0021}_{-0.0018}$	26004
	$\mathcal{U}$-deformation			Timing		
Statistic	$\epsilon_{95\%\text{CL}}$	$\epsilon_{99\%\text{CL}}$	t (s)	t^{null} (s)		
t_{SW}	$0.2705^{+0.047}_{-0.059}$	$0.30356^{+0.045}_{-0.029}$	**1396**	**271**		
$t_{\overline{\text{KS}}}$	$0.23391^{+0.024}_{-0.049}$	$0.25939^{+0.027}_{-0.024}$	8297	324		
t_{SKS}	$0.20877^{+0.051}_{-0.082}$	$0.24937^{+0.042}_{-0.046}$	10459	870		
t_{FGD}	$0.17213^{+0.014}_{-0.026}$	$0.19235^{+0.016}_{-0.015}$	2864	497		
t_{MMD}	$0.55884^{+0.088}_{-0.11}$	$0.6454^{+0.072}_{-0.07}$	10439	949		
t_{NPLM}	$\mathbf{0.02112^{+0.0039}_{-0.0086}}$	$\mathbf{0.02524^{+0.003}_{-0.0037}}$	24756	4468		

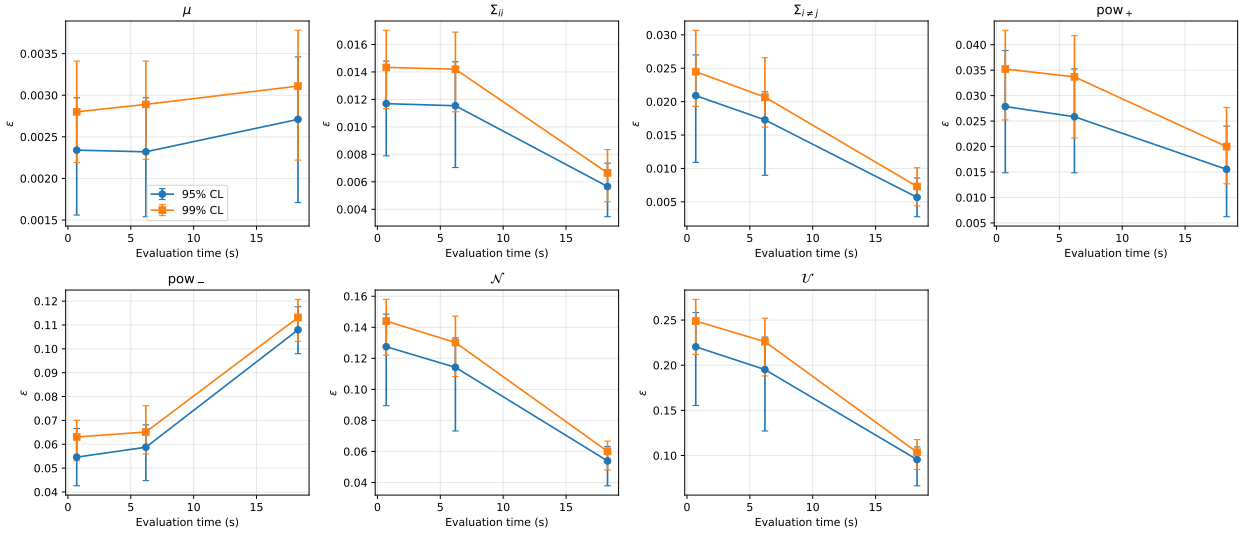
Table 6.7: Upper bounds and associated uncertainties on ϵ at 95% and 99% confidence levels, computed for different metrics and deformations. The table also reports the computation times required to estimate these values and to construct the $f(t_0)$ distribution. For each deformation, the best performing metric is indicated in bold.

JetNet - Particle features with $\mathbf{n = m = 2 \cdot 10^4}$						
	μ-deformation			Σ_{ii}-deformation		
Statistic	$\epsilon_{95\%\text{CL}}$	$\epsilon_{99\%\text{CL}}$	t (s)	$\epsilon_{95\%\text{CL}}$	$\epsilon_{99\%\text{CL}}$	t (s)
t_{SW}	$0.02051^{+0.0059}_{-0.0073}$	$0.02921^{+0.0058}_{-0.003}$	**1482**	$0.02531^{+0.0089}_{-0.01}$	$0.03339^{+0.01}_{-0.0057}$	**1030**
$t_{\overline{\text{KS}}}$	$0.00993^{+0.0052}_{-0.0078}$	$0.01668^{+0.0052}_{-0.0032}$	1875	$0.02465^{+0.016}_{-0.019}$	$0.03767^{+0.017}_{-0.0092}$	2464
t_{SKS}	$0.01919^{+0.0052}_{-0.0065}$	$0.02798^{+0.0013}_{-0.0067}$	3630	$0.03319^{+0.0078}_{-0.013}$	$0.04406^{+0.0065}_{-0.014}$	4096
t_{FGD}	$0.02609^{+0.0054}_{-0.011}$	$0.03056^{+0.0057}_{-0.0049}$	4986	$0.02305^{+0.0077}_{-0.0093}$	$0.02833^{+0.0068}_{-0.0078}$	6023
t_{MMD}	$0.028^{+0.0042}_{-0.011}$	$0.03484^{+0.0041}_{-0.0056}$	3875	$0.02671^{+0.011}_{-0.011}$	$0.03206^{+0.012}_{-0.0088}$	4796
t_{NPLM}	$\mathbf{0.00271^{+0.00075}_{-0.001}}$	$\mathbf{0.00311^{+0.00067}_{-0.00089}}$	111013	$\mathbf{0.00566^{+0.0017}_{-0.0022}}$	$\mathbf{0.00665^{+0.0017}_{-0.0021}}$	112034
	$\Sigma_{i\neq j}$-deformation			pow_+-deformation		
Statistic	$\epsilon_{95\%\text{CL}}$	$\epsilon_{99\%\text{CL}}$	t (s)	$\epsilon_{95\%\text{CL}}$	$\epsilon_{99\%\text{CL}}$	t (s)
t_{SW}	$0.05216^{+0.0084}_{-0.021}$	$0.06924^{+0.01}_{-0.012}$	**2288**	$0.03068^{+0.01}_{-0.014}$	$0.03771^{+0.009}_{-0.0064}$	**968**
$t_{\overline{\text{KS}}}$	$1.03785^{+0.009}_{-0.024}$	$1.04608^{+0.0082}_{-0.016}$	2734	$0.01587^{+0.0085}_{-0.013}$	$0.02425^{+0.011}_{-0.0046}$	3768
t_{SKS}	$0.06232^{+0.014}_{-0.028}$	$0.0766^{+0.018}_{-0.0079}$	7333	$0.04819^{+0.016}_{-0.017}$	$0.06357^{+0.02}_{-0.01}$	5176
t_{FGD}	$\mathbf{0.00395^{+0.0012}_{-0.0014}}$	$\mathbf{0.00521^{+0.0016}_{-0.0012}}$	10771	$0.0243^{+0.0079}_{-0.0098}$	$0.02987^{+0.0072}_{-0.0082}$	4923
t_{MMD}	$0.03602^{+0.013}_{-0.015}$	$0.04256^{+0.014}_{-0.012}$	41440	$0.03037^{+0.011}_{-0.012}$	$0.03734^{+0.0089}_{-0.011}$	5925
t_{NPLM}	$0.00567^{+0.0029}_{-0.0029}$	$0.0073^{+0.0028}_{-0.0029}$	110945	$\mathbf{0.01552^{+0.0085}_{-0.0093}}$	$\mathbf{0.01998^{+0.0077}_{-0.0073}}$	114774
	pow_--deformation			$\mathcal{N}$-deformation		
Statistic	$\epsilon_{95\%\text{CL}}$	$\epsilon_{99\%\text{CL}}$	t (s)	$\epsilon_{95\%\text{CL}}$	$\epsilon_{99\%\text{CL}}$	t (s)
t_{SW}	$0.03574^{+0.0092}_{-0.016}$	$0.04394^{+0.011}_{-0.011}$	**973**	$0.13338^{+0.031}_{-0.023}$	$0.15296^{+0.033}_{-0.0054}$	**787**
$t_{\overline{\text{KS}}}$	$\mathbf{0.01675^{+0.0096}_{-0.013}}$	$\mathbf{0.0256^{+0.011}_{-0.0076}}$	4374	$0.09466^{+0.019}_{-0.023}$	$0.10855^{+0.023}_{-0.021}$	3825
t_{SKS}	$0.05724^{+0.011}_{-0.025}$	$0.07036^{+0.017}_{-0.0072}$	5169	$0.13338^{+0.031}_{-0.023}$	$0.15296^{+0.033}_{-0.016}$	4479
t_{FGD}	$0.02527^{+0.0089}_{-0.011}$	$0.03169^{+0.0083}_{-0.0076}$	4913	$0.06571^{+0.0056}_{-0.012}$	$0.07484^{+0.003}_{-0.011}$	4537
t_{MMD}	$0.03366^{+0.01}_{-0.015}$	$0.04192^{+0.011}_{-0.01}$	6008	$0.35498^{+0.028}_{-0.049}$	$0.40428^{+0.016}_{-0.043}$	4996
t_{NPLM}	$0.10796^{+0.0097}_{-0.01}$	$0.11313^{+0.0076}_{-0.01}$	79016	$\mathbf{0.0539^{+0.0093}_{-0.016}}$	$\mathbf{0.06008^{+0.0066}_{-0.012}}$	87358
	$\mathcal{U}$-deformation			Timing		
Statistic	$\epsilon_{95\%\text{CL}}$	$\epsilon_{99\%\text{CL}}$	t (s)	t^{null} (s)		
t_{SW}	$0.22562^{+0.048}_{-0.041}$	$0.2808^{+0.033}_{-0.036}$	**765**	**269**		
$t_{\overline{\text{KS}}}$	$0.1547^{+0.025}_{-0.043}$	$0.19254^{+0.023}_{-0.046}$	4034	353		
t_{SKS}	$0.22562^{+0.048}_{-0.041}$	$0.2808^{+0.033}_{-0.054}$	4777	857		
t_{FGD}	$0.11465^{+0.0084}_{-0.027}$	$0.11995^{+0.016}_{-0.004}$	4174	1336		
t_{MMD}	$0.61572^{+0.041}_{-0.1}$	$0.70123^{+0.029}_{-0.081}$	2536	862		
t_{NPLM}	$\mathbf{0.09551^{+0.014}_{-0.029}}$	$\mathbf{0.10364^{+0.014}_{-0.019}}$	84708	18270		

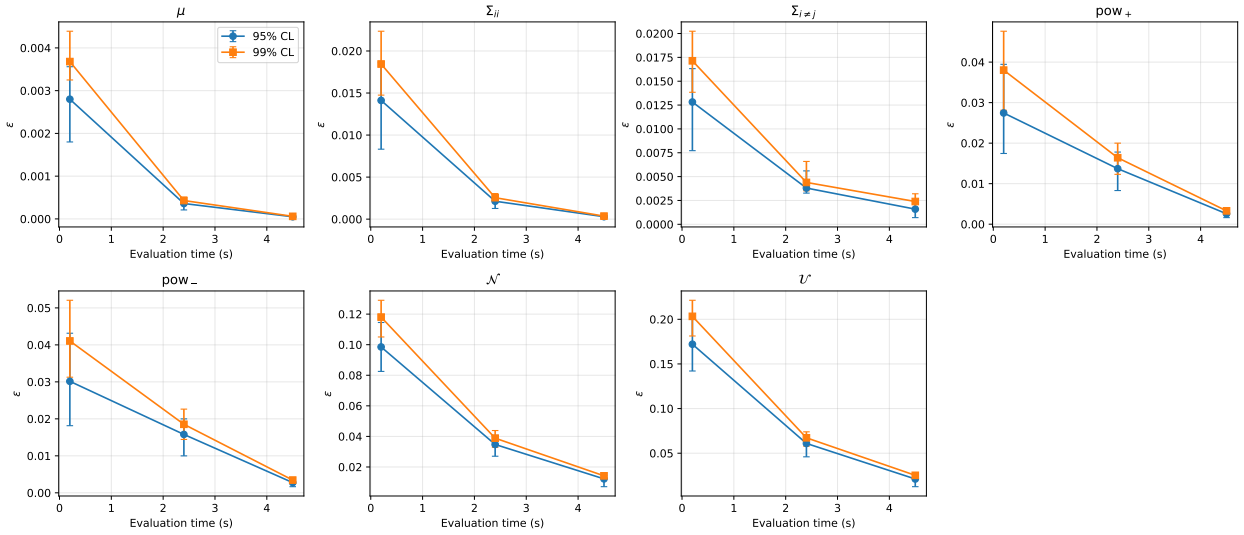
Table 6.8: Upper bounds and associated uncertainties on ϵ at 95% and 99% confidence levels, computed for different metrics and deformations. The table also reports the computation times required to estimate these values and to construct the $f(t_0)$ distribution. For each deformation, the best performing metric is indicated in bold.

6.D NPLM: performance against evaluation time

We report here the performance of the NPLM method as quantified by the parameter ϵ. The evaluation time is computed as the average runtime under the null hypothesis. In general, performance tends to plateau as the model complexity increases. This behaviour is less evident for particle-level features, which reside in a higher-dimensional space, where kernel-based methods, such as the one underlying this implementation of the NPLM test, are known to be less effective, as discussed in the main text. Moreover, sensitivity does not necessarily increase monotonically with model complexity, as certain deformations can be more efficiently detected by smoother models, as discussed in the main text.



(a) Particle-level features



(b) Jet-level features

Figure 6.10: NPLM performance on the JetNet dataset ($n = m = 2 \cdot 10^4$) as a function of evaluation time.

Chapter 7

Conclusions and outlook

In part I of this thesis, we exploited the Standard Model Effective Field Theory (SMEFT) framework to perform an indirect and model independent study of possible New Physics (NP) scenarios in the context of the Drell-Yan (DY) process. Our analysis was focused on providing realistic LHC predictions, taking into account the typical energy and luminosity benchmarks of the collider, as well as the experimental uncertainties expected in current and future measurements. We considered different classes of dimension-six operators, each characterized by their energy scaling and by whether or not they interfere with the Standard Model (SM) contributions.

In Chapter 2, we focused on the seven four-fermion current–current operators involving two quarks and two leptons with the same chirality structure as the SM. These operators provide the leading contributions to the cross section, since they interfere with the SM amplitudes and grow with the energy, making them especially relevant in the high-energy tails of kinematic distributions. We derived the 95% confidence level (CL) bounds on the corresponding Wilson coefficients, performing both one-dimensional and two-dimensional analyses. We also considered different binning strategies, focusing in particular on comparing the reach of multi-differential analyses with that of finely binned single-differential distributions.

In Chapter 3, we considered the sub-leading classes of chirality breaking dimension-6 operators, whose effects are suppressed in the inclusive DY cross section. In this case, we focused both on the cross section and on a specific angular observable, the so-called $A_0 - A_2$, where the SM contribution is suppressed up to $\mathcal{O}(\alpha_s^2)$. This observable provides sensitivity to chirality-breaking operators belonging to two distinct classes: the dipole operators, whose contributions do not grow with the energy and do not interfere with the SM, which are better constrained through the p_T distribution around the Z-boson peak and the four-fermion operators with a chirality structure different from that of the SM, which grow with the energy but also do not interfere with the SM, which are better constrained through the $m_{\ell\ell}$ distribution.

These two studies highlight the importance of selecting an appropriate analysis strategy when constraining a given NP operator. In particular, in Chapter 2 we observed that a finely binned single differential analysis can provide constraints comparable to those obtained from a multi-differential analysis. In our specific case the latter proved to be overall more sensitive, but a

single channel was used to constrain several operators simultaneously. Our results indicate that, for operators that are particularly sensitive to a specific observable, a fine binned analysis can already provide stringent bounds. This observation is particularly relevant when multiple channels are combined to constrain a set of operators and it motivates the inclusion of fine-binned single-differential measurements alongside multi-differential ones in experimental analyses. Finally, in Chapter 3, we identified an observable where the SM contribution is naturally suppressed, offering a clean and independent probe of chirality-breaking operators, free from contamination by dimension-six operators that interfere with the SM. Finding such observables can become particularly valuable in global SMEFT fits, where multiple operators are constrained simultaneously and degeneracies in parameter space must be resolved.

In part II of this thesis, we introduced a robust and general framework to evaluate the performance of different test statistics in two-sample testing. Our aim was to provide a fair and systematic comparison between various metrics, taking into account a wide range of data dimensionalities and sample sizes. Developing a clear and reliable validation framework is crucial for the modern High Energy Physics (HEP) community, since Machine Learning (ML) based generators are becoming increasingly important as tools to assist traditional Monte Carlo generators in improving the efficiency of synthetic data production. Ensuring the reliability of such data will become even more important in view of the advent of the High-Luminosity (HL) phase of the LHC, which will pose even more stringent precision requirements.

In Chapter 5, we quantified the ability of different metrics to discriminate between reference distributions and deformed ones. We introduced three test statistics based on one-dimensional distances: the mean Kolmogorov-Smirnov distance, the sliced Kolmogorov-Smirnov distance and the sliced Wasserstein distance and compared their performances with that of more complex metrics used in literature. Being multi-dimensional generalization of one-dimensional distances, the proposed metrics exhibit a low computational cost even in high-dimensional settings, in contrast to the more sophisticated ones, which are limited by the curse of dimensionality. Moreover, these metrics provide interpretable results, making them suitable as an initial validation step before turning to more complex architectures such as ML-based classifiers. We found that the performances of our proposed metrics were was comparable to the other, while being generally more efficient. Also, the reach for any considered deformation was stable across the different metrics, showing that our framework is a robust evaluation of the tests themselves, rather than merely assessing the specific models and deformations considered.

In Chapter 6, we extended the same framework to include a test-statistics provided by the kernel-based classifier known as New Physics Learning Machine (NPLM). Comparing its performance to those of the interpretable and low computing cost metrics introduced before, we could identify the regimes in which the adoption of such a more advanced tool is advantageous. Our results show that NPLM is generally more sensitive across all considered deformations, albeit at a significantly higher computational cost. This leads to the conclusion that ML-based classifiers are best suited for scenarios in which maximum sensitivity is required, while simpler metrics remain preferable

in situations where computational efficiency is a priority. In particular, one-dimensional distance based metrics can be used as a first validation step, either to quickly reject inadequate models or during the tuning of generative-model hyperparameters, before turning to more computationally demanding classifiers.

These two studies provide a benchmark for evaluating the performance of test statistics in two-sample testing. We showed that there is not a universally best performing metric, as the choice depends on the specific goals and constraints of the analysis. When there are no time limitations and high sensitivity is required, NPLM is the most powerful option, while for faster but less demanding applications, one-dimensional distance–based tests offer an efficient and reliable alternative.

A natural extension of this framework would be the evaluation of the null hypothesis for widely used benchmark datasets in the HEP community, providing a common reference for assessing and comparing the performance of different generative models trained on such references.

Bibliography

[1] S. Grossi and R. Torre, *More variables or more bins? Impact on the EFT interpretation of Drell–Yan measurements*, *Eur. Phys. J. C* **84** (2024), no. 7 713, [arXiv:2404.10569].

[2] S. Grossi, X. Li, L. Rolla, and R. Torre, *Isolating chirality-breaking smeft operators with drell-yan angular analysis*, 2025.

[3] S. Grossi, M. Letizia, and R. Torre, *Refereeing the referees: evaluating two-sample tests for validating generators in precision sciences*, *Mach. Learn. Sci. Tech.* **6** (2025), no. 1 015052, [arXiv:2409.16336].

[4] S. Grossi, M. Letizia, and R. Torre, *Comparing generative models with the new physics learning machine*, *Nucl. Phys. B* **1024** (2026) 117349, [arXiv:2508.02275].

[5] S. Weinberg, *Phenomenological Lagrangians*, *Physica A* **96** (1979), no. 1-2 327–340.

[6] H. Georgi, *Effective field theory*, *Ann. Rev. Nucl. Part. Sci.* **43** (1993) 209–252.

[7] J. Polchinski, *Effective field theory and the Fermi surface*, in *Theoretical Advanced Study Institute (TASI 92): From Black Holes and Strings to Particles*, pp. 0235–276, 6, 1992. hep-th/9210046.

[8] D. B. Kaplan, *Five lectures on effective field theory*, 10, 2005. nucl-th/0510023.

[9] A. V. Manohar and M. B. Wise, *Heavy quark physics*, vol. 10. 2000.

[10] A. Pich, *Effective field theory: Course*, in *Les Houches Summer School in Theoretical Physics, Session 68: Probing the Standard Model of Particle Interactions*, pp. 949–1049, 6, 1998. hep-ph/9806303.

[11] S. Davidson, P. Gambino, M. Laine, M. Neubert, and C. Salomon, eds., *Proceedings, Les Houches summer school: EFT in Particle Physics and Cosmology: Les Houches (Chamonix Valley), France, July 3-28, 2017*, vol. 108 of *Lecture Notes of the Les Houches Summer School*. Oxford University Press, Oxford, 4, 2020.

[12] T. Cohen, *As Scales Become Separated: Lectures on Effective Field Theory*, *PoS* **TASI2018** (2019) 011, [arXiv:1903.03622].

[13] J. Polchinski, *Renormalization and effective lagrangians*, *Nuclear Physics B* **231** (1984), no. 2 269–295.

[14] G. F. Giudice and M. McCullough, *A Clockwork Theory*, *JHEP* **02** (2017) 036, [arXiv:1610.07962].

[15] M. E. Peskin and D. V. Schroeder, *An Introduction to quantum field theory*. Addison-Wesley, Reading, USA, 1995.

[16] C. Quigg, *Gauge Theories of the Strong, Weak, and Electromagnetic Interactions: Second Edition*. Princeton University Press, USA, 9, 2013.

[17] C. P. Burgess and G. D. Moore, *The standard model: A primer*. Cambridge University Press, 12, 2006.

[18] W. N. Cottingham and D. A. Greenwood, *An Introduction to the Standard Model of Particle Physics*. Cambridge University Press, 2007.

[19] M. D. Schwartz, *Quantum Field Theory and the Standard Model*. Cambridge University Press, 3, 2014.

[20] C. M. Becchi and G. Ridolfi, *An introduction to relativistic processes and the standard model of electroweak interactions*. 2006.

[21] G. Altarelli, *The Standard electroweak theory and beyond*, in *Theoretical Advanced Study Institute in Elementary Particle Physics (TASI 98): Neutrinos in Physics and Astrophysics: From 10**(-33) to 10**(+28) cm*, pp. 27–93, 11, 1998. hep-ph/9811456.

[22] A. Pich, *The Standard Model of Electroweak Interactions*, in *2010 European School of High Energy Physics*, pp. 1–50, 1, 2012. arXiv:1201.0537.

[23] F. Quevedo and A. Schachner, *Cambridge Lectures on The Standard Model*, arXiv:2409.09211.

[24] P. W. Higgs, *Broken symmetries, massless particles and gauge fields*, *Phys. Lett.* **12** (1964) 132–133.

[25] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, *Phys. Rev. Lett.* **13** (1964) 508–509.

[26] **ATLAS** Collaboration, G. Aad et al., *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, *Phys. Lett. B* **716** (2012) 1–29, [arXiv:1207.7214].

[27] **CMS** Collaboration, S. Chatrchyan et al., *Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC*, *Phys. Lett. B* **716** (2012) 30–61, [arXiv:1207.7235].

[28] R. Barbieri, *Electroweak theory after the first Large Hadron Collider phase*, *Phys. Scripta T* **158** (2013) 014006, [arXiv:1309.3473].

[29] G. F. Giudice, *Naturalness after LHC8*, *PoS* **EPS-HEP2013** (2013) 163, [arXiv:1307.7879].

[30] **Super-Kamiokande** Collaboration, Y. Fukuda et al., *Evidence for oscillation of atmospheric neutrinos*, *Phys. Rev. Lett.* **81** (1998) 1562–1567, [hep-ex/9807003].

[31] A. Loureiro, A. Cuceu, F. B. Abdalla, B. Moraes, L. Whiteway, M. McLeod, S. T. Balan, O. Lahav, A. Benoit-Lévy, M. Manera, R. P. Rollins, and H. S. Xavier, *Upper bound of neutrino masses from combined cosmological observations and particle physics experiments*, *Phys. Rev. Lett.* **123** (Aug, 2019) 081301.

[32] S. Weinberg, *ULTRAVIOLET DIVERGENCES IN QUANTUM THEORIES OF GRAVITATION*, pp. 790–831. 1980.

[33] A. D. Sakharov, *Violation of CP Invariance, C asymmetry, and baryon asymmetry of the universe*, *Pisma Zh. Eksp. Teor. Fiz.* **5** (1967) 32–35.

[34] R. D. Peccei, *The Strong CP problem and axions*, *Lect. Notes Phys.* **741** (2008) 3–17, [hep-ph/0607268].

[35] G. Bertone, D. Hooper, and J. Silk, *Particle dark matter: Evidence, candidates and constraints*, *Phys. Rept.* **405** (2005) 279–390, [hep-ph/0404175].

[36] H. P. Nilles, *Supersymmetry, Supergravity and Particle Physics*, *Phys. Rept.* **110** (1984) 1–162.

[37] R. D. Peccei and H. R. Quinn, *CP Conservation in the Presence of Instantons*, *Phys. Rev. Lett.* **38** (1977) 1440–1443.

[38] P. Minkowski, $\mu \to e\gamma$ *at a Rate of One Out of* 10^9 *Muon Decays?*, *Phys. Lett. B* **67** (1977) 421–428.

[39] I. Brivio and M. Trott, *The Standard Model as an Effective Field Theory*, *Phys. Rept.* **793** (2019) 1–98, [arXiv:1706.08945]. Semantic Scholar.

[40] W. Buchmuller and D. Wyler, *Effective Lagrangian Analysis of New Interactions and Flavor Conservation*, *Nucl. Phys. B* **268** (1986) 621–653.

[41] B. Grzadkowski, M. Iskrzynski, M. Misiak, and J. Rosiek, *Dimension-Six Terms in the Standard Model Lagrangian*, *JHEP* **10** (2010) 085, [arXiv:1008.4884]. Semantic Scholar.

[42] R. Franceschini, G. Panico, A. Pomarol, F. Riva, and A. Wulzer, *Electroweak Precision Tests in High-Energy Diboson Processes*, *JHEP* **02** (2018) 111, [arXiv:1712.01310].

[43] M. Farina, G. Panico, D. Pappadopulo, J. T. Ruderman, R. Torre, and A. Wulzer, *Energy helps accuracy: electroweak precision tests at hadron colliders*, *Phys. Lett. B* **772** (2017) 210–215, [arXiv:1609.08157]. Semantic Scholar.

[44] S. Alioli, M. Farina, D. Pappadopulo, and J. T. Ruderman, *Catching a New Force by the Tail*, *Phys. Rev. Lett.* **120** no. 10 101801, [arXiv:1712.02347]. Semantic Scholar.

[45] S. Alioli, M. Farina, D. Pappadopulo, and J. T. Ruderman, *Precision Probes of QCD at High Energies*, *JHEP* **07** (2017) 097, [arXiv:1706.03068]. Semantic Scholar.

[46] T. Han, S. C. I. Leung, and M. Low, *Higgs to $b\bar{b}$ from Vector Boson Fusion for High-Scale Physics*, arXiv:2305.01010. Semantic Scholar.

[47] **SMEFiT** Collaboration, J. J. Ethier, G. Magni, F. Maltoni, L. Mantani, E. R. Nocera, J. Rojo, E. Slade, E. Vryonidou, and C. Zhang, *Combined SMEFT interpretation of Higgs, diboson, and top quark data from the LHC*, *JHEP* **11** (2021) 089, [arXiv:2105.00006]. Semantic Scholar.

[48] G. F. Giudice, C. Grojean, A. Pomarol, and R. Rattazzi, *The Strongly-Interacting Light Higgs*, *JHEP* **06** (2007) 045, [hep-ph/0703164]. Semantic Scholar.

[49] D. Liu, A. Pomarol, R. Rattazzi, and F. Riva, *Patterns of Strong Coupling for LHC Searches*, *JHEP* **11** (2016) 141, [arXiv:1603.03064]. Semantic Scholar.

[50] **LHC Higgs Cross Section Working Group** Collaboration, D. de Florian et al., *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*, arXiv:1610.07922. Semantic Scholar.

[51] **CMS** Collaboration, A. M. Sirunyan et al., *Combined measurements of Higgs boson couplings in proton–proton collisions at $\sqrt{s} = 13$ TeV*, *Eur. Phys. J. C* **79** (2019), no. 5 421, [arXiv:1809.10733]. Semantic Scholar.

[52] A. N. Rossia, M. O. A. Thomas, and E. Vryonidou, *Diboson production in the SMEFT from gluon fusion*, *JHEP* **11** (2023) 132, [arXiv:2306.09963]. Semantic Scholar.

[53] C. Degrande and H.-L. Li, *Impact of dimension-8 SMEFT operators on diboson productions*, *JHEP* **06** (2023) 149, [arXiv:2303.10493]. Semantic Scholar.

[54] **ATLAS** Collaboration, G. Aad et al., *Interpretations of the ATLAS measurements of Higgs boson production and decay rates and differential cross-sections in pp collisions at $\sqrt{s} = 13$ TeV*, arXiv:2402.05742. Semantic Scholar.

[55] P. Englert, *New Physics Probes at Present/Future Hadron Colliders via Vh Production*. PhD thesis, Humboldt U., Berlin, 2023. InSpireHEP.

[56] J. Ellis, M. Madigan, K. Mimasu, V. Sanz, and T. You, *Top, Higgs, Diboson and Electroweak Fit to the Standard Model Effective Field Theory*, *JHEP* **04** (2021) 279, [arXiv:2012.02779]. Semantic Scholar.

[57] **CMS** Collaboration, T. C. Collaboration et al., *Search for new physics in top quark production with additional leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV using effective field theory*, *JHEP* **03** (2021) 095, [arXiv:2012.04120]. Semantic Scholar.

[58] S. Tong, J. Corcoran, M. Fieg, M. Fenton, and D. Whiteson, *New Physics in Single Resonant Top Quarks*, arXiv:2311.00121. Semantic Scholar.

[59] N. Elmer, M. Madigan, T. Plehn, and N. Schmal, *Staying on Top of SMEFT-Likelihood Analyses*, arXiv:2312.12502. Semantic Scholar.

[60] C. Englert, G. F. Giudice, A. Greljo, and M. Mccullough, *The $\hat{H}$-Parameter: An Oblique Higgs View*, *JHEP* **09** (2019) 041, [arXiv:1903.07725]. Semantic Scholar.

[61] R. Torre, L. Ricci, and A. Wulzer, *On the W&Y interpretation of high-energy Drell-Yan measurements*, *JHEP* **02** (2021) 144, [arXiv:2008.12978]. Semantic Scholar.

[62] G. Panico, L. Ricci, and A. Wulzer, *High-energy EFT probes with fully differential Drell-Yan measurements*, *JHEP* **07** (2021) 086, [arXiv:2103.10532]. Semantic Scholar.

[63] S. Alioli, P. Nason, C. Oleari, and E. Re, *NLO vector-boson production matched with shower in POWHEG*, *JHEP* **07** (2008) 060, [arXiv:0805.4802]. Semantic Scholar.

[64] S. Frixione, P. Nason, and C. Oleari, *Matching NLO QCD computations with Parton Shower simulations: the POWHEG method*, *JHEP* **11** (2007) 070, [arXiv:0709.2092]. Semantic Scholar.

[65] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands, *An introduction to PYTHIA 8.2*, *Comput. Phys. Commun.* **191** (2015) 159–177, [arXiv:1410.3012]. Semantic Scholar.

[66] Y. Li and F. Petriello, *Combining QCD and electroweak corrections to dilepton production in FEWZ*, *Phys. Rev. D* **86** (2012) 094034, [arXiv:1208.5967]. Semantic Scholar.

[67] J. Gao, L. Harland-Lang, and J. Rojo, *The Structure of the Proton in the LHC Precision Era*, *Phys. Rept.* **742** (2018) 1–121, [arXiv:1709.04922]. Semantic Scholar.

[68] E. Hammou, Z. Kassabov, M. Madigan, M. L. Mangano, L. Mantani, J. Moore, M. M. Alvarado, and M. Ubiali, *Hide and seek: how PDFs can conceal new physics*, *JHEP* **11** (2023) 090, [arXiv:2307.10370]. Semantic Scholar.

[69] S. Amoroso et al., *Snowmass 2021 Whitepaper: Proton Structure at the Precision Frontier*, *Acta Phys. Polon. B* **53** (2022), no. 12 12–A1, [arXiv:2203.13923]. Semantic Scholar.

[70] A. Buckley, J. Ferrando, S. Lloyd, K. Nordström, B. Page, M. Rüfenacht, M. Schönherr, and G. Watt, *LHAPDF6: parton density access in the LHC precision era*, *Eur. Phys. J. C* **75** [arXiv:1412.7420]. Semantic Scholar.

[71] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, *Eur. Phys. J. C* **71** (2011) 1554, [arXiv:1007.1727].

[72] C. S. Lam and W.-K. Tung, *Structure Function Relations at Large Transverse Momenta in Lepton Pair Production Processes*, *Phys. Lett. B* **80** (1979) 228–231. [InSpire] [Semantic Scholar].

[73] C. S. Lam and W.-K. Tung, *A Systematic Approach to Inclusive Lepton Pair Production in Hadronic Collisions*, *Phys. Rev. D* **18** (1978) 2447. [InSpire] [Semantic Scholar].

[74] J. C. Collins and D. E. Soper, *Angular Distribution of Dileptons in High-Energy Hadron Collisions*, *Phys. Rev. D* **16** (1977) 2219. [InSpire] [Semantic Scholar].

[75] X. Li, B. Yan, and C. P. Yuan, *Lam-Tung relation breaking in Z boson production as a probe of standard model effective field theory effects*, *Phys. Rev. D* **111** (2025), no. 7 073007, [arXiv:2405.04069]. [InSpire] [Semantic Scholar].

[76] G. Li, X. Li, and B. Yan, *Lam-Tung relation breaking effects and weak dipole moments at lepton colliders*, *Phys. Lett. B* **870** (2025) 139931, [arXiv:2503.17663]. [InSpire] [Semantic Scholar].

[77] R. Gauld, U. Haisch, and J. Weiss, *A tale of Z+jet: SMEFT effects and the Lam-Tung relation*, *SciPost Phys.* **18** (2025), no. 5 148, [arXiv:2412.13014]. [InSpire] [Semantic Scolar.

[78] V. E. Lyubovitskij, A. S. Zhevlakov, and I. A. Anikin, *Angular coefficients of the Drell-Yan process across different rapidity and kinematical ranges*, *Phys. Rev. D* **112** (2025), no. 5 054023, [arXiv:2503.16008]. [InSpire] [Semantic Scholar].

[79] Y. B. Bandeira, V. P. Goncalves, and W. Schäfer, *Dilepton angular distributions in the color-dipole S-matrix framework*, *JHEP* **10** (2025) 010, [arXiv:2507.06207]. [InSpire] [Semantic Scholar].

[80] F. Petriello and K. Şimşek, *Naive T-odd Drell-Yan angular coefficients as a probe of the dimension-8 SMEFT*, arXiv:2511.19617. [InSpire] [Semantic Scholar].

[81] P. F. Monni, P. Nason, E. Re, M. Wiesemann, and G. Zanderighi, *MiNNLO$_{PS}$: a new method to match NNLO QCD to parton showers*, *JHEP* **05** (2020) 143, [arXiv:1908.06987]. [InSpire] [Semantic Scholar].

[82] P. F. Monni, E. Re, and M. Wiesemann, *MiNNLO$_{PS}$: optimizing $2 \to 1$ hadronic processes*, *Eur. Phys. J. C* **80** (2020), no. 11 1075, [arXiv:2006.04133]. [InSpire] [Semantic Scholar].

[83] P. Nason, *A New method for combining NLO QCD with shower Monte Carlo algorithms*, *JHEP* **11** (2004) 040, [hep-ph/0409146]. [InSpire] [Semantic Scholar].

[84] S. Alioli, P. Nason, C. Oleari, and E. Re, *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, *JHEP* **06** (2010) 043, [arXiv:1002.2581]. [Inspire] [Semantic Scholar].

[85] T. Sjostrand, S. Mrenna, and P. Z. Skands, *A Brief Introduction to PYTHIA 8.1*, *Comput. Phys. Commun.* **178** (2008) 852–867, [arXiv:0710.3820]. [InSpire] [Semantic Scholar].

[86] C. Bierlich et al., *A comprehensive guide to the physics and usage of PYTHIA 8.3*, *SciPost Phys. Codeb.* **2022** (2022) 8, [arXiv:2203.11601]. [InSpire] [Semantic Scholar].

[87] J. Alwall et al., *A Standard format for Les Houches event files*, *Comput. Phys. Commun.* **176** (2007) 300–304, [hep-ph/0609017]. [InSpire] [Semantic Scholar].

[88] V. Shtabovenko, R. Mertig, and F. Orellana, *FeynCalc 9.3: New features and improvements*, *Comput. Phys. Commun.* **256** (2020) 107478, [arXiv:2001.04407]. [InSpire] [Semantic Scholar].

[89] T. Hahn, *Generating Feynman diagrams and amplitudes with FeynArts 3*, *Comput. Phys. Commun.* **140** (2001) 418–431, [hep-ph/0012260]. [InSpire] [Semantic Scholar].

[90] Wolfram Research, Inc., "Mathematica, Version 13.0." Champaign, IL, 2021 [Website].

[91] D. B. Clark, E. Godat, and F. I. Olness, *Maneparse : A mathematica reader for parton distribution functions*, *Comput. Phys. Commun.* **216** (2017) 126–137, [arXiv:1605.08012].

[92] **ATLAS** Collaboration, G. Aad et al., *Precise measurements of W- and Z-boson transverse momentum spectra with the ATLAS detector using pp collisions at $\sqrt{s} = 5.02$ TeV and 13 TeV*, *Eur. Phys. J. C* **84** (2024), no. 10 1126, [arXiv:2404.06204]. [InSpire] [Semantic Scholar].

[93] **ATLAS** Collaboration, G. Aad et al., *A precise measurement of the Z-boson double-differential transverse momentum and rapidity distributions in the full phase space of the decay leptons with the ATLAS experiment at $\sqrt{s} = 8$ TeV*, *Eur. Phys. J. C* **84** (2024), no. 3 315, [arXiv:2309.09318]. [InSpire] [Semantic Scholar].

[94] **CMS** Collaboration, A. Tumasyan et al., *Measurement of the mass dependence of the transverse momentum of lepton pairs in Drell-Yan production in proton-proton collisions at $\sqrt{s} = 13$ TeV*, *Eur. Phys. J. C* **83** (2023), no. 7 628, [arXiv:2205.04897]. [InSpire] [Semantic Scholar].

[95] **LHCb** Collaboration, R. Aaij et al., *First Measurement of the Z→μ+μ- Angular Coefficients in the Forward Region of pp Collisions at s=13 TeV*, *Phys. Rev. Lett.* **129** (2022), no. 9 091801, [arXiv:2203.01602]. [InSpire] [Semantic Scholar].

[96] **ATLAS** Collaboration, G. Aad et al., *Measurement of the transverse momentum distribution of Drell–Yan lepton pairs in proton–proton collisions at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, *Eur. Phys. J. C* **80** (2020), no. 7 616, [arXiv:1912.02844]. [InSpire] [Semantic Scholar].

[97] **CMS** Collaboration, A. M. Sirunyan et al., *Measurements of differential Z boson production cross sections in proton-proton collisions at* $\sqrt{s}$ *= 13 TeV*, *JHEP* **12** (2019) 061, [arXiv:1909.04133]. [InSpire] [Semantic Scholar].

[98] **CMS** Collaboration, A. M. Sirunyan et al., *Measurement of the differential Drell-Yan cross section in proton-proton collisions at* $\sqrt{\text{s}}$ *= 13 TeV*, *JHEP* **12** (2019) 059, [arXiv:1812.10529]. [InSpire] [Semantic Scholar].

[99] **CMS** Collaboration, A. M. Sirunyan et al., *Measurement of differential cross sections in the kinematic angular variable* ϕ^* *for inclusive Z boson production in pp collisions at* $\sqrt{s} = 8$ *TeV*, *JHEP* **03** (2018) 172, [arXiv:1710.07955]. [InSpire] [Semantic Scholar].

[100] **ATLAS** Collaboration, M. Aaboud et al., *Measurement of the Drell-Yan triple-differential cross section in pp collisions at* $\sqrt{s} = 8$ *TeV*, *JHEP* **12** (2017) 059, [arXiv:1710.05167]. Semantic Scholar.

[101] **CMS** Collaboration, V. Khachatryan et al., *Measurement of the transverse momentum spectra of weak vector bosons produced in proton-proton collisions at* $\sqrt{s} = 8$ *TeV*, *JHEP* **02** (2017) 096, [arXiv:1606.05864]. [InSpire] [Semantic Scholar].

[102] **ATLAS** Collaboration, G. Aad et al., *Measurement of the double-differential high-mass Drell-Yan cross section in pp collisions at* $\sqrt{s} = 8$ *TeV with the ATLAS detector*, *JHEP* **08** (2016) 009, [arXiv:1606.01736]. Semantic Scholar.

[103] **ATLAS** Collaboration, G. Aad et al., *Measurement of the angular coefficients in Z-boson events using electron and muon pairs from data taken at* $\sqrt{s} = 8$ *TeV with the ATLAS detector*, *JHEP* **08** (2016) 159, [arXiv:1606.00689]. [InSpire] [Semantic Scholar].

[104] **ATLAS** Collaboration, G. Aad et al., *Measurement of the transverse momentum and* ϕ^*_η *distributions of Drell–Yan lepton pairs in proton–proton collisions at* $\sqrt{s} = 8$ *TeV with the ATLAS detector*, *Eur. Phys. J. C* **76** (2016), no. 5 291, [arXiv:1512.02192]. [InSpire] [Semantic Scholar].

[105] **CMS** Collaboration, V. Khachatryan et al., *Angular coefficients of Z bosons produced in pp collisions at* $\sqrt{s}$ *= 8 TeV and decaying to* $\mu^+\mu^-$ *as a function of transverse momentum and*

rapidity, *Phys. Lett. B* **750** (2015) 154–175, [arXiv:1504.03512]. [InSpire] [Semantic Scholar].

[106] **CMS** Collaboration, V. Khachatryan et al., *Measurement of the Z boson differential cross section in transverse momentum and rapidity in proton–proton collisions at 8 TeV*, *Phys. Lett. B* **749** (2015) 187–209, [arXiv:1504.03511]. [InSpire] [Semantic Scholar].

[107] N. Arteaga-Romero, A. Nicolaidis, and J. Silva, *Z^0 Production at the $p\bar{p}$ Collider and the Spin of the Gluon*, *Phys. Rev. Lett.* **52** (1984) 172. [InSpire] [Semantic Scholar].

[108] S. Alioli, R. Boughezal, E. Mereghetti, and F. Petriello, *Novel angular dependence in Drell-Yan lepton production via dimension-8 operators*, *Phys. Lett. B* **809** (2020) 135703, [arXiv:2003.11615].

[109] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer, and T. Stelzer, *MadGraph 5: Going Beyond*, *JHEP* **06** (2011) 128, [arXiv:1106.0522]. [InSpire] [Semantic Scholar].

[110] N. Valori and O. Vives, *Beyond the Standard Model contributions to dipole moments*, arXiv:2505.06345. [InSpire] [Semantic Scholar].

[111] J. de Blas, J. C. Criado, M. Perez-Victoria, and J. Santiago, *Effective description of general extensions of the Standard Model: the complete tree-level dictionary*, *JHEP* **03** (2018) 109, [arXiv:1711.10391]. Semantic Scholar.

[112] I. Doršner, S. Fajfer, A. Greljo, J. F. Kamenik, and N. Košnik, *Physics of leptoquarks in precision experiments and at particle colliders*, *Phys. Rept.* **641** (2016) 1–68, [arXiv:1603.04993]. [InSpire] [Semantic Scholar].

[113] O. U. Shanker, *$\pi\ell$ 2, $K\ell$ 3 and $K^0 - \bar{K}0$ Constraints on Leptoquarks and Supersymmetric Particles*, *Nucl. Phys. B* **204** (1982) 375–386. [InSpire] [Semantic Scholar].

[114] **Particle Data Group** Collaboration, S. Navas et al., *Review of particle physics*, *Phys. Rev. D* **110** (2024), no. 3 030001. [InSpire].

[115] M. Leurer, *A Comprehensive study of leptoquark bounds*, *Phys. Rev. D* **49** (1994) 333–342, [hep-ph/9309266]. [InSpire] [Semantic Scholar].

[116] M. Leurer, *Bounds on vector leptoquarks*, *Phys. Rev. D* **50** (1994) 536–541, [hep-ph/9312341]. [InSpire] [Semantic Scholar].

[117] S. Bond-Taylor, A. Leach, Y. Long, and C. G. Willcocks, *Deep generative modelling: A comparative review of vaes, gans, normalizing flows, energy-based and autoregressive models*, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **44** (2022), no. 11 7327–7347.

[118] S. Mohamed and B. Lakshminarayanan, *Learning in Implicit Generative Models*, arXiv:1610.03483.

[119] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, *Generative Adversarial Nets*, in *Advances in Neural Information Processing Systems* (Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K. Weinberger, eds.), vol. 27, Curran Associates, Inc., 2014. arXiv:1406.2661. [Semantic Scholar].

[120] A. Radford, L. Metz, and S. Chintala, *Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks*, 11, 2015. arXiv:1511.06434.

[121] D. P. Kingma and M. Welling, *Auto-Encoding Variational Bayes*, arXiv:1312.6114.

[122] C. Doersch, *Tutorial on Variational Autoencoders*, arXiv:1606.05908.

[123] D. J. Rezende and S. Mohamed, *Variational Inference with Normalizing Flows*, 5, 2015. arXiv:1505.05770.

[124] L. Dinh, J. Sohl-Dickstein, and S. Bengio, *Density estimation using Real NVP*, arXiv:1605.08803.

[125] J. Ho, A. Jain, and P. Abbeel, *Denoising Diffusion Probabilistic Models*, arXiv:2006.11239.

[126] L. Yang, Z. Zhang, Y. Song, S. Hong, R. Xu, Y. Zhao, W. Zhang, B. Cui, and M. Yang, *Diffusion models: A comprehensive survey of methods and applications*, *ACM Computing Surveys* **56** (Apr., 2024). Publisher Copyright: © 2023 held by the owner/author(s). Publication rights licensed to ACM.

[127] Y. Du and I. Mordatch, *Implicit generation and modeling with energy based models*, in *Advances in Neural Information Processing Systems* (H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, eds.), vol. 32, Curran Associates, Inc., 2019.

[128] J. Brehmer and K. Cranmer, *Simulation-based inference methods for particle physics*, arXiv:2010.06439.

[129] 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** (2014) 079, [arXiv:1405.0301].

[130] T. Gleisberg, S. Hoeche, F. Krauss, M. Schonherr, S. Schumann, F. Siegert, and J. Winter, *Event generation with SHERPA 1.1*, *JHEP* **02** (2009) 007, [arXiv:0811.4622].

[131] J. Bellm et al., *Herwig 7.0/Herwig++ 3.0 release note*, *Eur. Phys. J. C* **76** (2016), no. 4 196, [arXiv:1512.01178].

[132] W. Kilian, T. Ohl, and J. Reuter, *WHIZARD: Simulating Multi-Particle Processes at LHC and ILC*, *Eur. Phys. J. C* **71** (2011) 1742, [arXiv:0708.4233].

[133] **GEANT4** Collaboration, S. Agostinelli et al., *GEANT4 - A Simulation Toolkit*, *Nucl. Instrum. Meth. A* **506** (2003) 250–303.

[134] S. Badger et al., *Machine learning and LHC event generation*, *SciPost Phys.* **14** (2023), no. 4 079, [arXiv:2203.07460].

[135] O. Amram et al., *CaloChallenge 2022: A Community Challenge for Fast Calorimeter Simulation*, arXiv:2410.21611.

[136] **ATLAS** Collaboration, *ATLAS Software and Computing HL-LHC Roadmap*, tech. rep., CERN, Geneva, 2022.

[137] CMS Collaboration, "CMS Offline Computing Results." https://twiki.cern.ch/twiki/bin/view/CMSPublic/CMSOfflineComputingResults, 2025. [Accessed: 7 Oct. 2025].

[138] C. Bishop, *Pattern recognition and machine learning*, vol. 4. Springer New York, 2006.

[139] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016. http://www.deeplearningbook.org.

[140] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. u. Kaiser, and I. Polosukhin, *Attention is all you need*, in *Advances in Neural Information Processing Systems* (I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, eds.), vol. 30, Curran Associates, Inc., 2017.

[141] C. Krause, I. Pang, and D. Shih, *Caloflow for calochallenge dataset 1*, arXiv:2210.14245.

[142] **NNPDF** Collaboration, R. D. Ball et al., *An open-source machine learning framework for global analyses of parton distributions*, *Eur. Phys. J. C* **81** (2021), no. 10 958, [arXiv:2109.02671].

[143] F. Vaselli, F. Cattafesta, P. Asenov, and A. Rizzi, *End-to-end simulation of particle physics events with flow matching and generator oversampling*, *Machine Learning: Science and Technology* **5** (jul, 2024) 035007.

[144] **CMS** Collaboration, A. Rizzi, F. Vaselli, F. Cattafesta, and P. Asenov, *CMS FlashSim: End-to-end simulation with Machine Learning*, *EPJ Web Conf.* **337** (2025) 01014.

[145] C. ATLAS, M. Beckingham, M. Duehrssen, E. Schmidt, M. Shapiro, M. Venturi, J. Virzi, I. Vivarelli, M. Werner, S. Yamamoto, and T. Yamanaka, *The simulation principle and performance of the ATLAS fast calorimeter simulation FastCaloSim*, tech. rep., CERN, Geneva, 2010. All figures including auxiliary figures are available at

https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2010-013.

[146] *Performance of the Fast ATLAS Tracking Simulation (FATRAS) and the ATLAS Fast Calorimeter Simulation (FastCaloSim) with single particles*, tech. rep., CERN, Geneva, 2014. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-SOFT-PUB-2014-001.

[147] S. Abdullin, P. Azzi, F. Beaudette, P. Janot, A. Perrotta, and (on behalf ofthe CMS Collaboration), *The fast simulation of the cms detector at lhc*, *Journal of Physics: Conference Series* **331** (dec, 2011) 032049.

[148] **CMS** Collaboration, M. Hildreth, V. N. Ivanchenko, and D. J. Lange, *Upgrades for the CMS simulation*, *J. Phys.: Conf. Ser.* **898** (2017), no. 4 042040.

[149] A. Rogachev and F. Ratnikov, *GAN with an Auxiliary Regressor for the Fast Simulation of the Electromagnetic Calorimeter Response*, *J. Phys. Conf. Ser.* **2438** (2023), no. 1 012086, [arXiv:2207.06329].

[150] D. Belayneh et al., *Calorimetry with deep learning: particle simulation and reconstruction for collider physics*, *Eur. Phys. J. C* **80** (2020), no. 7 688, [arXiv:1912.06794].

[151] A. Xu, S. Han, X. Ju, and H. Wang, *Generative machine learning for detector response modeling with a conditional normalizing flow*, *JINST* **19** (2024), no. 02 P02003, [arXiv:2303.10148].

[152] J. Neyman and E. S. Pearson, *On the problem of the most efficient tests of statistical hypotheses*, *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* **231** (1933), no. 694-706 289–337. [Semantic Scholar].

[153] R. T. D'Agnolo and A. Wulzer, *Learning new physics from a machine*, *Phys. Rev. D* **99** (2019), no. 1 015014, [arXiv:1806.02350]. [Semantic Scholar].

[154] R. T. D'Agnolo, G. Grosso, M. Pierini, A. Wulzer, and M. Zanetti, *Learning multivariate new physics*, *Eur. Phys. J. C* **81** (2021), no. 1 89, [arXiv:1912.12155]. [Semantic Scholar].

[155] C. Krause and D. Shih, *Fast and accurate simulations of calorimeter showers with normalizing flows*, *Phys. Rev. D* **107** (2023), no. 11 113003, [arXiv:2106.05285]. [Semantic Scholar].

[156] M. Letizia, G. Losapio, M. Rando, G. Grosso, A. Wulzer, M. Pierini, M. Zanetti, and L. Rosasco, *Learning new physics efficiently with nonparametric methods*, *Eur. Phys. J. C* **82** (2022), no. 10 879, [arXiv:2204.02317]. [Semantic Scholar].

[157] G. Grosso, M. Letizia, M. Pierini, and A. Wulzer, *Goodness of fit by Neyman-Pearson testing*, *SciPost Phys.* **16** (2024) 123, [arXiv:2305.14137]. [Semantic Scholar].

[158] R. Das, L. Favaro, T. Heimel, C. Krause, T. Plehn, and D. Shih, *How to understand limitations of generative networks*, *SciPost Phys.* **16** (2024), no. 1 031, [arXiv:2305.16774].

[159] P. Cappelli, G. Grosso, M. Letizia, H. Reyes-Gonzalez, and M. Zanetti, *In preparation*, 2024.

[160] A. Kolmogorov, *Sulla determinazione empirica di una legge didistribuzione*, *Giornale Dell'instituto Italiano Degli Attuari* **4** (1933) 89–91. [Semantic Scholar].

[161] N. Smirnov, *Table for estimating the goodness of fit of empirical distributions*, *The annals of mathematical statistics* **19** (1948), no. 2 279–281. [Semantic Scholar].

[162] G. Peyré, M. Cuturi, et al., *Computational optimal transport*, *Foundations and Trends® in Machine Learning* **11** (2019), no. 5-6 355–607. [Semantic Scholar].

[163] C. Villani, *Topics in optimal transportation*, vol. 58. American Mathematical Soc., 2021. [Semantic Scholar].

[164] F. J. Massey Jr, *The kolmogorov-smirnov test for goodness of fit*, *Journal of the American statistical Association* **46** (1951), no. 253 68–78. [Semantic Scholar].

[165] S. S. Wilks, *The large-sample distribution of the likelihood ratio for testing composite hypotheses*, *Annals Math. Statist.* **9** (1938), no. 1 60–62. [Semantic Scholar].

[166] R. J. Tibshirani and B. Efron, *An introduction to the bootstrap*, *Monographs on statistics and applied probability* **57** (1993), no. 1 1–436. [Semantic Scholar].

[167] E. Edgington and P. Onghena, *Randomization tests (4th ed.)*. Chapman and Hall/CRC, 2007. [Semantic Scholar].

[168] L. Wasserman, *All of statistics: a concise course in statistical inference*. Springer Science & Business Media, 2013. [Semantic Scholar].

[169] M. Bińkowski, D. J. Sutherland, M. Arbel, and A. Gretton, *Demystifying MMD GANs*, in *International Conference on Learning Representations*, 2018. [Semantic Scholar].

[170] M. J. Chong and D. Forsyth, *Effectively unbiased fid and inception score and where to find them*, in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 6070–6079, 2020. [Semantic Scholar].

[171] A. Gretton, K. Borgwardt, M. Rasch, B. Schölkopf, and A. Smola, *A kernel method for the two-sample-problem*, *Advances in neural information processing systems* **19** (2006). [Semantic Scholar].

[172] A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Schölkopf, and A. Smola, *A kernel two-sample test*, *The Journal of Machine Learning Research* **13** (2012), no. 1 723–773. [Semantic Scholar].

[173] R. Kansal, C. Pareja, Z. Hao, and J. Duarte, *JetNet: A Python package for accessing open datasets and benchmarking machine learning methods in high energy physics*, *Journal of Open Source Software* **8** (Oct., 2023) 5789. [Semantic Scholar].

[174] "GMetrics: A tensorflow2 implementation of the metrics discussed in the paper, 2024,." GitHub.

[175] "GenerativeModelsMetrics: code and results for the two-sample tests with correlated gaussians and mixture of gaussians models, 2024,." GitHub.

[176] JetNetMetrics*: code and results for the two-sample tests with jetnet data, 2024,*, 2024. GitHub.

[177] R. Torre, *Metrics for two-sample tests: results on Mixture of Gaussians and Correlated Gaussians models*, Sept., 2024. [Zenodo].

[178] R. Torre, *Metrics for two-sample tests: results on JetNet dataset*, Sept., 2024. [Zenodo].

[179] J. Rabin, G. Peyré, J. Delon, and M. Bernot, *Wasserstein Barycenter and Its Application to Texture Mixing*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012. [Semantic Scholar].

[180] N. Bonneel, J. Rabin, G. Peyré, and H. Pfister, *Sliced and Radon Wasserstein Barycenters of Measures*, *Journal of Mathematical Imaging and Vision* **51** (2015) 22–45. [Semantic Scholar].

[181] F. Santambrogio, *Optimal transport for applied mathematicians*, *Birkäuser, NY* **55** (2015), no. 58-63 94. [Semantic Scholar].

[182] A. Ramdas, N. García Trillos, and M. Cuturi, *On wasserstein two-sample testing and related families of nonparametric tests*, *Entropy* **19** (2017), no. 2 47. [Semantic Scholar].

[183] P. J. Bickel, *A distribution free version of the smirnov two sample test in the p-variate case*, *The Annals of Mathematical Statistics* **40** (1969), no. 1 1–23. [Semantic Scholar].

[184] A. Justel, D. Peña, and R. Zamar, *A multivariate kolmogorov-smirnov test of goodness of fit*, *Statistics & probability letters* **35** (1997), no. 3 251–259. [Semantic Scholar].

[185] A. Coccaro, M. Pierini, L. Silvestrini, and R. Torre, *The DNNLikelihood: enhancing likelihood distribution with Deep Learning*, *Eur.Phys.J.C* **80** (2020) 664, [arXiv:1911.03305]. [Semantic Scholar].

[186] H. Reyes-Gonzalez and R. Torre, *Testing the boundaries: Normalizing Flows for higher dimensional data sets*, *J. Phys. Conf. Ser.* **2438** (2023), no. 1 012155, [arXiv:2202.09188]. [Semantic Scholar].

[187] A. Coccaro, M. Letizia, H. Reyes-Gonzalez, and R. Torre, *Comparative Study of Coupling and Autoregressive Flows through Robust Statistical Tests*, *Symmetry* **16** (2024) 942, [arXiv:2302.12024]. [Semantic Scholar].

[188] R. Kansal, A. Li, J. Duarte, N. Chernyavskaya, M. Pierini, B. Orzari, and T. Tomei, *Evaluating generative models in high energy physics*, *Phys. Rev. D* **107** (2023), no. 7 076017, [arXiv:2211.10295]. [Semantic Scholar].

[189] B. K. Sriperumbudur, A. Gretton, K. Fukumizu, B. Schölkopf, and G. R. Lanckriet, *Hilbert space embeddings and metrics on probability measures*, *The Journal of Machine Learning Research* **11** (2010) 1517–1561. [Semantic Scholar].

[190] S. Grossi, M. Letizia, and R. Torre, "Full numerical results on mog and cg." GitHub .

[191] S. Grossi, M. Letizia, and R. Torre, "Full numerical results on the jetnet dataset." GitHub .

[192] S. Grossi, M. Letizia, and R. Torre, "NPLM_Parameters_Tuning: code and results for the hyper parameters tuning of nplm.." GitHub .

[193] J. Neyman and E. S. Pearson, *On the Problem of the Most Efficient Tests of Statistical Hypotheses*, *Phil. Trans. Roy. Soc. Lond. A* **231** (1933), no. 694-706 289–337.

[194] T. Hastie, *The elements of statistical learning: data mining, inference, and prediction*, 2009.

[195] R. Barlow, *Extended maximum likelihood*, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **297** (1990), no. 3 496–506.

[196] A. J. Smola and B. Schölkopf, *Learning with kernels*, vol. 4. Citeseer, 1998.

[197] G. Meanti, L. Carratino, L. Rosasco, and A. Rudi, *Kernel methods through the roof: handling billions of points efficiently*, *Advances in Neural Information Processing Systems* **33** (2020) 14410–14422.

[198] A. Rudi, R. Camoriano, and L. Rosasco, *Less is more: Nyström computational regularization*, *Advances in Neural Information Processing Systems* **28** (2015). arXiv:1507.04717 [stat.ML].

[199] G. Grosso and M. Letizia, *Multiple testing for signal-agnostic searches for new physics with machine learning*, *Eur. Phys. J. C* **85** (2025), no. 1 4, [arXiv:2408.12296].

[200] K. Metzger, L. Xu, M. Sodini, T. K. Arrestad, K. Govorkova, G. Grosso, and P. Harris, *Anomaly-preserving contrastive neural embeddings for end-to-end model-independent searches at the LHC*, *Phys. Rev. D* **112** (2025), no. 7 072011, [arXiv:2502.15926].

[201] G. Grosso, N. Lai, M. Letizia, J. Pazzini, M. Rando, L. Rosasco, A. Wulzer, and M. Zanetti, *Fast kernel methods for Data Quality Monitoring as a goodness-of-fit test*, *Machine Learning: Science and Technology* **4** (2023), no. 3 035029, [arXiv:2303.05413].